# RNA: A JACK OF ALL TRADES.
# Studying the regulatory role of 6S RNA in *E. coli* and the impact of exosomal RNA in parasite pathogenesis

**by**

**Mariana Eugenia Oviedo Ovando**

M.Sc. (Molecular Biology and Genetics), University of Guelph, 2006
Microbiologa, Universidad Nacional de Rio Cuarto, 2002

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

in the
Department of Molecular Biology and Biochemistry
Faculty of Science



**SIMON FRASER UNIVERSITY**

**Summer 2015**

# Approval

**Name:** **Mariana Eugenia Oviedo Ovando**

**Degree:** **Doctor of Philosophy**

**Title:** ***RNA: A JACK OF ALL TRADES. Studying the regulatory role of 6S RNA in E. coli and the impact of exosomal RNA in parasite pathogenesis***

**Examining Committee:** **Chair:** Dr. Lisa Craig
Associate Professor

**Dr. Peter J. Unrau**
Senior Supervisor
Professor

**Dr. Carl Lowenberger**
Supervisor
Professor

**Dr. Jack Chen**
Supervisor
Professor

**Dr. Mark Paetzel**
Internal Examiner
Professor
Department of Molecular Biology and Biochemistry

**Dr. Ute Kothe**
External Examiner
Associate Professor
Department of Chemistry & Biochemistry
University of Lethbridge

**Date Defended/Approved:** July 24, 2015

# Abstract

During the past two decades we have seen an explosion in our understanding of RNA dependent gene regulation. We now know that RNA is involved in every major event in the life of cells, from the Okazaki fragments involved in DNA replication to programmed cell death. The work described here explores two situations in which RNA plays an important role; how 6S RNA helps ensure bacterial survival and the role of RNA in helping the intracellular parasite, *Leishmania,* escape the immune system by taking refuge inside mammalian macrophages.

6S RNA is a non-coding RNA that regulates bacterial transcription by sequestering the RNA polymerase holoenzyme ($E\sigma^{70}$) in low nutrient conditions. In high nutrient environments, $E\sigma^{70}$ is released by the synthesis of a short product RNA (pRNA) using the 6S RNA as a template. A range of 6S RNA release-defective mutants were selected and characterized from a highly diverse *in vitro* pool. There is complex crosstalk between regions of the 6S RNA large open bubble that interact with $E\sigma^{70}$ in a cooperative manner so as to ensure efficient pRNA-dependent release. When a group of 6S RNA mutants was over-expressed in *E. coli*, they significantly delayed growth and decreased cell survival indicating that 6S RNA release rate plays a key role in regulating normal transcriptional dynamics and ultimately cell division. Interestingly, cells resumed normal growth rates approximately 6 hours after mutant 6S RNA overexpression. This growth pattern might be correlated with the accumulation of a protein factor that binds strongly to the 6S and mutant 6S RNA, and data suggest that 6S RNA also might bind to an RNase.

RNA may contribute directly to parasite pathogenesis in trypanosomatids. *Leishmania spp*. uses exosomes to weaken mammalian host cells. Exosomes are known to be involved in intercellular communication. We examined the use of exosomes and their RNA from two species of *Leishmania* and how that RNA reprograms host cells. Exosome RNA cargo is delivered to host cell cytoplasm during *in vitro* studies. Sequencing of exosomal RNA indicated that the majority of cargo sequences were derived from non-coding RNA, while Northern blotting confirmed the specific and selective enrichment of tRNA-derived small RNAs in exosomes. We also identified a number of novel transcripts, which appeared to be specifically enriched in exosomes compared to total cell RNA. To our knowledge this is the first report that exosomes are used by a pathogen to invade new host cells. These findings also open up a new avenue of research on non-canonical, small RNA pathways in trypanosomatid parasites, which may elucidate pathogenesis factors and identify novel therapeutic targets.



**Keywords**: 6S RNA; transcription regulation; regulatory RNA, *E. coli* growth, *Leishmania*, exosomes

*A los mejores papas del mundo. Negrita y Quico, por siempre agradecida con uds.*

*To the best parents one could ask for. To Negrita and Quico, I'm forever thankful to you.*

## Acknowledgements

Being a graduate student has been unlike any other job. Being a graduate student is not something you do during working hours, it's who you are. Life and school became one and navigating both, has been an amazing experience that has taught me a lot more than I ever thought I would learn.

First of all, I would like to thank my supervisor, Dr. Peter Unrau for his constant support, patience and trust through my time in working in his laboratory. I am also grateful to my committee members, Dr. Chen and Dr. Lowenberger for their advice and guidance throughout the years as well as their advice during the preparation of this manuscript.

I also thank the external and internal examiner, Dr. Ute Kothe and Dr. Paetzel for taking time out of their busy schedules to attend my defense. I appreciate the time spent and the things learned from present and past members of the Unrau Lab; Alex Ebhardt, Matthew Lau, Lena Dolgosheina, Lindsay Shephard, Neil Dobson, and Lyssa Martin. Thank you Sunny Jeng and Ulrike Lambertz for your feedback and comments on previous versions of this thesis.

Thanks to previous MBB Chair Dr. Bruce P. Brandhorst for helping me during a time of exceptional circumstances. Also, thanks to current Chair, Dr. Lynne Quarmby for her advice and support. Finally, I would like to thank the staff of the MBB office: Nancy Suda, Mimi Fourie, and Christine Beauchamp for their constant help with a friendly smile.

The people that supported me outside the lab were equally important to encourage me to complete this work. I am grateful for my son Liam, whose smile gives me the strength and motivation to keep going everyday. To my family Cory, my sisters Luciana and Claudia. My friends and loved ones were also a big part of my life, Gabriela, Ulrike, Viviana, among many others with whom I've shared different aspects of this unique experience.

# Table of Contents

# List of Tables

# List of Figures

# List of Acronyms

| | |
|---|---|
| Å | Angstrom |
| AGO | Argonaute |
| BLAST | basic local alignment search tool |
| cDNA | complementary DNA |
| CDS | (protein) coding sequence |
| CE | cell extract |
| CIP | calf intestinal phosphatase |
| CIR74 | chromosomal internal repeats, 74-nucleotide long |
| CS | Cold Shock |
| DNA | deoxyribonucleic acid |
| ds | double stranded |
| E | core RNA polymerase |
| EBV | Epstein-Barr virus |
| ENS | enteric nervous system |
| FBS | fetal bovine serum |
| $_{f}$C | Final Concentration |
| HBSS | Hank's buffered salt solution |
| KDa | kilo Dalton |
| LINE | long interspersed elements |
| lncRNA | long non-coding RNA |
| miRNA | micro RNA |
| mRNA | messenger RNA |
| MW | molecular weight |
| nt | Nucleotide |
| PAGE | polyacrylamide gel electrophoresis |
| PBS | phosphate buffered saline |
| piRNA | Piwi interacting RNA |
| PNK | polynucleotide kinase |
| PSTVd | potato spindle viroid |
| RNA | ribonucleic acid |
| RNAi | RNA interference |

| | |
|---|---|
| RNAP | RNA polymerase |
| rRNA | ribosomal RNA |
| RT | reverse transcription |
| scRNA | small cytoplasmic RNA |
| SDS | sodium dodecyl sulfate |
| SFU | Simon Fraser University |
| SINE | short interspersed elements |
| siRNA | small interfering RNA |
| SL | spliced leader |
| SLACS | spliced leader-associated conserved sequence |
| snoRNA | small nucleolar RNA |
| snRNA | small nuclear RNA |
| sRNA | small RNA |
| srpRNA | signal recognition particle RNA |
| ss | single stranded |
| TAP | tobacco acid phosphatase |
| TAS | telomere-associated sequence |
| TATE | telomere-associated transposable element |
| tRF | tRNA-derived RNA fragment |
| tRNA | transfer RNA |
| tsRNA | tRNA-derived small RNA |
| TSS | Transcription start site |
| U | Unit |
| UTR | untranslated region |
| Vol | Volume |
| vRNA | vault RNA |

Chapter 1.

# Introduction

## 1.1. The flow of information is tightly regulated at many different levels in living organisms

### 1.1.1. DNA and RNA, the basics

Nucleic acids are biomolecules formed by polymers of nucleotides. Nucleotides are formed by three main components: a nitrogenous base, a ribose sugar and a phosphate. Nitrogenous bases (or nucleobases) are derived from purines formed by two N and C containing rings, while pyrimidines consist of a single ring (Figure 1.1). The main purine bases used in nucleic acids are adenine (Ade) and guanine (Gua). The pyrimidine bases in nucleic acids are cytocine (Cy) and thymine (Thy) in DNA and uracil (Ura) in RNA. N1 in the pyrimidine and N9 in purines are linked to a sugar via the $C_1$ carbon to form nucleosides. In RNA the pentose sugar is a ribose, while in DNA it is a deoxyribose. Nucleotides are composed by the nucleoside with one or more phosphate groups that are attached to the nucleosides by nucleoside monophosphate kinases (Van Rompay, Johansson *et al.* 2000, Kuhnel and Luisi 2001). The pH inside cells is ~ 7 and varies slightly depending on the type of cell. Due to a slightly basic pH, the phosphate groups are negatively charged. Phosphodiester bonds are the links between nucleotide bases. The ester bond is formed between the 5' carbon of a sugar has a hydroxyl or phosphate group that reacts with the free hydroxyl group on the 3' carbon in the sugar of the newly incorporated nucleotide. In nature, nucleic acid synthesis always proceeds 5' to 3'. Although the primary structures of DNA and RNA are similar, their conformations are very different and are critical to determining their unique functions (Van Rompay, Johansson *et al.* 2000, Alberts 2002, Lehninger, Nelson *et al.* 2008).

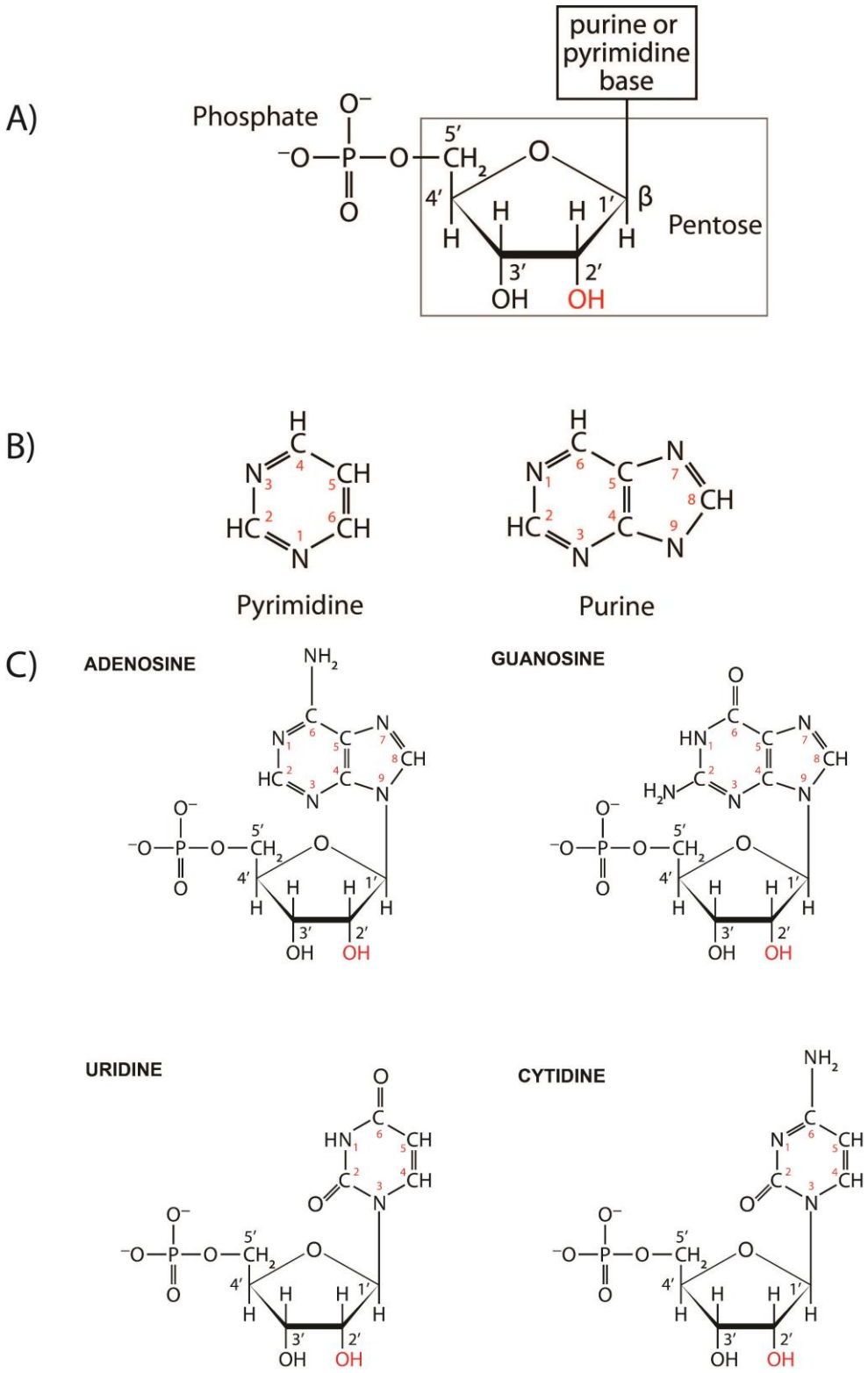



***Figure 1.1. General structure of nucleotides***

A) General nucleotide structure, they are the backbone of nucleic acids. Nucleotides are formed by a ribose (which includes the conventional numbering), a phosphate group and a purine or pyrimidine base. The main structure in black represents a ribonucleotide, while in the deoxyribonucleotides a H replaces the 2' OH (in red). B) Purine and pyrimidine rings are the scaffolds for the bases that will form the shown nucleotides. The base is bound to either ribose or deoxyribose via a beta-glycosidic linkage. C) RNA nucleotides as monophosphate versions.

Structurally, the most stable conformation for DNA consists of two intertwined polynucleotide strands. DNA is a double helix with the sugar-phosphate backbone on the outside and the bases stacked on top of each other in the interior of the double helix. The strands have an anti-parallel orientation and two factors are mainly responsible for the stability of the DNA double helix base pairing between complementary strands and stacking between adjacent bases (Yakovchuk, Protozanova *et al.* 2006). The absence of hydrogen bonds between consecutive nucleotides allows DNA to bend and interact with DNA-binding proteins. This is an essential property that enables the DNA to be tightly packaged. The base pairs are stacked about 0.34 nm apart in the DNA double helix. The planar sides of the base pairs are relatively non polar and provide the hydrophobic core that contributes considerably to the stability of the DNA double strand. In principle, the two DNA strands can form either a right-handed or left handed helix (Lodish and Darnell 2000, Richmond and Davey 2003). Due to the geometry of the sugar-phosphate backbone, natural DNA is a right-handed helix known as the B form of the DNA. The B form helix has stacked bases 0.34 nm apart along the helix axis, as shown by X-ray diffraction. This helix makes a complete turn every 3.4 nm, so there are about 10 base pairs per turn. The spaces existing between the intertwined DNA strands have different widths and are described as minor and major grooves (Richmond and Davey 2003). These two binding surfaces are used by different DNA-binding proteins. In very low humidity conditions, the crystallographic structure of B DNA shifts to the A form where bases are more tightly packed than the B form, having 11 bases per turn, and consequently, the stacked bases are slightly tilted. RNA-DNA, RNA-RNA duplexes, and dehydrated DNA exist naturally in the A configuration (Ghosh and Bansal 2003, Richmond and Davey 2003). The Z form of DNA is substantially different from the conformations previously described as it consists of two strands of DNA coiled in left handed helices with a zig-zag pattern of the phosphate backbone. DNA can adopt this configuration when the DNA sequence alternates between purine and pyrimidine bases

forming sequences like GCGCGCGC or when segments of the DNA become methylated (Saenger 1984, Lodish and Darnell 2000, Richmond and Davey 2003, Lehninger, Nelson *et al.* 2008). The sum of all of the DNA characteristics make it highly stable and resistant to degradation and in consequence the ideal molecule for maintaining and transmitting genetic information from generation to generation.

RNA is structurally more similar to protein than DNA. Due to its linear primary structure, RNA can fold in on itself to form highly structured areas connected by linear, flexible stretches of nucleotides (Lodish and Darnell 2000). As previously mentioned, RNA nucleotides have a ribose sugar with a 2' hydroxyl, which makes RNA more chemically labile and also more chemically reactive than DNA. Unlike DNA – which usually exists as a long double strand – RNA can exhibit a variety of lengths and conformations (primary or linear, secondary or tertiary) that allow them to execute different functions in the cell. The simplest secondary structures formed by single-stranded RNA are called "Hairpins" and "Stemloops". These relatively simple structures arise from the pairing of complimentary bases within the same RNA strand that occur when the nucleotides are separated by ~ 5-10 nt or by ~ 50 to several hundred nucleotides. They cooperate to form more complex tertiary order structures, such as "Pseudoknots" (Cech and Bass 1986, Clancy 2014). More examples of the structures and functions of RNA will be discussed in further detail later in this chapter.

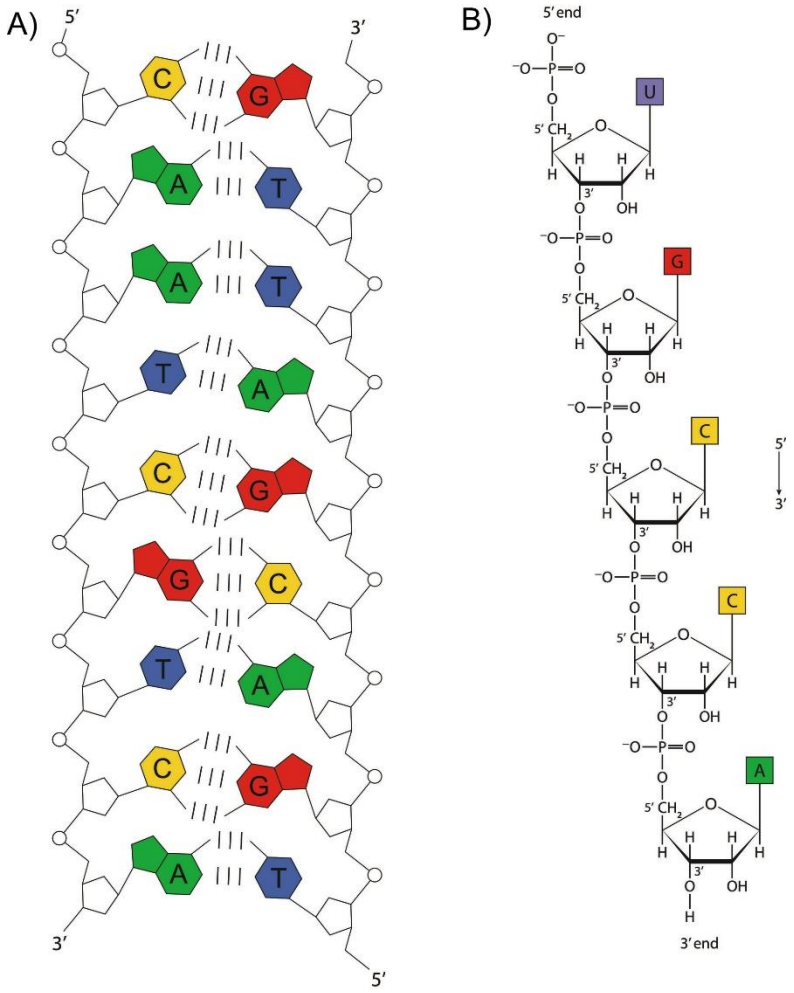



***Figure 1.2. DNA and RNA structures***

The presence or absence of an OH group in the 2' carbon of the pentose makes RNA chemically and structurally very different from DNA. A) Given its biological function, DNA is biologically a double-stranded molecule in which anti-parallel strands remain joined by hydrogen bond interaction indicated by vertical bars between coloured nucleotide bases. B) RNA is commonly found in the cell with a single-stranded structure. The above panel shows a linear RNA arrangement but it is often folded into secondary and tertiary structures.

RNA's structural flexibility makes it a molecule capable of executing a wide variety of enzymatic and regulatory roles, as well as functioning as the intermediate between DNA and proteins. The two most prominent examples of RNA's involvement in catalysis are the ribosome and RNase P. These are the only RNAs capable of carrying out multiple turnover reactions and processing multiple substrates. Both RNAs are part of a protein-RNA complex which has the RNA at its catalytic core. The ribosome is

essential during translation, while RNase P is involved in the processing and maturation of tRNAs and other types of RNA. These structures will be also discussed in further detail in this chapter.

Nucleic acids (DNA and RNA) as well as proteins are the heart of a cycle that ensures the storage and transfer of information in living organisms. The regulation of the delicate biological balance maintained by all sorts of organisms – from viruses to plants and human beings – is carried out mainly by proteins and regulatory RNAs. The next sections will describe these events, while retaining the dissertation's primary focus on the importance of the regulation performed by RNAs.

### 1.1.2. Biology's central dogma and the flow of information

Crick's 1958 and 1970 papers provided the framework for understanding how biopolymers carry and transmit information. The major classes of bio-molecules DNA, and RNA have a unique biochemical make-up and encode information in a way that can be transferred residue by residue. The central dogma proposes that genetic information is shared unidirectionally in a living system and flows from DNA → RNA → Protein (Crick 1970).

DNA's main role in the cell is to store and carry information. Transcription is the process that produces RNA, using DNA as a template. The main enzyme involved in this process is RNA polymerase. Proteins are the main molecules that provide structure and function as the regulators and catalyzers of different reactions in the cell. Proteins are produced in a process called translation, in which the ribosome is the heart of the process (Figure 1.3). The delicate balance between the above mentioned events ensures an accurate and efficient flow of information that in turn ensures continuity of life (Crick 1970, Alberts 2002). Early research on mainly bacterial organisms indicated that information flows unidirectionally and that genes encode for proteins using RNA as an intermediate. Prokaryotic genomes are largely composed of protein coding genes, 5' and 3' cis-regulatory elements, tRNA, and rRNA genes. Only about 0.5% of these prokaryotic genomes contain non-protein coding RNAs (Argaman, Hershberg *et al.* 2001). In more recent years, research has shown that RNA has an unexpectedly wide

variety of regulatory and catalytic roles, which suggest that the flow of information is not exclusively unidirectional, and that there is still much left to be discovered about the RNA's roles (Novina and Sharp 2004, Sharp 2009). Many examples of the exception to the classical view of RNA as a mere intermediary between DNA and protein are currently known. As illustrated in Figure 1.3, RNA can function as a template for its own replication or a template for DNA production during reverse transcription (Alberts 2002, Novina and Sharp 2004). It is worth mentioning that the DNA/RNA/protein central dogma is too complex to have arisen *de novo* since it requires both nucleic acids and proteins to be simultaneously generated from abiotic organic components (Orgel 2004). However, the discussion of the origins of these molecules and the origins of life itself are beyond the scope of this dissertation.

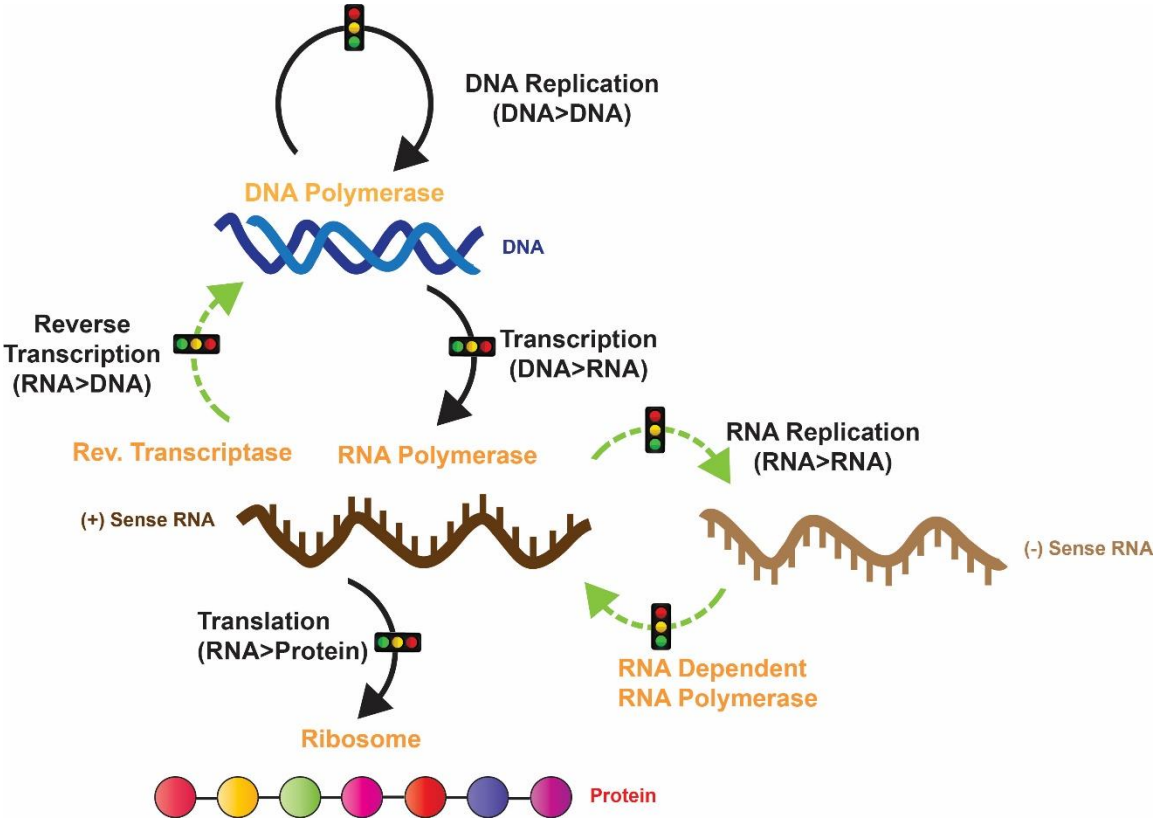



***Figure 1.3. Biology's central dogma***

The black arrows describe the unidirectional flow of information from DNA→ RNA→ Protein as initially proposed by Francis Crick (1970). The dotted green lines indicate the more recently discovered pathways that produce DNA using RNA as a template as well as the RNA replication pathway. The name of the main enzymes involved in each one of the pathways, is in orange. The traffic light on top of the arrows indicated that they are critical points tightly regulated by either Proteins or RNAs.

Crucial events such as DNA replication, transcription, and translation, are tightly regulated and will be later discussed. Early on it was assumed that most of this regulation was conducted by proteins at different levels (e.g., lac operon or tryptophan operon) (Matthews 1996, Alberts 2002). During the last two decades, research has shown that non-protein coding RNA plays an essential role in regulating an ever growing number of genomic programming events, especially in higher organisms (Mattick 2001). Small RNA (sRNA) has various roles related to growth, the regulation of metabolism or adaptation to stress (Michaux, Verneuil *et al.*), quorum sensing, biofilm formation, adaptation to growth under different conditions (such as different carbon sources), and pathogenesis (Hammer and Bassler 2007, Bardill, Zhao *et al.* 2011, Revelles, Millard *et al.* 2013, Michaux, Verneuil *et al.* 2014). Throughout this chapter, the relevance of RNA and its regulatory capabilities will be highlighted, and in some cases, compared to protein-mediated regulatory events.

### *DNA replication*

During replication two identical copies of a double stranded DNA (dsDNA) molecule will be generated and one of them will be used as a template in a process referred to as semiconservative replication. DNA polymerase is the DNA-dependent enzyme at the heart of this process. This enzyme can only extend existing fragments, a limitation that will be overcome by using short RNA fragments as primers to start the replication process. DNA polymerases are highly accurate, with an intrinsic error rate of less than one mistake every $10^7$ nucleotides added. Some DNA polymerases also have the proofreading ability to remove nucleotides from the end of a growing strand in order to correct mismatched bases. Finally, post-replication mismatch repair mechanisms scan the DNA looking for mismatches in the newly synthesized DNA strand. Taking advantage of these three mechanisms helps to maintain replication fidelity and generally results in the polymerase making less than one mistake for every $10^9$ nucleotides added (McCulloch and Kunkel 2008).

The replication process is generally divided into three steps: initiation, elongation and termination. The initiation process starts in specific areas of the DNA known as “replication origins”, which are labelled by DnaA initiation protein in *E. coli*. This protein recognizes A/T rich areas, which are easier to unwind, and recruits proteins to the site to

form the pre-replication complex (Weigel, Schmidt *et al.* 1997, Lodish and Darnell 2000). During the elongation phase the new DNA strand is synthesized in a 5' → 3' direction. Most organisms also have an enzyme called primase, which will generate short fragments of RNA with a free 3' OH end that DNA polymerase can hold onto, in order to proceed with the synthesis. The leading strand will receive one primer, while the lagging strand receives several RNA primers. The leading strand is extended continuously by DNA polymerase III, which is characterized by high processivity, while the lagging strand is extended discontinuously from each primer, forming Okazaki fragments. An RNase then removes the primer RNA fragments, and the low processivity DNA polymerase I, enters to fill in the gaps. The nicks remaining will be eventually filled in by the Ligase that completes the newly replicated DNA molecule.

As the DNA synthesis progresses, the DNA continues to unwind with the help of helicases that break the hydrogen bonds, while holding the two DNA strands together and forming a replication fork. Since the bacteria has a single origin for replication, the opening of the circular chromosome creates a "theta structure". As the helicases unwind the DNA, they generate additional torsion on the DNA ahead of them, which builds up resistance that could stall the formation of the replication fork (which is also ahead of them). Topoisomerases (including DNA gyrase) relieve this tension by temporarily breaking the strands of DNA, unwinding them, and adding negative supercoils to the DNA helix. Single strand binding proteins interact with the individual strands to prevent them from folding back into secondary structures that could stall the movement of the DNA polymerase. Since bacteria have circular chromosomes, replication terminates when the two replication forks meet on the opposite end of the parental chromosome. *E. coli* uses termination sequences that bind to the Tus protein and allow only one direction of the replication fork to pass through. As a result, the replication forks are always constrained to meet within the chromosome's termination region (Lodish and Darnell 2000, Alberts 2002, Lehninger, Nelson *et al.* 2008).

In bacteria, the replication origin and the initiator protein DnaA are the main targets for regulation. In *E. coli,* the replication origin activity is regulated by DNA methylation and specific oriC-binding proteins. Regulation performed by oriC-binding proteins and *dnaA* gene expression is conserved in bacteria. The replication origin has

several DnaA binding sites, while DnaA contains ATP/ADP-binding as well as DNA-binding domains. When enough ATP-DnaA has accumulated in the cell, an active initiation complex can be formed at the origin of replication, resulting in strand opening and the recruitment of helicase. Proteins that stimulate ATP-DnaA hydrolysis yield inactive ADP-DnaA, while DnaA-binding DNA elements control the subcellular localization of DnaA or stimulate the ADP-to-ATP exchange of the DnaA-bound nucleotide. The regulation of dnaA gene expression is also important for initiation (Skarstad and Katayama 2013).

**Regulation of DNA replication by RNA**

**Okazaki Fragments**

The generation of short RNA fragments that will form the Okazaki fragments is a good example of the importance of RNA regulating essential cellular events. These RNA fragments are indispensable, given that DNA polymerase cannot initiate *de novo* synthesis. The RNA fragments are especially important in the lagging strand, to which they are added in order to overcome the requirement that DNA polymerase must synthesize in the 5' → 3' direction. In bacteria the RNA fragments are generated by DnaG, which has primase activity. Primases are recruited by an interaction with the replicative helicase DnaB, which increases DnaG affinity for ssDNA (Corn and Berger 2006). The polymerase proceeds with the discontinuous polymerization and generates the Okazaki fragments, which are approximately 1200 nt long in bacteria and about 200 nt long in eukaryotes (Balakrishnan and Bambara 2013). The polymerase III adds about 1200 nt/sec, which allows bacteria to copy their entire genome and divide every 30 min or so. The ligation process in bacteria is much slower than the time required to actually synthesize the fragments, so if the Okazaki fragments in the lagging strand were shorter, joining them would take too long and the pace of the synthesis would not be able to keep up with the cellular division (Balakrishnan and Bambara 2013). Once the synthesis of the lagging strand has been completed, RNase H removes the RNA primers and polymerase I, which has a low processivity, performs its proofreading function to fill in the gaps. Finally, a ligase brings the fragments together to complete the synthesis of the lagging strand (Alberts 2002). The processing of the Okazaki fragments is one of life's fundamental processes because without it, semi-conservative DNA replication could not

occur. The generation and processing of the Okazaki fragments can be optimized in any particular organism for speed, fidelity, and energy consumption (or a combination of these). Such optimization is achieved by the crosstalk between different proteins, – especially helicases and primases – and by post-translational regulation of replication proteins. Speed and energy consumption would appear to be the most important factors in bacteria because they are competing with other rapidly growing cells.

**Long non- coding RNAs (lncRNAs) and chromatin remodelling**

Chromatin remodeling is another means by which RNAs are involved in the regulation of cellular events. The development of new bioinformatics approaches to analyzing the large amounts of data generated by different high-throughput methods has allowed for the identification of an unexpectedly large number of non-protein coding RNAs. Some scientists think that unless a specific function is determined, the mere fact that these RNA motifs are transcribed is not enough to indicate that they have a function (Kowalczyk, Higgs *et al.* 2012). Those who hold a contrasting point of view suggest that there must be a reason for this transcription, especially when cells like bacteria run such a tight economy and only use resources in events that will benefit them in one way or the other. Long non-coding RNAs (lncRNAs) are defined as RNA genes longer than 200 bp that do not appear to encode for proteins. This somewhat arbitrary cut-off will help to distinguish lncRNAs from small RNAs (sRNA) such micro RNAs (miRNA), Piwi interacting RNAs (piRNA) or small interfering RNAs (siRNA) (Rinn and Chang 2012).

Many lncRNAs facilitate regulatory proteins' access to the chromatin in order to influence the expression of individual genes or group of genes. In lot of cases, that role is achieved by working hand on hand with DNA methylation (Lai and Shiekhattar 2014). This was the first epigenetic mechanism identified and involves the modification of DNA by the addition of a methyl group (-CH3) to the cytosine or adenine of DNA nucleotides. Methylation typically occurs at CpG sites, usually clustered together close to transcription start sites (TSS) where they are mainly un-methylated (Riggs 1975, Bird 1986, Bunz 2005). X chromosome dosage compensation in mammals is a very well studied example of regulation via lncRNA. The genetic expression of two X chromosomes in female cells is levelled to the single X chromosome present in male cells via the expression of the lncRNAs Xist, which alters the chromatin structure of the

entire X chromosome and inhibits the transcription of most of its genes (Wutz 2011). Air is an lncRNA involved in the imprinting of paternally inherited genes. It is transcribed from the paternal copy of a silenced gene and recruits histone H3 and lysine 9 methylase G9a in order to mediate the transcriptional silencing of the *Kcnq 1* loci (Nagano, Mitchell *et al.* 2008). In the case of mammals, DNA methylation can also be regulated by lncRNAs working independently. In the case of plants, it can be regulated by IncRNAs working in concert with siRNAs (Law and Jacobsen 2010). Further research is required, however to understand the relationship between structure and function and detect common motifs in the RNA structure that could govern specific protein interactions.

### *Transcription*

RNA is synthesized by the DNA dependent RNA polymerase (RNAP) in all cells. In the past decade there has been a huge expansion of the amount of structural and biophysical data available on RNAPs from *Thermus aquaticus* (Campbell, Korzheva *et al.* 2001) and RNAP II from *Saccharomyces cerevisiae* (Cramer, Bushnell *et al.* 2001) or *Escherichia coli* (Murakami 2013), to name just a few model organisms. Since RNA is present in all cellular organisms it is expected that RNAPs are all derived from a final, common universal ancestor. This is apparent in their amino acid sequence, subunit composition, structure, function and molecular mechanisms, which are very similar from human to bacteria (Werner and Grohmann 2011). The X-ray structures for yeast RNAP II and *Taq* core RNAP are very similar in shape and size. These structures have a crab claw-like shape with an internal channel of 27 Å in diameter. The enzyme's active site is located on the back wall of the channel, where an essential $Mg^{+2}$ ion is bound (Cramer, Bushnell *et al.* 2001, Murakami and Darst 2003). The sequence, structure and function for the RNAP's catalytic core consist of 5 subunits ($\alpha$I, $\alpha$II, $\beta$, $\beta$', $\omega$) that are highly conserved (Archambault and Friesen 1993, Zhang, Campbell *et al.* 1999, Cramer, Bushnell *et al.* 2001). Promoter-specific initiation is the step that requires a σ factor to bind to the core RNAP in order to form the holoenzyme. The interaction between the RNAP core and the σ is very stable, with a dissociation constant of $10^{-9}$ M. However, when the RNAP enters the elongation phase the σ factor is rapidly released. These somewhat conflicting properties can be explained by the fact that the binding with RNAP is determined not by one – but by several – discrete associations between the structural

elements of σ ($\sigma_{1.1}$, $\sigma_2$, $\sigma_3$, $\sigma_4$ domains and the $\sigma_{3.2}$ loop) and the different parts of the core, with none of these interactions being particularly stable (Murakami and Darst 2003). Unlike DNA polymerase, RNAP is capable of *de novo* synthesis. It is able to do this because specific interactions with the initiating nucleotide hold RNAP rigidly in place, facilitating a chemical attack on the incoming nucleotide. Such specific interactions explain why RNAP prefers to start transcripts with ATP (followed by GTP, UTP, and then CTP) (Kennedy, Momand *et al.* 2007). In contrast to DNA polymerase, RNAP includes helicase activity, so no additional enzyme is needed to unwind the DNA (Yin and Steitz 2004).

Bacterial transcription occurs in the cytoplasm simultaneously with translation. Transcription is tightly controlled by a variety of regulators in prokaryotes. Many of these transcription factors are homodimers containing helix-turn-helix DNA-binding motifs (Huffman and Brennan 2002), while others are the more recently discovered RNAs (Breaker and Joyce 2014). Bacterial transcription happens in three stages, and the two last stages are highly dynamic (Figure 1.4). The bacterial RNA polymerase binds to a σ factor to form a holoenzyme which is capable of recognizing promoters through their specific '-35' and '-10' areas (Figure 1.4a). The holoenzyme will transcribe the sequence between the +1 position and the terminator sequence. At this point, the DNA and holoenzyme form a "closed complex" as the DNA has not been unwound yet (Figure 1.4b). Next, the DNA unwinds and the β subunit of the RNA polymerase starts transcription in the "open complex". The first attempts at transcription produce short abortive products (~ 10 nt long, Figure 1.4c) that are unable to leave the RNA polymerase because the exit channel is blocked by the σ factor (Alberts 2002, Goldman, Ebright *et al.* 2009). Promoter clearance (Figure 1.4d) occurs mechanistically through a scrunching mechanism in which the strain energy built up by DNA folding onto itself provides the energy needed to break the interactions between RNA polymerase holoenzyme and the promoter that are thought to occur by contacts between the nascent RNA and the proteins blocking the exit channel (Revyakin, Liu *et al.* 2006). During elongation, transcription becomes highly processive and the RNA polymerase continues to use base pair complementarity with the DNA template to create an RNA copy. Transcription occurs in a 5'→3' direction and produces an exact copy of the coding strand where T's have been replaced by U's. Elongation also involves a proofreading

mechanism that can replace incorrectly incorporated bases, but this mechanism is not as thorough as the proofreading mechanisms at work during DNA replication (Alberts 2002, Lehninger, Nelson *et al.* 2008, Chen, Darst *et al.* 2010). RNA polymerase lacks nuclease activity, and in contrast with DNA polymerase, RNA polymerase does not correct the nascent polynucleotide chain. Consequently, the fidelity of transcription is much lower than that of replication with error rate in the order of one mistake per $10^4$ or $10^5$ nucleotides, about $10^5$ times as high as that of DNA synthesis.(Sydow and Cramer 2009). This much lower fidelity of RNA synthesis can be tolerated because mistakes are not transmitted to progeny. For most genes, many RNA transcripts are synthesized; a few defective transcripts are unlikely to be harmful.

There are two common ways of finishing bacterial transcription, intrinsic termination or Rho independent termination and Rho-dependent termination. During intrinsic termination, transcription stops occurring when the new RNA forms a hairpin loop rich in G-C's, followed by a run of U's. When the hairpin loop forms it contributes to the breakdown of the bond between the RNA and DNA, which pulls the poly U transcript out of the RNA polymerase site, effectively terminating transcription. During Rho-dependent termination, a protein factor called 'Rho' destabilizes the interaction between the template and the mRNA, releasing the newly synthesized mRNA from the elongation complex (Richardson 2002, Lehninger, Nelson *et al.* 2008).

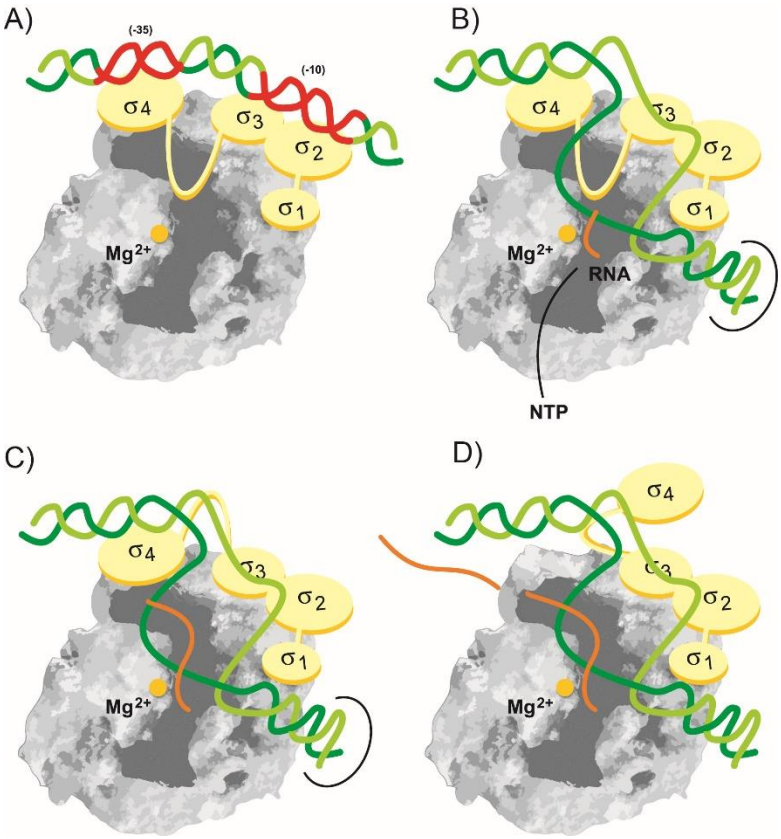



***Figure 1.4. Transcription is a combination of static and highly dynamic events***

During transcription there are two largely static states– binding (A) and initiation (B)– while the elongation stage is highly dynamic and processive. Transcription starts with the binding of the σ factor (shown schematically in yellow) to the core RNA polymerase (in grey). The holoenzyme, particularly its σ factor, recognizes specific promoters through their conserved '-35' and '-10' domains (in red). The initiation of RNA synthesis is abortive (C) and inefficient, but after the polymerase has synthesized about 10 nt of RNA, it shifts to the Elongation mode (D), and transcription becomes highly processive after the σ factor becomes disengaged disengages from the core RNA polymerase [Modified from Murakami and Darst (Murakami and Darst 2003)].

## Global regulation of transcription

### Classic modes of transcriptional regulation

Although RNAP is the main player in the transcription process, its activity is complex and highly regulated. More than 100 transcription factors have been identified In *E. coli* alone that modify the activity of RNAP (Ishihama 2000). In addition, different types of promoter and σ factors contribute to the modulation of individual gene expression. The global regulation of transcription is mainly carried out by a few main

elements, promoters, specific sigma (σ) factors, and transcription factors. The regulation of the *lac* operon or *tryptophan* operon provide examples of regulation by proteins or transcription factors (Matthews 1996, Alberts 2002). The focus of this chapter will be mainly on global regulation, not the individual regulation of genes or group of genes.

Promoters are regions of the DNA in which the reading of a transcriptional unit is initiated. Bacterial promoters contain two areas of conserved sequence: at '-10' and at '-35' (Alberts 2002). As previously mentioned, these are the areas that will be recognized by the holoenzyme. Promoters are typically classified as strong or weak depending on how closely their sequence resembles that of the ideal consensus sequence for a specific holoenzyme complex. Strong promoters will have a nucleotide sequence very similar to the known consensus sequence, while weak promoters will diverge from the consensus sequence. Weak promoters are usually associated with genes that have low transcriptional rates (Harley and Reynolds 1987).

The role of σ factors in transcriptional initiation is critical in eubacteria, as they confer promoter-binding specificity to the bacterial RNA polymerase (Qiu, Nagarajan *et al.* 2013). Different sigma factors are specific to different promoters and environmental conditions (Dove, Darst *et al.* 2003, Dong and Schellhorn 2009). Several σ factors in bacteria can bind to the core RNA polymerase and are ultimately responsible for recognizing DNA promoters. Factors such as $\sigma^{70}$ in *E. coli* (SigA in other bacteria) belong to the group 1 of σ factors also known as primary or "housekeeping" factors. The factors in this group contain highly conserved regions of amino acid sequence divided into four domains: σ1.1 (region 1.1), σ2 (regions 1.2 to 2.4), σ3 (regions from 3.0 to 3.1), and σ4 (region 4.1 to 4.2) (Murakami and Darst 2003). Other σ factors have more defined roles in the cell. The flagellar sigma factor is $\sigma^{28}$ (RpoF), which along with $\sigma^{70}$ and $\sigma^{54}$ have been found in cells that are actively growing or in the stationary phase. During the exponential phase, $\sigma^{28}$ is the second highest level of sigma factor after $\sigma^{70}$. The factor $\sigma^{19}$ (FecI), or ferric citrate sigma factor, regulates the *fec* gene involved in iron transport. The extracytoplasmic/extreme heat stress sigma factor $\sigma^{24}$ (RpoE) poorly transcribes regulons in the absence of (p)ppGpp. Also known as guanosine pentaphosphate, (p)ppGpp causes the inhibition of RNA synthesis during amino acid shortage, which causes translation rates to decrease hence conserving amino acids.

Furthermore, ppGpp triggers the up-regulation of many other genes involved in stress response such as the genes for amino acid uptake and biosynthesis (Srivatsan and Wang 2008). (p)ppGpp, along with DksA and 6S RNA, form a group of effectors that facilitate the transcription switch during conditions of starvation and stress. The nitrogen-limitation sigma factor $\sigma^{54}$ (RpoN) also exhibited a dependency on DksA and (p)ppGpp, and represents 10% of the total σ factor present during exponential growth. The heat shock sigma factor $\sigma^{32}$ (RpoH) is active when bacteria are exposed to heat. Due to its enhanced expression, it is highly probable that the $\sigma^{32}$ factor will bind to the core RNAP. As a result, other heat shock proteins are expressed that will enable the cell to survive at higher temperatures. Chaperones, proteases and DNA-repair enzymes (Gruber and Gross 2003, Sharma and Chatterji 2010) are some of the enzymes expressed when $\sigma^{32}$ is activated. The starvation/stationary phase sigma factor is $\sigma^{38}$ or $\sigma^{s}$ (RpoS), which transitions from undetectable levels during exponential phase to at least 30% of the σ factor present during the stationary phase. This factor is the one with the lowest affinity for the RNAP, so it can be thought of as the "last resort" for cell survival. When $\sigma^{s}$ binds, it will activate genes that help the cell to respond to starvation and various stress conditions like DNA damage, the presence of reactive oxygen species, or osmotic shock (Farewell, Kvint *et al.* 1998, Hengge-Aronis 2002, Sharma and Chatterji 2010).

Competition between sigma factors e for a limited amount of core RNAP has been experimentally proven while varying the level of a particular sigma factor by regulated expression or by generating low-affinity sigma factor mutants (Zhou, Walter *et al.* 1992). These specific sigma factors will interact with their corresponding promoters and the strength or weakness of the promoters will determine what genes will get transcribed under a specific set of environmental conditions (Sharma and Chatterji 2010). This is one of the many strategies that cell use to distribute a limited amount of RNAP to promoters or maintenance or survival related genes (Sharma and Chatterji 2010). A 2014 report describes the use of a theoretical model that looks into the binding between sigma factors and core RNAP, transcription, non-specific binding to DNA and the modulation of the availability of the molecular components. The model is validated by comparison with *in vitro* competition experiments and found that transcription is affected via the modulation of the concentrations of the different types of holoenzymes, so

saturated promoters are only weakly affected by sigma factor competition. Interestingly, in the case of promoters recognized by two types of sigma factors, they found that even saturated promoters are strongly affected. Active transcription effectively lowers the affinity between the sigma factor driving it and the core RNAP, resulting in complex cross-talk effects. The finally also looked at the role of increased core RNAP availability upon the shut-down of ribosomal RNA transcription (which represents 75% of the total transcription in rapidly growing bacteria) during the stringent response, and found that passive up-regulation of alternative sigma-dependent transcription is not only possible, but also displays hypersensitivity based on the sigma factor competition (Mauri and Klumpp 2014). This report provides an interesting view of the global switches of genetic expression in bacteria, but it should be kept in mind that the report considers only steady state situations and has been validated against *in vitro* experiments. The authors emphasize that having more data about *in vivo* concentrations of different holoenzymes under specific conditions (different growth rates, or time points, entry into stationary phase or responding to different kinds of stress) would allow a better understanding of sigma factor competition in physiological situations.

## RNAs as modulators during transcription

### *6S RNA*

The non-coding RNA 6S RNA seems to be the only RNA capable of modulating global transcription rates (Nitzan, Wassarman *et al.* 2014). The non-coding RNA 6S RNA was one of the first RNAs to be sequenced (Brownlee 1971). It took more than three decades to understand its function (Wassarman and Storz 2000). The *E. coli* chromosome has one copy of the gene ssRS that encodes the 6S RNA (Hsu, Zagorski *et al.* 1985). The mature 6S RNA derives from a longer precursor or primary transcript. Interestingly, the synthesis of the 6S RNA is not tightly regulated as its expression persists even during amino acid starvation. Two promoters that regulated expression were found upstream of the gene (Lee, Park *et al.* 2013). Transcription from the P1 promoter generates a shorter transcript with 9 additional nucleotides in the 5' end that are processed in a well-known process by RNase G and E. The transcript generated from the P2 promoter contain 224 additional nucleotides and is processed exclusively by RNase E (Kim and Lee 2004, Steuten, Hoch *et al.* 2014). P1 is a canonical $\sigma^{70}$

dependent promoter, while P2 is dependent on both, $\sigma^s$ and $\sigma^{70}$. Transcription of the two promoters is terminated in a Rho-dependent manner 90 nt downstream of the mature 3' end (Chae, Han *et al.* 2011). The regulation of the ssrS gene P1 and P2 promoters is intimately related to the bacterial growth phase. In addition, the depletion of 6S RNA did not have an observable effect on growth-phase-dependent transcription from either promoter, whereas the overexpression of 6S RNA increased P1 transcription and decreased P2 transcription (Hsu, Zagorski *et al.* 1985). These results suggest that transcription from P1 and P2 is subject to feedback activation and feedback inhibition, respectively. The differential feedback regulation may provide a means for maintaining appropriate cellular concentrations of 6S RNA throughout varying growth conditions (Lee, Park *et al.* 2013).

A report by Barrick in 2005 described hundreds of 6S RNA bacterial homologs using a computational approach (Barrick, Sudarsan *et al.* 2005). Recent data confirms that 6S RNA might be more widespread than previously thought. A more updated review of all bacterial genomes found a set of 1,750 6S RNA genes. Some 1,367 6S RNA genes are novel, and distributed among 1,610 bacteria (Wehner, Damm *et al.* 2014). Structural probing also revealed that the secondary structure was a common feature among 6S RNA homologs. The conserved structure is precisely what allows 6S RNA to bind to the $E\sigma^{70}$ by mimicking the structure of a DNA template (Figure 1.5) in an open promoter complex conformation (Barrick, Sudarsan *et al.* 2005). Furthermore, the action mechanisms are basically identical in organisms as diverse as the hyperthermophile *Aquifex aeolicus,* and *E. coli* or *B. subtillis* (Köhler, Duchardt-Ferner *et al.* 2015).

The majority of the nc-RNAs participate in post-transcriptional regulation by base pairing to mRNA or by sensing and binding small ligands in the same manner that riboswitches do. One of the characteristics that make 6S RNA unique is the fact that unlike other nc-RNA, 6S RNA affects global transcription by binding directly to the housekeeping $E\sigma^{70}$ and ultimately has the power to directly control transcriptional rates. Secondly, 6S RNA prompts the $E\sigma^{70}$ to function as an RNA-dependent RNA polymerase when producing an RNA product (pRNA). When *E. coli* is under low nutrient conditions, 6S RNA binds to $E\sigma^{70}$ (Wassarman and Storz 2000). Shephard and colleagues used *in*

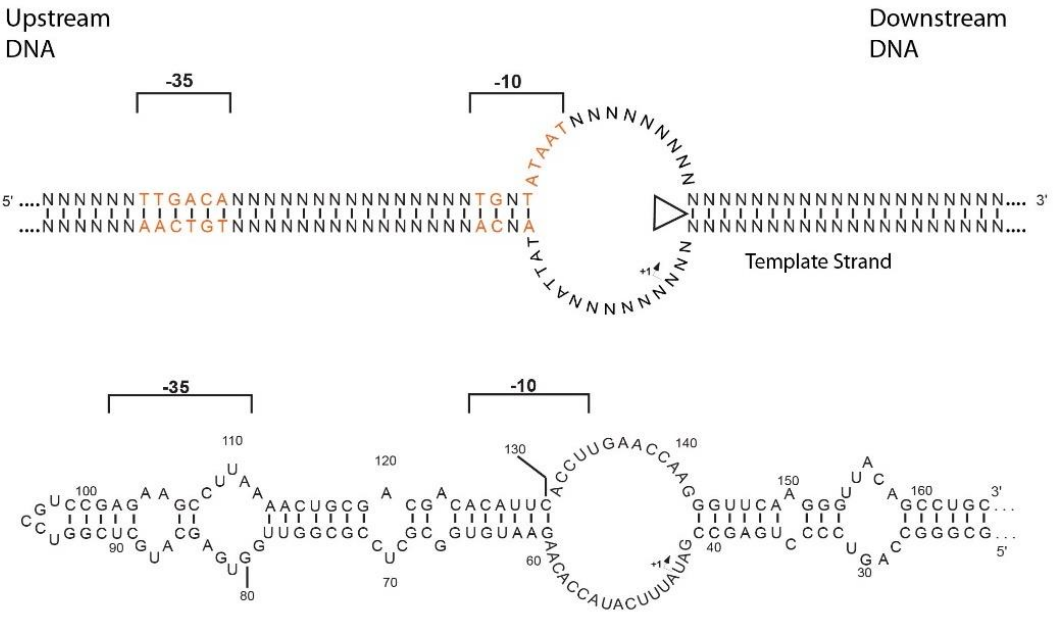



***Figure 1.5 The 6S RNA resembles a transcriptional elongation complex***

The 6S RNA secondary structure resembles a transcriptional elongation complex because it has '-35' and '-10' like areas (in red in the top panel) with conserved bases (Barrick, Sudarsan *et al.* 2005), a transcription start site (indicated by +1 and a black arrow), and a central large open bubble. The bottom strand will be used as a template during pRNA synthesis.

*vitro* selection to screen a high diversity library containing ~4 X $10^{12}$ sequences in order to understand how the 6S RNA sequence controls binding to the E$\sigma^{70}$. Residues that played a critical role in binding were found in the '-35' region, which is phylogenetically conserved. Mutating some of the residues in this region led to decreases in binding, and removing them abolished binding completely. Unexpectedly, the mutation of residues found in the conserved '-10' region were found to influence 6S RNA release rates, in addition to modulating -35 binding. These results suggest that 6S uses a natural mechanism to differentiate between "strong" and "weak" promoters. 6S RNA binding and release seems to be a dynamic process (Figure 1.6) in which the release rates have been fine-tuned over time in order to correctly regulate cellular levels of transcription (Shephard, Dobson *et al.* 2010). When nutritional conditions change, the 6S RNA is used as a template for the synthesis of a short product RNA (pRNA) that triggers a conformational change. This change will allow the RNA polymerase to become available and resume transcription under the control of $\sigma^{70}$ or other σ factors (Wassarman and Saecker 2006). The 6S RNA conformational change includes the

formation of a release hairpin when the pRNA reaches the critical length of 9 nt (Panchapakesan and Unrau 2012, Steuten and Wagner 2012), which triggers the ejection of $\sigma^{70}$. As the pRNA continues growing in length enough scrunching forces are accumulated to ultimately trigger the release of 6S, freeing the core enzyme (E) (Panchapakesan and Unrau 2012).

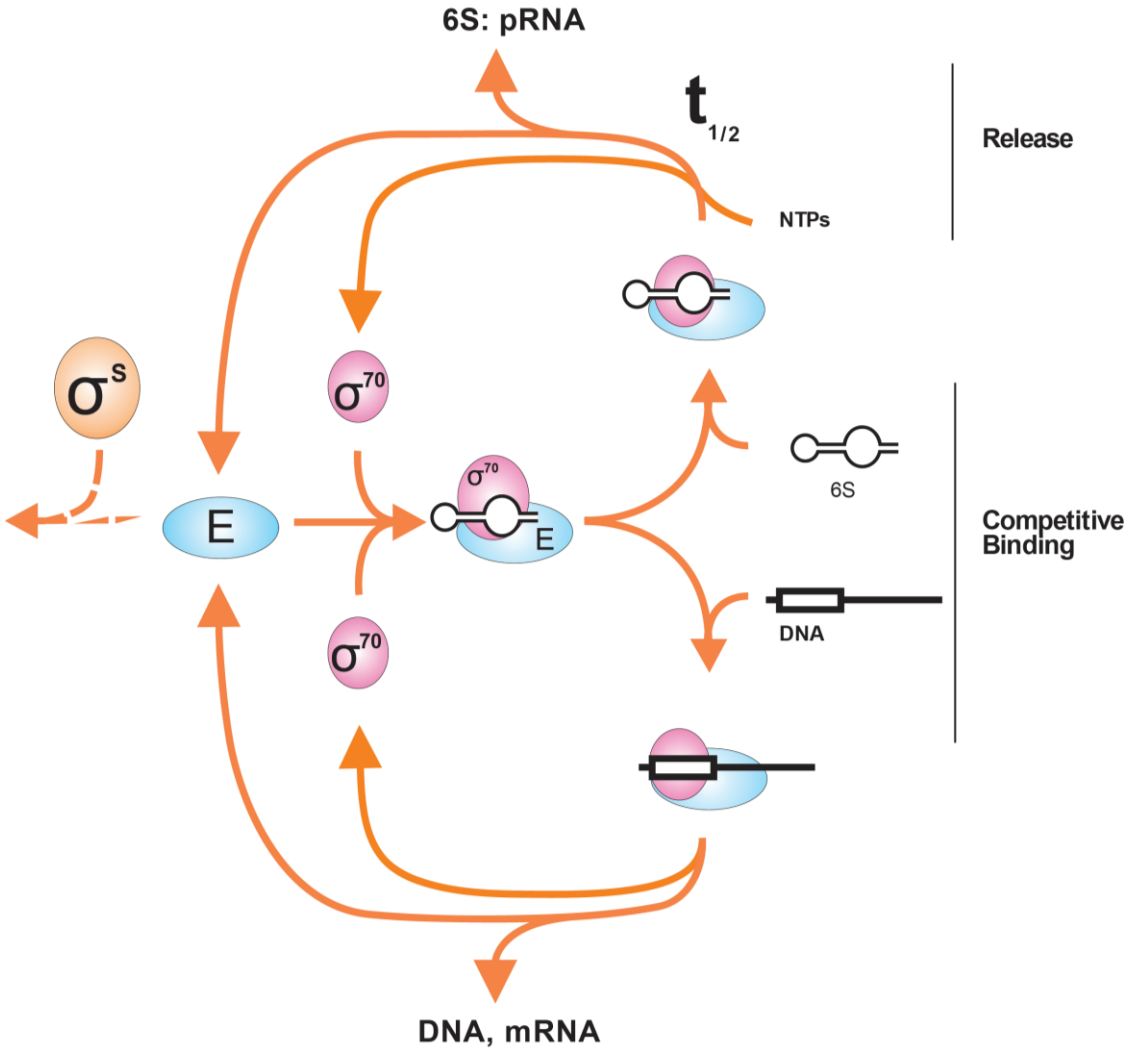



***Figure 1.6. Regulation of transcription during stationary phase.***

A summary of the two modes of 6SRNA-dependent regulation; 6S RNA binding affinity to $E\sigma^{70}$, and pRNA-induced release rate. Levels of free core enzyme (E, large circle in light blue) and free sigma factors ($\sigma^{70}$ in pink and $\sigma^{S}$ in orange) determine the formation rate of the $E\sigma^{70}$complex (superimposed circles). Such complex is depleted by the formation of either a bound 6SRNA: $E\sigma^{70}$ (top path) or a DNA: $E\sigma^{70}$ complex (bottom path). The characteristic time for pRNA-induced 6S RNA release ($t_{1/2}$), is sensitive to free magnesium and NTP levels, provides a mechanism to sequester core polymerase from use, effectively lowering the concentration of free polymerase (E) in the cell (Adapted from (Shephard, Dobson *et al.* 2010)).

*In vivo,* 6S RNA has a wide range of effects. 6S RNA deficient cells are viable – but less capable of surviving – during the stationary phase, have growth defects, and show a decrease in competitive fitness (Trotochaud and Wassarman 2004). A genome-wide transcriptional analysis of wild type and 6S RNA deficient cells found 245 genes during the exponential phase and 273 genes during the early stationary phase, to be differentially expressed by at least 1.5 fold. Up- and down-regulated genes include many transcriptional regulators, stress-related proteins, transporters, and enzymes involved in purine metabolism. There was a reduction in the expression of genes that encode for the translational machinery in the 6S RNA deficient strain. Furthermore, the lack of 6S RNA is compensated by an increase of ppGpp concentration, which is known to affect growth adaptation and ribosome biogenesis. The overall effect of 6S RNA on transcription is not strictly confined to $\sigma^{70}$ dependent promoters (Neusser, Polen *et al.* 2010).

### *Translation*

During translation the information contained in the mRNA, which was produced from transcription using the DNA as template, is decoded by the ribosome to produce a specific polypeptide chain. Although most biological activities are carried out by proteins, rRNA, tRNAs and mRNA are at the heart of the protein production pathway.

The ribosome is a ribonucleoprotein (RNP) complex that carries out protein synthesis in both eukaryotes and prokaryotes. This ribosome is the ultimate ribozyme, given the complexity of the reaction it catalyzes and is also capable of executing multiple catalytic cycles. It was first identified by Claude (Claude 1943) while he performed cell fractionation, and again identified by one of his students while that individual performed electron microscopy studies on mammalian and avian cells (Palade 1955). Later research showed that these particles, initially called “microsomes”, contained RNA and were essential for protein production. By the end of the 60's these particles had been isolated, chemically characterized, and named ‘ribosomes’. Actual proof that these substructures were involved in protein synthesis was obstained by using *in vitro* systems derived from eukaryotic and prokaryotic cells (Zamecnik 1966). Ribosomes are formed by two subunits of different sizes: the large subunit (LSU) and the small subunit (SSU). These two subunits are conserved among bacteria and eukaryotic cells, but there are

substantial differences in their detailed structures (Lake 1981, Lake 1985). Sizes for the two subunits vary between prokaryotes and eukaryotes, mainly due to the varying numbers and sizes of the ribosomal proteins present (Brimacombe, Stoffler *et al.* 1978, Wittmann 1983, Sommerville 1986). X-ray crystallography has made it possible to determine ribosome structure with atomic resolution. The interactions of tRNA molecules with the binding sites A, P, and E in the ribosome were determined in *Thermus thermophilus* (Yusupov, Yusupova *et al.* 2001). The discovery of RNA molecules with catalytic activity such as RNase P and the self-splicing introns (Cech, Zaug *et al.* 1981, Guerrier-Takada, Gardiner *et al.* 1983) promoted the extensive study of rRNAs. As a consequence, rRNAs were found to be indispensable during the *in vitro* assembly of functional ribosomes, whereas ribosomal proteins were shown to be largely dispensable. The omission of ribosomal proteins would produce ribosomes with decreased activity, but never a complete loss of function. The first crystal structure of the LSU, which is responsible for the peptidyl transferase reaction, demonstrated that there are no proteins in the vicinity (20 Å) of the active site – as indicated by the transition state analogue CCdAp-Puromycin, further confirming that the ribosome is indeed a ribozyme (Ban, Nissen *et al.* 2000, Nissen, Hansen *et al.* 2000). Earlier studies showed that rRNAs participate in all the steps of the translation process: the selection of mRNA, the binding of tRNA, the association between ribosomal subunits, translocation, proofreading, frame-shift suppression, antibiotic interactions, termination, and peptidyl-transferase activity (Nogi, Yano *et al.* 1991, Noller, Hoffarth *et al.* 1992, Green and and Noller 1997).

Four stages are involved in the translation that occurs in the bacterial cytoplasm: initiation, elongation, translocation, and termination. Initiation involves the formation of the translation system by the two ribosomal subunits (50S LSU, 30S SSU), the mature mRNA, the tRNA charged with N-formylmethionine (the first amino acid in the nascent peptide), guanosine triphosphate (GTP) as a source of energy, the prokaryotic elongation factor EF-P, and the three prokaryotic initiation factors IF1, IF2, and IF3 (which help in the assembly of the initiation complex). The A site in the ribosome is the point of entry for the aminoacyl tRNA (excepting the first aminoacyl tRNA, which enters at the P site). The P site is where the peptidyl tRNA is formed in the ribosome. The E site is the exit site for the tRNA after it gives its amino acid to the growing peptide chain.

The selection of the starting site, usually an AUG codon, will depend on the interaction between the 30S subunit and the mRNA. The 30S subunit binds to the mRNA template in a purine-rich region called the Shine-Dalgarno sequence, which is located upstream of the initiation codon. The Shine-Dalgarno sequence is complementary to a pyrimidine rich region on the 16S rRNA component of the 30S subunit. The conformation of the double-stranded RNA structure between the mRNA and the ribosome places the initiation codon precisely at the P site. Elongation starts when the fMet-tRNA enters the P site, triggering a conformational change that opens the A site for binding by the next aminoacyl-tRNA. This binding is facilitated by elongation factor-Tu (EF-Tu). The P site contains the beginning of the peptide chain, while the A site contains the next amino acid to be added. The growing polypeptide connected to the tRNA in the P site is detached and a peptide bond is formed between the last amino acid of the polypeptide and the amino acid that is still attached to the tRNA in the A site. The peptide bond formation is catalyzed by the ribozyme 23S rRNA in the 50S ribosomal subunit. At this point, the growing peptide chain is in the A site, while the P site contains an uncharged deacylated tRNA. During translocation the deacylated tRNA and the di-peptidyl tRNA – along with their corresponding codons – move to the E and P sites respectively, making room in the A site for the entry of a new aminoacyl-tRNA. The ribosome will continue to translate the rest of the codons until it encounters a termination codon. Termination starts when one of the termination codons moves onto the A site. These codons are recognized by release factor (RF-1 and RF-2) proteins that trigger the hydrolysis of the ester bond in peptidyl-tRNA, which results in the ribosome's release of the newly synthesized protein. A third release factor, RF-3, catalyzes the release of RF-1 and RF-2 at the end of the termination process (Green and and Noller 1997, Lodish and Darnell 2000).

The translation process has a slower rate when compared to DNA replication or transcription. Proteins in prokaryotes are synthesized at a rate of 18 amino acid residues per second, while (as mentioned earlier) DNA is synthesized at a rate of 1200 nucleotides per second (Martin, Bilgin *et al.* 1998, Balakrishnan and Bambara 2013). This large difference is due at least in part to the difference between polymerizing only four types of nucleotides to make nucleic acids and polymerizing 20 types of amino acids to make proteins. Testing and rejecting incorrect aminoacyl-tRNAs takes time and slows down protein synthesis. Two phenomena speed up the rate of protein synthesis in

bacteria, so the cells are still able to divide every 30 minutes or so. Therein, translation and transcription are coupled in bacteria and, translation initiation occurs as soon as the 5’ end of an mRNA has been synthesized. In addition, a single mRNA molecule can be simultaneously translated by several ribosomes, forming structures called polysomes that can be seen in the electronic microscope. This is not possible in eukaryotes because transcription and translation are carried out in separate compartments of the cell (the nucleus and cytoplasm) (Lodish and Darnell 2000).

**Translation regulation by RNA**

As indicated, RNA is a central part of protein synthesis. A few RNAs that have not been previously mentioned and are involved in the modulation and control of translation will be discussed below.

**Transfer RNA (tRNA)**

Transfer RNAs (tRNAs) are an essential component of the translational machinery. They are small molecules of about 75 to 94 nt long and cells contain several types of tRNA. Most tRNAs serve as carriers of amino acids and participate in protein synthesis. However, tRNAs also participate in other reactions that are not related to protein synthesis (Raina and Ibba 2014). The tRNA that binds to alanine ($tRNA^{Ala}$) was the first nucleic acid to be sequenced, earning Holley the 1968 Nobel prize in Physiology or Medicine (Holley 1966). All tRNAs have two main functions: to base pair with a codon in the mRNA and to be chemically linked to amino acids. The primary sequence in tRNAs allows them to fold into the characteristic clover leaf secondary structure. The 3' end of all tRNAs have the CCA sequence of 3' overhanging nucleotides that are usually added after the molecule has been synthesized and processed. Several other bases in the 3' end are typically modified as well. The attachment of the correct amino acid to the 3' end of the tRNA is carried out by aminoacyl-tRNA synthetase. The enzyme links the amino acid to the 2' or 3' hydroxyl of the terminal adenosine in a two-step ATP-dependent reaction (Arnez and Moras 1997, Lodish and Darnell 2000).

**Ribonuclease P**

Ribonuclease P (RNase P) is found in all bacteria, archaea and eukaryotic cells, and has a common origin and similar highly conserved elements. Is composed of two subunits, a 400 nt long RNA and a protein of about 120 residues, and is completely dependent on $Mg^{2+}$divalent ions (Frank and Pace 1998, Kazantsev and Pace 2006). Ribonuclease P plays a critical role for the translational apparatus catalyzing the maturation of the 5' pre-tRNA substrates (Konarska and Sharp, 1998). RNase P is a RNA-protein complex esential for cell viability. Along with the 23S rRNA in the ribosome, RNase P is the only ribozyme capable of carrying out multiple catalytic cycles and processing multiple substrates like all the pre-tRNAs, the 4.5 S RNA in bacteria, the bacteriophage induced Φ 80 RNA, the transfer messenger RNA (tmRNA), and the polycistronic mRNA from the *his* operon (among others).

**Shine-Dalgarno (SD) sequences**

Shine-Dalgarno (SD) sequences are the binding site for ribosomes in prokaryotic mRNA located 8 nt upstream of the start codon AUG. The sequence is present in bacteria and archaea as well as chloroplast and mitochondrial transcripts and helps to initiate protein synthesis by aligning ribosomes with the start codon. Shine and Dalgarno showed that the 3' end of the 16S rRNA in *E. coli* recognizes a complimentary purine-rich stretch of sequence in the 5' UTR (Shine and Dalgarno 1974). These sequences are important not only for translation initiation, but also for determining expression levels. A positive correlation between the presence of an SD sequence and the predicted expression level of a gene based on codon usage biases predicted that highly expressed genes are more likely to possess a strong SD sequence than average genes (Ma, Campbell *et al.* 2002).

**Riboswitches and local regulation**

Riboswitches are regulatory elements within the mRNA sequence that sense metabolite concentration and consequently control gene regulation (Grundy, Winkler *et al.* 2002, Winkler, Nahvi *et al.* 2002). They can control the expression of certain mRNA's during translation. Riboswitches are also involved in Rho-independent termination of transcription. Translation initiation is regulated in a similar way to intrinsic transcriptional

termination because of mutually exclusive base-paired structures that limit the ribosome's access to the Shine-Dalgarno sequences (Breaker 2012).

Besides the presence of specific promoters and sigma factors, termination is another way in which overall transcription rates can be modulated. Intrinsic termination, also known as Rho-independent transcription termination, is an example of how RNA helps to regulate transcription. It is a frequent mechanism underlying the activity of other *cis*-acting RNA regulatory elements, such as the riboswitches. The modulation of transcription termination is arguably one of the most common mechanisms for riboswitch mediated gene control. It occurs through the formation of a strong stem loop (7-20 bp long), that is followed by a run of uridine (U) residues. The formation of the stem pulls the oligo (U) region out of the RNA polymerase active site so as to terminate transcription (Gusarov and Nudler 1999, Yarnell and Roberts 1999) (Alberts 2002).

Although most known riboswitches have been found in bacteria, they have also been described in plant and fungi. Thousands of riboswitches have been identified in databases and the number of these that have been experimentally validated keeps growing. It is likely that only a small amount of the existing classes of riboswitches has been discovered and that many new classes await discovery due to the increasing number of DNA databases and improved methods of comparative analysis (Ames and Breaker 2010). The sequence for the ligand-binding aptamer area is the most conserved throughout the structure of the riboswitch. The sequence and structure of the riboswitch regulatory portion can vary considerably due to the many ways in which RNA structures can influence transcription, translation, and RNA processing (Barrick and Breaker 2007). Riboswitches are almost exclusively located in the 5' UTR of the mRNA they control, which allows them to be synthesized first and then respond to the metabolite present in the environment before the full length RNA or even entire operons are transcribed. Although riboswitches provide more of the local modulation during transcription for a specific set of genes or operons, they are worth mentioning because of their widespread use in bacteria.

**Transfer messenger RNA (tmRNA)**

Transfer messenger RNA is encoded by the SsrA gene in bacteria and has both, tRNA and mRNA-like properties. A ribonucleoprotein complex is formed around tmRNA with the EF-Tu, small protein B and ribosomal protein S1. The protein complex binds to bacterial ribosomes that have stalled in the middle of protein synthesis and can add a proteolysis-inducing tag to the unfinished polypeptide that will also facilitate the degradation of the defective mRNA (Keiler 2008). Trans-translation is essential in many bacteria to help them survive under stressful growth conditions (Thibonnier, Thiberge *et al.* 2008). The tmRNA occupies de A site in the stalled ribosome. Next, the ribosome shifts from the 3' end of the truncated mRNA onto the resume codon of the MLR. Then a slippage-prone stage in which translation resumes normally occurs, continuing until the in-frame tmRNA stop codon is encountered (Keiler, Waller *et al.* 1996, Thibonnier, Thiberge *et al.* 2008).

All tmRNAs have a common feature: a tRNA-like domain (TLD) with two helices 1, 12 and 2a that are analogs of the tRNA acceptor stem, T-stem and variable stem, respectively. Transfer messenger RNAs (tmRNA) also contain a 5' monophosphate and alanylatable 3' CCA ends (Williams and Bartel 1996, Felden, Himeno *et al.* 1997). The mRNA-like region (MLR) in a tmRNA is a large loop containing pseudoknots and a coding sequence (CDS) for the tag peptide that is circumscribed by the resume codon and the stop codon. The encoded tag peptide varies among bacteria, probably due to the fact that the available proteases and adaptors vary among them (Gur and Sauer 2008). Most tmRNAs are transcribed as large precursors and processed very similarly to tRNAs. Processing at the 5' end is carried out by Ribonuclease P, while the 3' is attacked by multiple exonucleases that will vary depending on the bacterial specie. Whether the 3' CCA is added by the tRNA nucleotidyltransferase or encoded in the tmRNA sequence will also depend on the bacterial species (Srivastava, Srivastava *et al.* 1992, Li, Pandit *et al.* 1998). The tmRNA, along with 6S RNA, is one of the few non-coding RNAs capable of widespread regulation, rather than the regulation of individual genes or group of genes.

### 1.1.3. Global regulation of RNA by recycling

The fact that bacteria have well established networks that help them to maximize the use and reuse of resources to repair cell injury in times of stress is another factor that contributes to the cell economy and the cell's capacity for fast growth and adaptability. Although any kind of stress or intrusion will alter the progression of the cell cycle and the growth rate for bacteria, the survival of the bacterial population is the ultimate goal.

The process by which a cell grows, accumulates the nutrients required for mitosis, and then duplicates its DNA, is called cell cycle. Traditionally, the cell cycle is divided into four stages. During the Gap 1 (G1) or growth phase, cells increase in size and biosynthetic processes actively produce enzymes and proteins that will be used in the next phase. The S phase starts with the duplication of DNA and ends when all chromosomes have been copied. DNA duplication is carried out as quickly as possible because the exposed DNA base-pairs are extremely sensitive to external agents like drugs or mutagens. Another Gap phase (G2) follows during which the cell resumes growth and G2 is a checkpoint to ensure that the cell is ready to enter Mitosis or the M phase, during which the cell will be divided into two identical daughter cells again. Mitosis is a highly regulated process that consists of several stages known as prophase, metaphase, anaphase, telophase, and cytokinesis (Cooper and Hausman 2007). When faced with a change in environment or stress (e.g. heat, chemical), bacteria need to rapidly adapt and adjust their cell cycle to the new conditions in order to survive. Bacteria are capable of adjusting and responding to stress and changes in the environment faster than any other living organism. Changes in bacterial growth rate are usually coupled with changes in the cell cycle to ensure that events such as mass doubling, chromosome replication and segregation remain synchronized. Multiple signalling pathways allow cells to sample their environments and adjust the progression of the cell cycle to changing conditions (Schaechter, Maaloe *et al.* 1958, Wang and Levin 2009) by eliminating the stress agent or repairing cell injury (Hengge-Aronis 2002).

Different kinds of stress will trigger specific kinds of adaptive responses. For example, when an exponentially growing culture is transferred from 37°C to 20°C, the cells stop growing and need to acclimatize. During this phase the cells synthesize a

small set of Cold Shock (CS) proteins that allow cells to resume growth at a lower rate (Gualerzi, Maria Giuliodori *et al.* 2003). These CS proteins include nucleic acid binding proteins (involved in processes like RNA degradation, transcription, DNA replication, translation) and five members of the Csp family (which associate with ribosomes). Transcript profiling during cold shock also showed that genes involved in sugar transport and metabolism as well as membrane synthesis and function also get induced (Phadtare and Inouye 2004). The mechanisms that regulate gene induction can be either transcriptional or post-transcriptional (Gualerzi, Maria Giuliodori *et al.* 2003). Exposure to Quinolones and a loss of virulence is another example of the stress bacteria can face. Quinolones are a broad spectrum chemical with antibacterial properties that were initially used to treat urinary infections. They inhibit topoisomerase II and thus interfere with DNA replication and transcription (Hooper 2001). Research indicates that quinolone-resistant strains express fewer virulence factors than quinolone-susceptible strains (Vila, Simon *et al.* 2002). Quinolones increase the frequency of loss of pathogenicity islands (PAI), which are the usual location for cytotoxic necrotizing factor and hemolysin. Exposure to quinolones induces the SOS system response [DNA repair mechanism (Giuliodori, Gualerzi *et al.* 2007)].

A multitude of factors like 6S RNA, tmRNAs, heat shock proteins, and the SOS repair system (among many others) are involved in rescuing a bacterial population under stress. It is very likely that a fraction of the bacterial population will be sacrificed in order to provide nutrients for the survivors, which will take advantage of metabolic pathways that allow cells to break down biomolecules and enable the building blocks to be used as emergency coping mechanisms.

### *Sigma Factor s ($\sigma^S$)*

The sigma factor $\sigma^S$ is one subunit that binds the RNA polymerase in *E. coli* as well as other bacteria during stationary phase and under many different stress conditions (Hengge-Aronis 1993). The $\sigma^S$ is viewed as a master regulator that gives cells the ability to survive existing and future stressors, which is called cross-protection (Hengge-Aronis 1996)]. $\sigma^S$ 's activity is a marked contrast to specific stress responses in that a single stimuli will result in the induction of a group of proteins that will help the cell to cope with a specific stress (Storz and Hengge 2000). The expression of $\sigma^S$ and all the genes

governed by it have a primarily preventive role. More than 70 gene products have been known to be activated by $\sigma^S$ and these provide resistance to oxidative stress, near-UV irradiation, potentially lethal heat shocks, hyperosmolarity, acidic pH, and EtOH. The $\sigma^S$ activated genes also alter the cell envelope and cell morphology, as stressed cells tend to be smaller and have an ovoid shape (Hengge-Aronis 2002). Metabolism is also affected when RNAP switches from a $\sigma^{70}$ to a $\sigma^S$ factor and goes from a metabolism of maximum growth to having a maintenance metabolism. Furthermore, the genes for programmed cell death get activated in order for a fraction of the cell population to be sacrificed in an attempt to provide nutrients and increase the bacterial population's chance of surviving extreme stress. Finally, genes associated with virulence have also been found to be under the control of $\sigma^S$, which agrees with the idea that invading host organisms provides a stressful environment for the pathogens (Emoődy, Pál *et al.* 2002).

The $\sigma^S$ is encoded by the *rpoS* gene and its transcription is stimulated by decreases in the growth rate, especially when these decreases are somewhat gradual (Lange and Hengge-Aronis 1994). Much of the regulation of *rpoS* occurs due to post-transcriptional mechanisms. Interestingly, relatively high levels of *rpoS* mRNA are present during exponential growth and do not seem to change in response to different stresses that elevate amounts of $\sigma^S$ protein. A large number of factors regulate *rpoS* transcription, such as promoters, ppGpp levels, trans-acting (BarA) factors, acetate and other weak acids, NADH/$NAD^+$ ratio (Hengge-Aronis 2002). The translational regulation of r*poS* is also a multi-factorial event. Under certain conditions such as hyper-osmolarity, low temperature, acidic pH, and/or reaching a critical cell density during the exponential phase, the translation of *rpoS* mRNA will be stimulated (Lange and Hengge-Aronis 1994). The secondary structure the *rpoS* mRNA adopts, seems to have an impact on translation regulation in that the TIR (translational initiation region) seems to be based-paired and inaccessible to ribosomes. It is hypothesized that certain stress signals trigger a change in the secondary structure of the mRNA in order to allow for more frequent translation initiation. However, there is little evidence that supports this theory. Predictions with MFOLD indicate that ~ 340 nt of the 5' end of the *rpoS* mRNA fold into a very stable cruciform structure, while the TIR further downstream can potentially adopt two different conformations. In the most plausible structure, the SD-sequence region is partially base-paired with an "internal antisense" region located

further upstream (Hengge-Aronis 2002). It is difficult to understand how the folding of the *rpoS* mRNA actually affects translation regulation. There are many protein factors and small RNAs that positively or negatively affect the RNA transcription, so theoretical calculations or *in vitro* probing will yield inaccurate or at least incomplete results on their own (Morita, Kanemori *et al.* 1999, Hengge-Aronis 2002). Hfq is one of the trans-acting factors known to interact with the *rpoS* mRNA and is necessary for its efficient translation. However, Hfq's exact role is still unclear (Tsui, Leung *et al.* 1994). As is the case during transcription, the translation of *rpoS* mRNA is affected by a wide variety of RNAs, protein factors, and small molecules. Finally, the control of $\sigma^S$ proteolysis will also strongly affect the overall levels of the sigma factor (Hengge-Aronis 2002).

The study of the ever-increasing number of factors that contribute to rpoS transcription and translation as well as to $\sigma^S$ proteolysis has shown that $\sigma^S$ appears to be one of the most complex regulation systems in *E. coli*. The basic control of rpoS translation relies on the use of the rpoS mRNA secondary structure, the Hfq and HU proteins, and small RNAs such as DsrA mRNA. The core $\sigma^S$ degradation machinery consists of the ClpXP protease and the phosphorylation-modulated $\sigma^S$ recognition factor RssB. There are many unanswered questions in addition to the basic ones addressed above. There is also a growing amount of evidence that $\sigma^S$ is connected to other complex regulatory circuits such as oxidative stress, which operates via OxyS RNA, or to the heat shock response, which involves the DnaK chaperone. These connections are highly relevant under simultaneous stress conditions, which are probably more natural situations than the single source of stress present in carefully controlled experiments. Taking the broad variety of genes that core $\sigma^S$ controls into account provides a good starting point for studying the cellular network landscape during stress.

### *Degradosome*

The half lives of many bacterial RNAs range from 40 seconds to 60 min, while some RNAs in eukaryotes can have half-lives as long as several days (Belasco and Brawerman 1993). It is becoming apparent that mRNA decay, RNA degradation as well as RNA processing, and maturation are somewhat interconnected processes (Deutscher 2006). The protein complex formed around the RNase E is called a degradosome and its main function is RNA degradation. The degradosome has major and minor

components. The major components are RNase E, PNPase (polynucleotide phosphorylase), RhlB helicase and enolase (Burger, Whiteley *et al.* 2011). The minor components include the cold shock helicase CsdA, protein regulators of ribonuclease activity RraA and RraB, the RNA chaperon Hfq, and an RNase E inhibitor r-protein L4 (Burger, Whiteley *et al.* 2011).

The structure of the degradosome's major components can be described as follows. The exoribonuclease PNPase is an 85 KDa protein that binds to RNase E as a trimer. The next major component is RhlB, which is an ATP-dependent helicase and part of the DEAD box helicase family. This protein has a size of 50 KDa, binds to the RNase E via its N-terminal end, and can associate with the PNPase as a dimer or multimer in the absence of RNase E. Enolase is made up of an N-terminal domain and a barrel C-terminal domain. Crystal structure analysis showed that the scaffold domain of the RNase E binds asymmetrically to the interface of the enolase dimer. Approximately 5-10% of all cellular enolase and 10-20% cellular PNPase co-purify with the RNase E as part of the degradosome (Py, Higgins *et al.* 1996, Kuhnel and Luisi 2001, Liou, Chang *et al.* 2002, Chandran and Luisi 2006).

The current understanding of the system indicates that RNase E works with Hfq by acting as a chaperon involved with small regulatory RNA-mediated cleavage or the gene-silencing of specific transcripts. Interestingly, other major degradosome components such as enolase and RhlB are absent from the Hfq-RNase E interaction. Continuous research on the degradosome shows that much more work needs to be done in order to fully understand this highly dynamic structure (Burger, Whiteley *et al.* 2011).

### *HFQ*

Hfq is a relatively small protein (11 KDa monomer) that participates in several cell mechanisms. Deleting Hfq in various species of γ-proteobacteria uncovered a set of pleiotropic effects related to bacterial fitness and responses to environmental stress. These results suggested that Hfq globally affects the physiology of the cell (Sobrero and Valverde 2012) and acts on various cellular mechanisms to affect functions such as growth rate, reduced stress tolerance, and attenuated virulence *(Le Derout, Folichon et*

*al. 2003, Sobrero and Valverde 2012)*. Hfq protein has no sequence specificity and favours binding to A/U rich single-stranded RNA regions *(Le Derout, Folichon et al. 2003)*.

Hfq is well-known primarily for its role in post-transcriptional gene regulation, which involves messenger RNAs (mRNAs) and small regulatory RNAs (sRNAs). In its hexameric form, Hfq mediates the binding of sRNAs to their trans-encoded mRNA targets, which affects the translational rate and lifetime of the mRNA, both positively and negatively (Zhang, Wassarman *et al.* 2002). The joining of a sRNA with its target mRNA often blocks the ribosome-binding site, preventing the initiation of translation (Storz, Opdyke *et al.* 2004) and allowing the mRNA to be degraded (Carpousis 2007). However, it also appears that Hfq utilizes several other mechanisms to exert its effects on translation and the turnover rates of particular transcripts.

Hfq's involvement as minor components of the degradosome and in RNA degradation *(Le Derout, Folichon et al. 2003)* is an interesting topic that is explored further in Chapter 3. When Hfq associates with RNase E, it acts as a chaperon involved with small regulatory RNA-mediated cleavage or the gene-silencing of specific transcripts. The hybrids mRNA-sRNA are rapidly degraded by the Hfq-RNase E interaction. The RNase E microdomain in which these two proteins interact is between residues 711 and 750 (Ikeda, Yagi *et al.* 2011).

## 1.2. RNA is not only an information containing molecule, it is also part of cellular regulatory machines in all domains of life

DNA replication, transcription and translation are the most important and essential cellular processes in which RNA plays a key role by directly regulating each step of the process. However, there are other cellular events that have different types of RNAs or foreign RNA fragments like RNA viruses or viroids playing essential roles that will have a widespread effect on the cell or organism.

### 1.2.1. CRISPR RNAs, the role in the bacterial immune system and as a surveillance mechanism

Clustered, regularly interspaced short palindromic repeats (CRISPR) are an essential component of the adaptive immune system present in bacteria and archaea. The selective pressure imposed by viruses that target both bacteria and archaea has driven the diversification of the microbial defense systems (Rodriguez-Valera, Martin-Cuadrado *et al.* 2009). In response to virus and plasmid challenges, bacteria and archaea integrate small fragments of the foreign nucleic acid into their own chromosome at the end of a CRISPR element. These loci maintain a molecular record of encounters with previous genetic parasites. The CRISPR loci are transcribed and processed into an array of short CRISPR-derived RNAs (crRNAs). Each crRNA is packaged into a large surveillance complex that mediates the detection and destruction of foreign nucleic acids (Brouns, Jore *et al.* 2008). CRISPRs were initially identified in the *E. coli* genome as sequence elements that consisted of a series of 29 nt repeats separated by a unique 32 nt long spacer sequence (Ishino, Shinagawa *et al.* 1987). Sequences with similar characteristics were found in phylogenetically diverse bacterial and archaeal genomes, but the meaning of these sequences became clearer when bioinformatics approaches identified the spacers as being identical to viral or plasmid sequences. This type of adaptive immunity was previously thought to be present only in eukaryotes and the discovery of short CRISPR-derived RNAs (crRNAs) suggested that there were similarities between CRISPR-based immunity and the RNA interference (RNAi) mechanism [Figure 1.7 (Wiedenheft, Sternberg *et al.* 2012)].

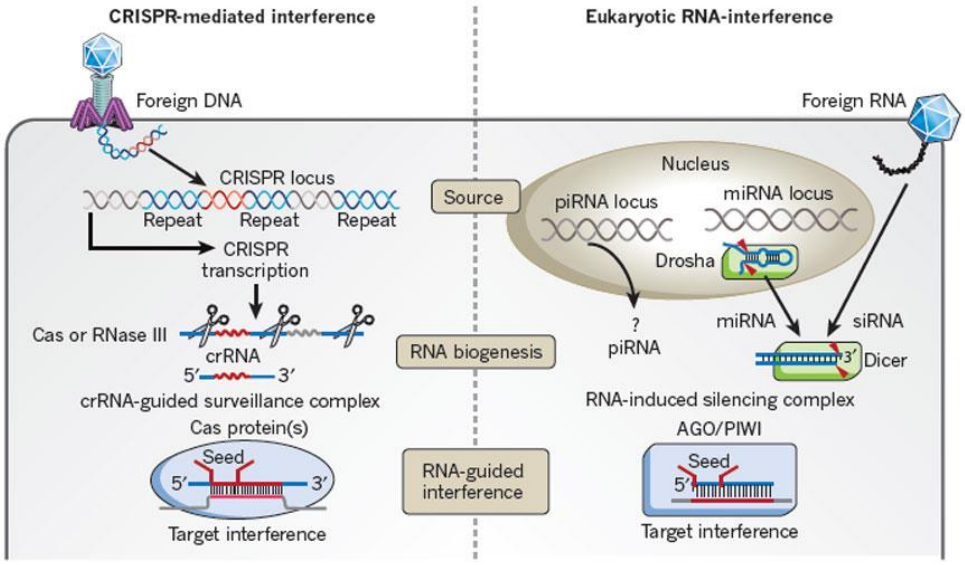



***Figure 1.7. Comparison between CRISPR RNA guided silencing systems and RNAi***

In CRISPR systems, foreign RNA is integrated into the CRISPR locus. Long transcripts from this loci are processed by Cas. The crRNAS are assembled into a ribonucleoprotein surveillance complex that destroys invading genetic material. In some eukaryotes, long dsRNA is recognized as foreign, Dicer enzymes and these cleave the RNAs into siRNAs that guide the immune system to invading viruses. PIWI-interacting RNAs are transcribed from repetitive clusters in the genome that often contain copies of retrotransposons and work mainly by restricting retrotransposon mobility. Micro RNAs are also encoded in the chromosome, and the primary transcripts are processed by Drosha or Dicer enzymes. Although the RNAs do not participate in genome defence, they are major players in gene regulation (Wiedenheft, Sternberg *et al.* 2012).

Bioinformatic analyses have shown that CRISPR loci have adjacent (A+T)-rich sequences that serve as promoters and are also flanked by a large set of *cas* genes. The *cas 1* gene is common to all CRISPR containing systems and encodes for a highly conserved protein (Makarova, Grishin *et al.* 2006). CRISPR systems have been divided into three major immune systems based on *cas* gene conservation and locus organization. The presence of one type of CRISPR system does not necessarily exclude the presence of others, indicating that they are compatible and can share functional components (Makarova, Haft *et al.* 2011). Type I systems have *cas3* genes and the protein contains an N-terminal HD phosphohydrolase domain and a C-terminal helicase domain. In some cases these two domains are encoded in different genes, but they participate in nucleic acid degradation regardless of that. The type II CRISPR system consists of 4 *cas* genes and one of these is always *cas9,* which encodes for a

large protein that has a RUVC-like nuclease domain and an HNH nuclease domain that may participate in CRISPR RNA processing and target destruction. Finally, there are two variations of type III CRISPR (A and B). Type III A has been reported in *S. epidermidis* targeting plasmid DNA *in vivo,* and the purified Type III B system found in *P. furiosus* only cleaves ssRNA substrates *in vitro (Makarova, Haft et al. 2011, Wiedenheft, Sternberg et al. 2012)*.

The acquisition of foreign DNA into the CRISPR loci is the first step during bacterial adaptive immunity. Several of the Cas proteins are involved in the process, but the mechanisms of integration and replication are still unclear. Mutational studies on *S. thermophilus* showed that *cas7* (also known as *csn2*) is required for the acquisition of new spacer sequences (Barrangou, Fremaux *et al.* 2007). Biochemical and structural data suggest that Cas1's role is related to the acquisition of new spacer sequences. The Escherichia coli Cas1–Cas2 complex mediates spacer acquisition *in vivo*. Doudna and colleagues showed that the purified Cas1–Cas2 complex integrates oligonucleotide DNA substrates into acceptor DNA to yield products similar to those generated by retroviral integrases and transposases. Cas1 is the catalytic subunit while Cas2 increases integration activity. Protospacer DNA with free 3'-OH ends and supercoiled target DNA are required. The integration occurs preferentially at the ends of CRISPR repeats and at sequences adjacent to cruciform structures bordering AT-rich regions, similar to the CRISPR leader sequence. These results demonstrate the Cas1–Cas2 complex to be the minimal machinery that catalyzes spacer DNA acquisition and explain the significance of CRISPR repeats in providing sequence and structural specificity for Cas1–Cas2-mediated adaptive immunity (Nunez, Doudna, *et al.*, 2015).

Purified Cas1 protein from *S. solfataricus* has shown to bind nucleic acids with high affinity, but without sequence preference. The same protein from *E.coli* binds to DNA with a preference for mismatched or abasic substrates. Activity assays for Cas1 from *E.coli* and *P. aeruginosa* indicate it is a metal-dependent nuclease. The crystal structures available for Cas1 from several species have diverse amino acid sequences, but the tertiary and quaternary conformations are similar. Cas 1 structure depicts a C-terminal domain with a divalent metal-ion binding site and alanine substitutions of the metal coordinating residues that inhibit Cas1-catalyzed DNA degradation (Han,

Lehmann *et al.* 2009, Wiedenheft, Zhou *et al.* 2009, Babu, Beloglazova *et al.* 2011). Undoubtedly, Cas1 is at the heart of the system that mediates bacterial immunity.

The production of crRNAs in CRISPR-mediated immunity was first observed while studying Cas6e, which specifically binds and cleaves within each repeat sequence of the long pre-crRNAs, producing crRNAs that contain a unique spacer sequence flanked by fragments of the adjacent repeats (Brouns, Jore *et al.* 2008). Although there was some clarity as to how Cas6 works in different species, the diversity of *cas* genes has made it difficult to investigate the protein factors that are involved in CRISPR RNA processing systems. For instance, none of the four *cas* genes identified in type II processing systems had similarities to an endoribonuclease. A different mechanism for processing CRISPR RNA has recently been reported. This mechanism relies on RNase III mediated cleavage of ds regions of RNA repeats. Further, Cas9 is required for RNA processing. A precise role for this enzyme has not been elucidated yet, which suggests that CRISPR RNA biogenesis is a complex process that requires several ancillary components (Deltcheva, Chylinski *et al.* 2011, Wiedenheft, Sternberg *et al.* 2012).

The third stage of CRISPR-mediated immunity involves interfering with the specific target. The crRNAs associate with Cas proteins to form a ribonucleoprotein complex that can recognize foreign nucleic acids. These acids are identified by specific base pairing interaction with the crRNA spacer sequence (Barrangou, Fremaux *et al.* 2007, Brouns, Jore *et al.* 2008). It has been demonstrated in *S. thermophilus* that the RNA guided ribonucleoprotein complex targets dsDNA and both strands are cleaved within the regions that are complementary to the spacer sequence in the crRNA (Garneau, Dupuis *et al.* 2010). The repeat sequence that flanks the CRISPR signals "self" and prevents the destruction of the host chromosome (Marraffini and Sontheimer 2010).

### 1.2.2. Regulation by RNA in eukaryotes through RNAi

The use of high throughput sequencing methods has revealed that the genomes of well-studied eukaryotes are almost entirely transcribed and generate a large amount of non-protein coding RNAs. There is increasing evidence of the functional role

assumed by some of these RNAs, but a lot of research still needs to be done (Amaral, Dinger *et al.* 2008). The ENCODE project, which aimed to identify all of the functional elements in the human genome sequence, found that ~ 93% of the nucleotides are transcribed in different types of cells (Birney 2007). This information suggests that the vast array of potentially functional ncRNAs could greatly surpass the 1.2% of the genome that encodes for proteins.

It has been more than a decade since Fire et al. (1998) described the effects of RNA injections into the well-characterized nematode *C. elegans*. RNAi is not only an ancient defense system that eukaryotic organisms use to protect their genomes, but is also a very powerful experimental tool. We now know that RNA silencing is a mechanism by which small RNAs regulate gene expression (Xie, Kasschau *et al.* 2003). Components of the RNA mediated pathway have been found in members of all eukaryotic super groups: Plantae, Excavates (Kinetoplastids), Rhizaria, Unikonts (Animals) and Chromoalveolates (Keeling, Burger *et al.* 2005). The widespread distribution of RNAi suggests that it was already present in the eukaryotes' last common ancestor. The silencing phenomena has been related to several processes, including post-transcriptional gene silencing, transcriptional gene silencing via heterochromatin formation or DNA methylation, DNA elimination, and translational arrestment (Cerutti and Casas-Mollano 2006).

Phylogenetic analyses suggest that the ancestral RNAi machinery present in the eukaryotes' last common ancestor had at least one Argonaute motif, a Piwi domain, one Dicer, and one RNA-dependent RNA polymerase (RdRP). The original RNAi function(s) was/were already present in this ancestor was/were probably a line of defense against genomic parasites like transposable elements and viruses. Functions that are now widely distributed among eukaryotes – like the post-transcriptional degradation of cognate RNAs, and the transcriptional repression of DNA sequences – are also likely to have been operative in the last common ancestor (Cerutti and Casas-Mollano 2006). Interestingly, even though RNA silencing is involved in a wide array of cellular processes, each process has a limited phylogenetic distribution. These facts question the evolutionary advantage of having the RNAi machinery over other species that lack the RNAi pathway.

There are two main executors in the RNAi pathway: RNaseIII (Dicer), which generates the active small RNAs, and the RNA-induced silencing complex (RISC) with its core enzyme Argonaute. Dicer belongs to the RNaseIII family, which has two RNaseIII motifs and an N-terminal helicase domain. Dicer produces two types of small RNAs: micro RNAs (miRNAs) and small interfering RNAs (siRNAs). These RNAs will be discussed in further detail in the next sections. Some of the core proteins that constitute the RISC belong to the Argonaute family and are highly conserved. They have two characteristic domains, PAZ and PIWI. Human Ago2 is essential for RNAi, it is the only protein that has been confirmed to have endonuclease activity (Sen and Blau 2006). The RISC complex might also contain other auxiliary proteins in addition to the Argonaute protein and the small RNA guide (Ghildiyal and Zamore 2009). There are a wide variety of small RNAs and RNA silencing proteins whose numbers vary greatly among organisms. For instance, *C.elegans* has 27 distinct Argonaute proteins compared with 5 in flies. In plants, *Arabidopsis thaliana* has 4 Dicer-like (DCL) proteins and 10 Argonauts that have both, unique and redundant functions.

In mammals and *C. elegans*, a single Dicer makes miRNA and siRNAs. In *Drosophila* there are two Dicers: DCR-1 that makes miRNAs while DCR-2 makes siRNAs (Hutvágner, McLachlan *et al.* 2001, Ketting, Fischer *et al.* 2001, Lee, Nakahara *et al.* 2004). The RNAi pathway defends against viral infection in *Drosophila*, so Dicer specialization might help to reduce the competition for Dicer between the pre-miRNA and the viral RNAs. Such specialization could also be a reflection of evolutionary pressure to counteract viral strategies in order to escape detection and neutralization. In fact, dcr-2 and ago2 are among the most rapidly evolving genes in *Drosophila (Obbard, Jiggins et al. 2006)*. This is probably due to the selective pressure that comes from having to protect from viruses that are constantly mutating and trying to escape detection. It seems like *C. elegans* does not fight natural viral infection via RNAi. It has been observed that mammals respond to viral attacks using a protein- (antibody-) based immune system, which might help explain the presence of only one Dicer in these systems (Shaham 2006, Vilcek 2006).

A large number of small RNA classes have been identified since the discovery of RNA interference and these vary in their biogenesis, mode of target regulation and the

biological pathways they control. Evidence suggests that the activities of the various small RNAs are interconnected and that the RNAs collaborate to regulate pathways and protect genomes from foreign and internal threats. The defining characteristics of small silencing RNAs are their length (20-30 nt long) and their association with Argonaute proteins, which guides them to their regulatory targets. Small RNA classes are also diverse and lead complex schemes of regulation. The following sections provide an overview of small silencing RNAs and provide examples of the variety and complexity of the processes they regulate (Ghildiyal and Zamore 2009).

### *Micro RNA (miRNA)*

Micro RNAs are transcribed from non-protein coding genes and their precursors form a hairpin–like structure. In *C. elegans,* miRNAs form imprecise base pairing with the target mRNA and arrest translation. However, miRNAs have a different role in plants in that they target mRNA with close to perfect complementarity for degradation. Some miRNAs in multicellular organisms like plants and animals have temporal and tissue-specific patterns of expression that are consistent with control of development (Xie, Kasschau *et al.* 2003). Micro RNAs binding to 3' UTR in mRNA are considered a typical miRNA mode of regulation in animals (Ghildiyal and Zamore 2009).

Micro RNAs (miRNA) derive from precursors called primary miRNAs that are usually transcribed by RNA polymerase II (Lee, Jeon *et al.* 2002). Next, two RNase III endonucleases (assisted by their dSRBD) will process the pri-miRNA to produce mature 20-24 nt long mature miRNAs. Further, the pri-miRNA is processed into a 60-70 nt long pre-miRNA by Drosha action in the nucleus of mammal cells (Figure 1.8). The resulting structure has a hairpin structure flanked by base-paired arms that form a stem. The arms are carried to the cytoplasm by Exportin 5 and Dicer completes the processing of the miRNAs (Lee, Jeon *et al.* 2002, Lee 2003, Yi, Qin *et al.* 2003, Jiang 2005).

The mechanisms by which a miRNA regulates its mRNA target reflects the specific nature of the Argonaute protein into which the RNA is loaded, as well as the extent of complementarity between the miRNA and the mRNA (Ghildiyal and Zamore 2009). Extensive complementarity between the mRNA and the miRNA is completely required in plants (Rhoades 2002, Yekta, Shih *et al.* 2004, Brennecke, Stark *et al.* 2005).

Only a few miRNA in flies and mammals are likewise completely complementary to their target mRNAs (Yekta, Shih *et al.* 2004, Brennecke, Stark *et al.* 2005). In flies and mammals most miRNAs pair with their target RNAs in a limited stretch of sequence called a 'seed region' in order to repress translation and trigger RNA degradation. The advantage of a small seed region is that the cell can regulate several genes with a single miRNA, if not hundreds of different genes (Lewis, Shih *et al.* 2003, Brennecke, Stark *et al.* 2005, Baek 2008).

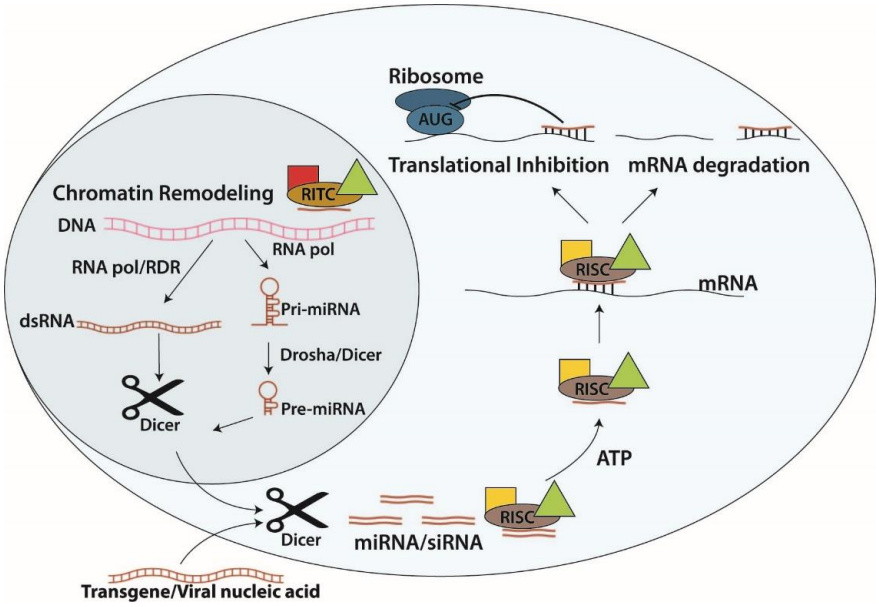



***Figure 1.8. Major RNA silencing pathways***

The three pathways are the small interfering RNA (siRNA), the microRNA (miRNA), and the Piwi-interacting RNA (piRNA). The pathways differ in terms of their substrates, biogenesis, effector proteins and modes of target regulation. **a**) On the small interfering RNA (siRNA) pathway, dsRNA precursors are processed by Dicer-2 (DCR-2) to generate siRNA duplexes with guide and passenger strands. DCR-2 and the dsRNA-binding protein R2D2 form the RISC-loading complex. **b**) On the microRNA (miRNA) pathway, the miRNAs which are encoded in the genome are transcribed to produce a primary miRNA (pri-miRNA) that is cleaved by Drosha/Dicer to yield a short precursor miRNA (pre-miRNA). miRNAs can also be present in introns (termed mirtrons) that are liberated following splicing to yield authentic pre-miRNAs. pre-miRNAs are exported from the nucleus to the cytoplasm and processed by DCR-1 that produces a duplex containing two strands: miRNA and miRNA*. The miRNA strand is loaded into AGO1 and guides the translational repression of target RNAs.

### *Small interfering RNA (siRNA)*

Small interfering RNAs are produced by the Dicer cleavage of long double-stranded (ds) RNAs. Mature siRNAs are loaded into RISC in an ATP-dependant

process and guide it to destroy mRNA in a sequence-specific manner (Novina and Sharp 2004). The mechanism described above is known as post-transcriptional gene silencing and is one of the better understood (Figure 1.8). In most Unikonts this mechanism silences transposable elements, repetitive genes, and viruses. siRNAs also guide nuclear events like histone and DNA methylation, which result in transcriptional silencing (Xie, Kasschau *et al.* 2003). In organisms such as worms and plants, this mechanism requires the activity of an RNA-dependent RNA polymerase (RdRP) that generates dsRNA from single stranded transcripts and results in an amplification of the RNAi response. In plants, the amplification of the RNAi response, contributes to spread gene silencing by cell-to-cell transfer of dsRNA. The outcomes of this include generating a widespread resistance to viral infection (Novina and Sharp 2004). Next, we classify siRNAs according to the molecules that trigger their production (Ghildiyal and Zamore 2009).

**siRNAs derived from exogenous agents**

Long exogenous dsRNA is cut into ds siRNA by Dicer, a dsRNA-specific RNase III family ribonuclease. The siRNAs produced are usually 21 nt long, with each strand bearing a 5' phosphate and a 3' hydroxyl group, and with a 2 nt overhang in the 3' end. The strand that directs the silencing is called a guide strand and the second strand is the passenger strand, which is ultimately destroyed (Zamore, Tuschl *et al.* 2000). The thermodynamic stability of each strand will determine the identity of the guide and passenger strands. In flies this difference is sensed by R2D2, a partner of AGO-2 which also forms part of the RISC loading complex. RLC recruits AGO2 and transfers it to the siRNA duplex. There, the passenger strand gets cleaved as if it was the target RNA at the phosphodiester bond between nt 10 and 11 (which are paired to the guide strand) (Elbashir, Lendeckel *et al.* 2001, Liu, Rand *et al.* 2003, Matranga, Tomari *et al.* 2005). The release of the passenger strand turns the pre-RISC complex into a mature RISC complex that only has the ss RNA guide strand 2'-O-methylated at the 3'end by the S-adenosyl methionine dependent methyltransferase HEN1 (Horwich, Li *et al.* 2007, Pélisson, Sarot *et al.* 2007). In plants, both miRNA and siRNAs are methylated at both ends, which is essential for the miRNA's and siRNAs's stability (Li, Yang *et al.* 2005, Yu, Yang *et al.* 2005) and is modulated by a range of viral proteins during infection (Ebhardt, Thi *et al.* 2005, Ebhardt, Ovando *et al.* 2012).

**siRNAs derived from endogenous agents (endo-siRNA)**

Endo-siRNA are ubiquitous among higher eukaryotes. They have been detected in plants, *C. elegans*, flies and mammals (Ghildiyal and Zamore 2009). Plant endo-siRNAs are called *cis*-acting siRNAs (casiRNAs), originate from transposons, repetitive elements and tandem repeats like rRNA genes, and comprise the majority of endo-siRNA (Xie 2004). These RNAs are 24 nt long, are methylated by HEN1, and are produced with the intervention of DCL3, RNA polymerases RDR2 and POL IV, and AGO6 or AGO4. casiRNAs promote DNA methylation and histone modification at the loci where they originated by promoting heterochromatin formation (Zilberman, Cao *et al.* 2003, Chan, Zilberman *et al.* 2004, Xie 2004, Herr, Jensen *et al.* 2005). This is just one example that highlights the diverse number of pathways used by different organisms to produce small regulatory RNA.

### *Piwi-interacting RNAs (piRNA)*

The piRNAs function mainly in the germ line and as their name suggests, they bind to the Piwi clade of Argonaute proteins. They are also longer than miRNAs and siRNAs on average (25-30 nt) (Aravin 2001). piRNAS were initially proposed to ensure germline stability by repressing transposable elements. In addition to binding Piwi proteins, piRNAs do not require DCR-1 or DCR-2 for their production (unlike miRNAs and siRNAs) (Vagin 2006, Pélisson, Sarot *et al.* 2007, Ghildiyal and Zamore 2009). piRNA sequences are surprisingly diverse. Over 1.5 million of sequences have been identified in Drosophila, and they map to only a few hundred genomic clusters. The best studied cluster is *flamenco*, which represses *gypsy* and other related transposons. Unlike siRNAs, piRNAs are mainly antisense, which suggests that they arise from a long single-stranded precursor (Saito 2006, Brennecke 2007, Nishida 2007, Czech 2008). As previously mentioned, piRNAS are mainly involved in silencing of transposons. In mammals, their activity is most important during embryo development (Aravin 2008). piRNAs exert their role via the formation of the RNA-induced silencing complex (RISC). The piRNAs interacting with the Piwi proteins are also required for germ cell and stem cell development in invertebrates (Aravin 2008). Further, it is thought that piRNA and endogenous small interfering RNA (endo-siRNA) may have comparable and even

redundant functionalities in the transposon control of mammalian oocytes (Malone and Hannon).

### *tRNA fragments (tRFs)*

Although tRNAs are best know for their canonical role as “adaptors” of the genetic code during translation, aminoacylated tRNAs have also been identified as substrates for non-ribosomal peptide bond formation, post-translational protein labeling, the modification of phospholipids in the cell membrane, and antibiotic biosynthesis (Raina and Ibba 2014). Recent studies have found that the cleavage products of tRNAs generate short RNA with microRNA-like features. The biogenesis pathway and function of tRNAs is not yet very clear.

A report looking into small RNAs in HIV-infected cells found small 20 nt long tRNA derived fragments that serve as primers for reverse transcription. The prevalence of the small RNA correlated with the HIV expression level and was bound to Ago 2 (Yeung, Bennasser *et al.* 2009). In 2009, another study described the deep sequencing of total small RNAs from two prostatic carcinoma cell lines. This study found 17 RNAs (18-22 nt long), that were not already entered in the Sanger database. These 18-22 nt RNAs aligned with tRNA sequences derived from either the 5' or the 3' ends of mature tRNAs or the 3' end region of the pre-tRNA. To further examine the potential function of tRNAs, one RNA tRF-10001 that corresponded to the 3' trailer of pre-tRNA $^{SerTGA}$ was selected. This RNA showed elevated expression in several human cell lines and its expression was correlated to cellular proliferation rate (Lee, Shibata *et al.* 2009). Although there is lot of interesting data suggesting that these tRNA fragments might have a role similar to that of miRNA during RNA silencing, it is important to address several questions; Is it possible that these tRNA fragments are a product of tRNA turnover? Are these bonafide microRNAs? What are the mRNA targets for these tRNA fragments? (Pederson 2010).

tRFs not only play a part in the regulation of genes, but have also been recently linked to other biological functions (Raina and Ibba 2014). A 2014 study described tRF involvement in viral infectivity when analyzing the data from large scale sequencing of $CD^{+4}$ T cells as well as the data from these same cells after they were infected with T-

cell leukemia virus type I (HTLV-1). The most abundant tRF found was derived from the 3' end of tRNAPro and had perfect complementarity to the primer binding site for HTLV-1. It has been proposed that one of the tRF fragment's functions is to support the initiation of the reverse transcription (Ruggero, Guffanti *et al.* 2014). Transfer RNA fragments (tRFs) have also been linked to progressive neuron motor loss in CLP-1 mutant cells (Hanada, Weitzer *et al.* 2013). The tRF generated in the CLP1 mutant cells sensitize cells to the oxidative stress induction of p53, which leads to the progressive loss of spinal motor neurons that cause muscle denervation and paralysis (Hanada, Weitzer *et al.* 2013). The study presents a possible link between tRFs and p53 dependent cell death. Most of these reports are largely based on data generated from high-throughput studies that have high sensitivity and the ability to detect unique molecules of RNAs in a cell. However, there is lack of studies describing the mRNA targets for tRFs, which would conclusively prove that tRFs have a biological role.

### *RNAi in Kinetoplastida*

Part of the research described in Chapter 4 will focus on the Trypanosomatidae family that contains well-known parasites. Understanding RNA interference in this group will provide useful background information for interpreting the discovery of RNA in leishmania-derived exosomes. The complex trypanosomatid life cycle will also be discussed in further detail later on this chapter. The trypanosomatides go through several developmental stages in order to adapt to the distinct environments they encounter during their life cycle, they need to survive inside cells in humans and as extracellular parasites in the vector (Barrett, Burchmore *et al.* 2003). Such adaptations require changes in morphology, surface composition, biochemical pathways, and are determined by differential gene expression.

The control of gene expression occurs mostly at a post-transcriptional level and uses several mechanisms, some of which are unique to trypanosomatides. RNAi seems like an obvious candidate for controlling gene expression in this group. The kinetoplastid *T. brucei* is the causal agent of the African sleeping disease, and is RNAi positive. The introduction of transient gene-specific dsRNA from α-tubulin mRNA 5' UTR led to the production of cells with morphological alterations that cannot undergo cytokinesis. Rapid and specific tubulin mRNA degradation was observed when dsRNA was

transfected, whereas individual RNA strands did not have the same effect (Ngo, Tschudi *et al.* 1998). Further research has shown that endogenous siRNAs are produced to silence retroposon-like elements in *Trypanosoma brucei's* genome and contribute to chromatin remodeling while in the nucleus. Both activities are consistent with those observed in plants and animals (Djikeng, Shi *et al.* 2001, Ullu, Tschudi *et al.* 2004). The actual RNAi machinery was later identified in *T. brucei*. An Argonaute-like protein (TbAGO1) was isolated from a ribonucleoprotein particle associated with siRNAs and knockout studies proved it to be essential in the RNAi pathway (Shi, Djikeng *et al.* 2004). A DICER-like protein (TbDcl1) in *T. brucei* was found by BLAST searches using the *Aquifex aeolicus* RNase III domain. TbDcl1 is highly divergent from the Dicer enzymes of animals, plants, fungi, and even protists. *T. brucei* is the only conclusive and well-studied example of RNA silencing in kinetoplastids.

### 1.2.3. Parasitic RNA sequences can hijack cellular machinery and re-program it to carry out additional functions

For the purposes of this dissertation, parasitic RNAs are understood as sequences of variable length that are capable of moving from one cell to another (within the same organism or between organisms) and exerting an effect in the host cell that triggers changes favourable to the survival of the foreign parasitic RNA molecule. The RNA sequences described below can move from one part of the genome to another, from one cell to another inside the same organism, or from one organism to other organisms, sometimes crossing interspecific barriers. The focus of the next sessions is on the RNA particles that contribute to disease or stress in the host cell/organism.

#### *RNA viruses*

RNA viruses are particles that carry ribonucleic acid as their genetic material. Human diseases such SARS, polio, influenza, measles are caused by RNA viruses. The Baltimore classification system of classifying viruses does not consider RNA viruses that have DNA at any point during their cycle to actually be RNA viruses (these are called retroviruses). The genetic material in RNA viruses can be single stranded (ss) or double stranded (ds) (Adams, Lefkowitz *et al.* 2013). Single-stranded viruses can have a positive (+) or negative (-) polarity, which means that the RNA has to be converted to a

positive RNA strand before the RNA polymerase can read it and transcribe it. Interestingly, purified negative (-) sense RNA strands are not infective by themselves (Bullido 2000), which highlights the fact that the (-) strand of RNA is purely an information carrier in the case of viruses. The dsRNA viruses are a diverse group with a wide variety of hosts (humans, fungi, bacteria) as well as virion organization.

One of the most commonly known RNA viruses is the Rotavirus. This is the leading causing agent of gastroenteritis in infants, is one of the renowned members of the dsRNA virus group (Patton 2008). The rotavirion has a nonenveloped, complex triple-layered capsid that surrounds a genome formed of 11 double-stranded RNA segments. There are six structural proteins and six non-structural proteins, each encoded in a unique genome segment with the exceptions of non-structural proteins 5 and 6 (NSP5 and NSP6), which are encoded in overlapping reading frames of a single segment. The model of diarrhoea induction by rotavirus starts when the virus binds and invades the enterocytes in the small intestine. The virus is internalized and loosens the outer capsid while the virion-associated transcriptase and viral macromolecular syntheses get activated. Cellular events triggered by NSP4 cause the release of $Ca^{+2}$ from the endoplasmic reticulum of host cells, which triggers the disruption of the microvillar cytoskeletal network and lowers the expression of disaccharidases and other enzymes from the apical surface of the small intestine, disrupting the $Na^{+}$ transport system and necrosis. These events lead to malabsorption in the epithelium, which is one of the components of diarrhoea. The secretory component of rotavirus diarrhoea appears to be a consequence of the changes induced in the villus epithelium governed by NSP4 and ENS (enteric nervous system). Although the precise role and targets of NSP4 are unknown, it seems to act on the crypt epithelium, increasing the $[Ca^{+2}]$ in the crypt cells and promoting the $Cl^{-}$ secretion, which causes an outflow of water. The ENS –which is also a target in cholera toxin induced diarrhoea– (Lundgren, Peregrin *et al.* 2000) is another target for NSP4. The ENS is also strongly linked to the increased intestinal motility but the molecular stimulator of motility is unknown (Ramig 2004). Although there are lot of unanswered questions regarding the molecular mechanisms that direct the pathogenesis of rothavirus, current evidence indicates that in the case of RNA viruses (and viruses in general), the pathogenic factors are protein based and the

RNA acts solely as an information carrier. This is a main difference with between RNA viruses and viroids, which are discussed below.

### *Plant viroids*

Viroids are small structures, RNA-based genomic material that ranges from 200 to 400 nt long – approximately 10 times smaller than the smallest genome of a regular virus. Viroids have a high degree of self-complementarity within their sequence, which is one of the reasons why there are highly compact particles. Viruses and viroids are considered parasites of the host cell's transcriptional machinery. In contrast to viruses, viroids do not seem to encode for proteins and consequently rely completely on the host's protein factors to complete their infectious cycle (Diener 1971, Diener 2001). This essentially means that viruses and viroids have developed different types of parasitism throughout their evolution. Interestingly, some viroids contain catalytic RNAs as they code for hammerhead ribozymes that mediate the self-cleavage of the multimeric RNA that accumulates during rolling cycle replication. The fact that no other virus has been classified as a catalytic RNA supports the idea that viroids evolved independently from other viruses. This idea was put forth some time ago in RNA world theory, which proposed that RNA preceded DNA and proteins (Diener 2001).

Most of the 30 viroids described so far belong to the *Pospiviroidae* family, which adopts a rod-like secondary structure with minimal free energy. The potato spindle viroid (PSTVd) was the first viroid to be characterized in the 70's (Diener 1971, Flores, Hernández *et al.* 2005). Viroids cause disease in a variety of economically important crops like potatoes, tomatoes, cucumbers, coconuts, and fruits grown in several subtropical or temperate climates. Only the *Tomato planta macho viroid* (TPMVd) is known to be transmitted by an aphid under specific ecological conditions. The rest of the viroids are transmitted mechanically in pollen grains, seeds and even more efficiently in the vegetative propagation of infected material (Flores, Hernández *et al.* 2005). Some viroids affect specific organs such leaves (chlorosis, epinasty, rugosity), fruits (discoloration, skin deformation), stems (dwarfing, internode shortening), flowers (size reduction or broken lines on petals) or seeds, while others have more general effects. Once the viroid has colonized the first cell, it needs to invade adjacent cells in order to

invade more distant parts of the plant and does so via the plasmodesmata, just as a virus does. Long distance spread occurs using the plant's vascular system.

The identity of the pathogenic effectors in viroids still remains unclear. Given that they lack the capacity to code protein, viroids must trigger disease by the direct interaction of their genomic RNA with host factors in a sequence of events that are still poorly understood. Specific 3-4 nt mutations in PSTVd have been mapped to a virulence modulating (VM) region that overlaps with a (PM) region in the rod-like structure and been found to have a marked effect on disease symptoms (Schnolzer, Haas *et al.* 1985). Similar evidence was found on *Citrus exocortis* viroid (CEVd) (Visvader and Symons 1986). Considering this evidence along with the findings from PSTVd suggests that variations in the geometry of the VM region lead to alterations in RNA-protein interactions that are the main cause of viroid pathogenicity. It is also hypothesized that initial interactions with the host proteins involved in replication, viroid movement or accumulation could be the initial event in pathogenesis (Sano, Candresse *et al.* 1992). Another line of evidence proposes that instead of host proteins being the main target, the molecular event triggering pathogenesis might be the base-pair interactions between PSTVd and host RNAs that result in the interference of rRNA maturation, mRNA maturation or the 7S assembly (into the signal recognition particle) (Diener 2001). This hypothesis fails to explain the differential symptoms observed in closely related host species. It is generally accepted that viroids trigger disease by directly interfering with the host gene expression machinery, although many questions regarding the underlying mechanism remain unanswered. Viroids are a great example of the regulatory capacity that RNAs can have on their own, even when the viroids interfere with the normal development of complex organisms such as plants.

### *Retrotransposons*

Retrotransposons are mobile DNA elements that use retro-viral-like reverse transcription strategies during the process of transposition. The retrotransposon DNA gets transcribed into an RNA template and is then reverse transcribed into DNA and inserted into a new site in the genome. These sequences have proliferated specially in animal and plant genomes while transposons proliferated in bacteria (Boeke 2003). Retrotransposons are divided into two main classes: viral and non-viral. Viral

retrotransposons are abundant in yeast and Drosophila, while non-viral retrotransposons are most abundant in mammals.

Retroelements are regulated by epigenetic controls, which generate multiple miRNAs that are involved in the induction and progression of genomic instability. Further research into the biological roles of retroelements will advance a better understanding of retroelements' evolutionary features and implications in disease. The movement of these elements around the genome causes insertional mutagenesis and changes in the genome structure as well as changes in the expression of neighbouring genes. Retroelements interfere with gene expression by inducing alternative splicing via exon skipping and exonization using cryptic splice sites, and by providing alternative polyadenylation signals. Such events have been connected to many diseases in humans, including Duchene Muscular dystrophy (disruption of the DMD gene), Colon Cancer (disruption of APC gene), Haemophilia A and B (disruption of F8 and F9 gene respectively) (Jung, Ahn *et al.* 2013, Kaer and Speek 2013).

## 1.3. Intercellular communication, and transfer of proteins and RNA between cells

Cells communicate with each in many different ways, depending on the complexity of the organism. Cell signalling can be seen in unicellular as well as multicellular organisms. Hormone secretion is the system used by animals and plants for long distance communication. Membrane vesicle trafficking is observed for intercellular communication for intra or inter species. It is particularly interesting to us when it involves inter kingdom cell signalling. Exosomes are 40-100 nm membrane vesicles traditionally associated with secretion in mammalian cells (Simpson, Jensen *et al.* 2008). The secretion of membrane-enclosed vesicles from tumour cells and platelets was described more than 40 years ago. Exosomes are different from membrane microvesicles in that exosomes are produced by blebbing and their release occurs through the fusion of multivesicular bodies from the endocytic/exocytic pathway and the plasma membrane of the cell (Kowal, Tkach *et al.* 2014). Early studies of extracellular vesicles suggested that they were carriers of cellular debris. However, in the last decade or so a large amount of research has indicated that exosomes are much more

than mere garbage carriers for the cell. Interestingly, proteomic studies of exosomal content reveals that exosomes contain a conserved set of proteins across species, which suggests that proteins like tubulin, actin or Hsp 70 and Hsp 90 have important and even life-saving roles (Simpson, Jensen *et al.* 2008).

Exosomes have been shown to carry several surface determinants of tumour cells that interact with monocytes to alter their immunotyphe and biological activity (Baj-Krzyworzeka, Szatanek *et al.* 2006). Exosomes have also been found *in vivo* in body fluids such as blood, urine, synovial fluids, and breast milk. Although the functions and mechanisms of exosomes are not completely clear, they promote intercellular communication by facilitating antigen presentation and trans-signalling to neighbouring cells. Exosome secretion has also been implicated in the transport and propagation of infectious cargo such as prions and retroviruses such as HIV, which suggests they may also be related to pathological situations (Simpson, Jensen *et al.* 2008). It has been reported that exosomes containing pathogen-derived factors have been be released by infected cells after being infected with the Epstein-Barr virus (EBV), *M. tuberculosis*, or toxoplasma (Giri, Kruh *et al.* 2010, Pegtel, Cosmopoulos *et al.* 2010, Mantel, Hoang *et al.* 2013) . New research provides further insight into diseases and offer new treatment options (Giri, Kruh *et al.* 2010). There are also several examples of pathogens releasing exosomes by themself. Two species of medically relevant *Mycobacterium* release vesicles were found to deliver immunologically active molecules that contribute to mycobacterial virulence (Prados-Rosales, Baena *et al.* 2011). The pathogenic fungi, *Cryptococcus neoformans* also produces exosomes, which provide a means of circumventing the fungi's thick cell wall and delivering virulence factors to host cells (Panepinto, Komperda *et al.* 2009).

Studies from Dr. Reiner's laboratory identified the first protozoan, *Leishmania donovani,* to release exosomes. A variety of proteins were described in the Leishmania exosomes (Silverman, Chan *et al.* 2008). The alternative exosome-based secretion mechanism is found to be responsible for exporting ~52% of the proteins present into the Leishmania culture medium (Silverman, Chan *et al.* 2008). Silverman also identified a variety of proteins with proteolytic activity that might participate in pathogenesis. The proteins degrade enzymes in the phagolysosomes, the major histocompatibility complex

I and II. Both actions prevent antigen loading and reduce the efficiency of antigen presentation. The two effects have been described previously in macrophages prior to parasite invasion. Interestingly, proteins related to the translational machinery were also isolated from the vesicles (Silverman, Chan *et al.* 2008, Silverman, Clos *et al.* 2010, Silverman, Clos *et al.* 2010). Finally, the proteins found in the exosome were also detected in the cytoplasmic compartment of infected macrophages, where they induced the production of IL-8, but did not induce the production of TNF-σ. These findings highlight exosomes' capabilities for long range intercellular communication as well as inmuno modulation (Silverman, Clos *et al.* 2010).

### 1.3.1. Exosomes as vehicles for RNA

Another interesting layer of complexity emerged when exosomes were observed to contain RNA which can be transferred to other cells and even be functional in the new environment.

A 2007 study described mouse exosomes' capacity to shuttle mRNA and miRNA from one cell to another. The exosomal content of mouse mast cells was transferred into human mast cells and resulted in the *de novo* synthesis of mouse proteins in the recipient cells {Pegtel, 2010 #154}. This new mechanism adds complexity to the way cells communicate, and the RNA exchange could potentially have regulatory effects (Valadi, Ekstrom *et al.* 2007). A more recent study found that EBV infected cells secreted exosomes which contained miRNAs. It was also verified that the miRNAs are functional when transferred to uninfected cells in a dose-dependent manner, and mediate the repression of confirmed EBV target genes. These results are consistent with miRNA-mediated gene silencing as a mechanism of intercellular communication (Pegtel, Cosmopoulos *et al.* 2010).

Non-coding regulatory RNA was detected in exosomes. The RNA showed to have either cellular or viral origin, and it is capable of controlling gene expressions as well as repressing the translation of mRNA. The exosomes might protect labile RNAs from degradation by RNases and increase their chances of being functional outside the

cell that generated them (Simpson, Jensen *et al.* 2008, Pegtel, Cosmopoulos *et al.* 2010).

### 1.3.2. *Leishmania*, exosomes and the RNA connection

*Leishmania sp.* are a group of single-celled eukaryotic parasites that belong to the Protozoa. Protozoans usually share characteristics associated with animals (such as mobility and heterotrophy). Traditionally, Protozoans have been divided on the basis of their means of locomotion. *Leishmania sp.,* belongs to the Flagelates group (Marquardt, Demaree *et al.* 1999). *Leishmania sp.* include species that cause some of the most neglected human diseases in the tropics and subtropics (Barrett, Burchmore *et al.* 2003). *Leishmania sp.* vary in type and can take a visceral, cutaneous or mucosal form, each causing different symptoms of infection. These infections are caused by about 21 different species that are transmitted by about 30 different species of phlebotomine sandflies. If the disease is clinically evident but is left untreated, visceral *Leishmaniasis* causes a systemic infection that can be life-threatening; cutaneous Leishmaniasis can cause chronic skin sores, and; mucocutaneous Leishmaniasis causes a metastatic complication of new world cutaneous Leishmaniasis that produces facial disfigurement (Herwaldt 1999). Chapter 4 of this dissertation focuses on two species in particular: *Leishmania donovani,* which causes visceral infection in vertebrates, and *Leishmania braziliensis,* which causes a mucosal form of the disease (Lambertz, Oviedo Ovando *et al.* 2015).

*Leishmania sp.* have a complex life cycle that involves at least two hosts. Female sand flies seek a blood meal at or after dusk and become infected if they suck the blood of infected vertebrates. Amastigotes transform in the sandfly gut and replicate as promastigotes, which are regurgitated and inoculated in the skin of mammals at a subsequent blood meal. The flagellated promastigotes are the infective form of the parasite. Promastigotes invade or are phagocytosed by macrophages into the phagolysosomes. In there, the parasite transforms and replicates as amastigotes that will infect additional macrophages either locally or in distant tissues. Promastigotes enter macrophages silently to evade triggering host responses in order to initially establish infection (Murray, Berman *et al.* 2005). The work initiated by Silverman

explores exosomes' capacity to induce changes in the macrophages in order to create a more hospitable environment for the *Leishmania sp.* by the time they invade it (Silverman, Chan *et al.* 2008, Silverman, Clos *et al.* 2010, Silverman, Clos *et al.* 2010). The last chapter of this thesis explores whether RNA exists in exosomes that could be delivered to the host cells and contribute to pathogenesis.

## 1.4. Research Objectives

The overall goal of my research was to better understand the role of RNAs in helping pathogens to survive under stressful conditions such as a lack of nutrients or even the invasion of a new host. In more detail, I want to observe how 6S RNA contributes to overall eubacterial transcriptional control and cell survival and provide a better understanding of one of the regulatory pathways that enable bacteria to survive so succesfully under extreme or unfavourable conditions (Chapter 2). I also want to explore the importance of sequence or structural 6S RNA features that determine binding, release and pRNA production capabilities while interacting with the holoenzyme (Chapter 2). The possibility of additional roles for 6S RNA during lag and exponential phases of growth as well as the potential pathways by which it might be degraded will be explored in Chapter 3. Finally I will shift the focus from bacteria to eukaryotes to explore the possibility that the protein containing exosome–like vesicles secreted by *Leishmania sp.* may also be transporting RNA and could be delivered to the host cells and contribute to pathogenesis, just like their protein counterparts (Chapter 4).

In chapter 1, I presented background information that highlights how ubiquitous RNA is and the functions it can carry out. Our current understanding of the flow of information in nature includes several functions for RNA. It is an essential intermediary in the flow of information from DNA → RNA → Protein, and also a carrier of information in RNA virus and viroids. Different types of RNA have regulatory roles in processes like DNA replication, transcription and translation.

The process of transcription and its regulation are presented in detail as they are closely connected to 6S RNA, which is a molecule that can efficiently regulate overall transcriptional rates. During nutritional stress, the 6S RNA binds to the holoenzyme and

when nutrient concentration increases, it releases from the holoenzyme in a manner similar to that of DNA. The 6S RNA is one of the regulatory molecules I am interested in, and will be focusing on in the next two chapters.

Chapter 2 focuses aims to understand what are the 6S RNA primary and secondary structure features that influence the rate of *in vitro* release from the holoenzyme, as well as the consequences of expressing *in vivo* release-defective 6S RNA variants. The mutant 6S RNA sequences studied, were the product of an *in vitro* selection procedure that also helped to narrow down the important residues in the 6S RNA sequence and structure. *In vivo* studies accentuated the importance of 6S RNA activity in bacteria when the expression of release defective 6S RNA stalled cellular growth in liquid media.

Chapter 2.

# *In vitro* selected 6S RNA release mutants reveal three steps required for regulating *E. coli* RNA polymerase

Oviedo Ovando, M., L. Shephard, and P. Unrau.

A modified version of this chapter was published in the RNA journal, 2014 20(5): 670-680

The binding and release selection, was performed by Lindsay Shephard. The remaining experiments were designed and carried out by Mariana Oviedo-Ovando.

## 2.1. Abstract

6S RNA is a non-coding RNA that inhibits bacterial transcription by sequestering RNA polymerase holoenzyme ($E\sigma^{70}$) in low nutrient conditions. This transcriptional block can be relieved by the synthesis of a short product RNA (pRNA) using the 6S RNA as a template. Here, we selected a range of 6S RNA release-defective mutants from a high diversity *in vitro* pool. Studying the release defective variant R9-33 uncovered complex interactions between three regions of the 6S RNA. As expected, mutating the transcriptional start site (TSS) slowed and partially inhibited release. Surprisingly, additional mutations near the TSS were found that rescued this effect. Likewise, three mutations in the top strand of the large open bubble (LOB) could considerably slow release, but were rescued by the addition of upstream mutations found between a highly conserved '-35' motif and the LOB. Combining the three top strand LOB mutations with mutations near the TSS however was particularly effective at preventing release and this effect could be further enhanced by inclusion of the upstream mutations. Overexpressing R9-33 and a series of milder release-defective mutants in *E. coli* resulted in a delayed entry into exponential phase together with a decrease in cell survival that correlated well with the severity of the *in vitro* phenotypes. This suggests

that 6S RNA release rate plays an important *in vivo* role. The complex crosstalk observed between distinct regions of the 6S RNA supports a scrunching type model of 6S RNA release, where at least three regions of the 6S RNA must interact with $E\sigma^{70}$ in a cooperative manner so as to ensure effective pRNA dependent release.

## 2.2. Introduction

*E. coli* 6S RNA binds to $\sigma^{70}$ RNA polymerase holoenzyme ($E\sigma^{70}$) in low nutrient conditions and globally suppresses transcription by reducing the amount of free cellular $E\sigma^{70}$ (Wassarman and Storz 2000, Cavanagh, Klocko *et al.* 2008, Neusser, Polen *et al.* 2010). The 6S RNA is released by the synthesis of a short product RNA (pRNA) when optimal nutrient conditions are restored allowing the bacteria to rapidly resume exponential growth (Wassarman and Saecker 2006, Wassarman 2007, Cavanagh, Sperger *et al.* 2012). The transcription of 6S RNA is controlled by $\sigma^{70}$ (housekeeping gene expression) and $\sigma^{38}$ (stationary phase gene expression) dependent promoters. Since 6S RNA release from $E\sigma^{70}$ depends on nutrient availability, the overall level of free core enzyme (E) is expected to be highly dynamic and largely determined by the turnover rate of the 6S RNA (Hsu, Zagorski *et al.* 1985, Wassarman and Storz 2000, Kim and Lee 2004, Shephard, Dobson *et al.* 2010). This form of RNA dependent transcriptional regulation, unlike protein mediated regulation (Hammerle, Beich-Frandsen *et al.* 2012), is intrinsically rapid and is one of several strategies that bacteria use to promptly respond to changes in their environment (Wassarman 2007). The focus of our investigation is to determine how the 6S RNA sequence contributes to the transcriptional dynamics of RNA polymerase.

A published model suggests that the process of 6S RNA release is analogous to the process of DNA-dependent transcriptional initiation (Panchapakesan and Unrau 2012). During transcriptional initiation, DNA 'scrunching' (Cheetham and Steitz 1999) steadily packs downstream DNA into the $E\sigma^{70}$ complex via the ratcheting effect of NTP polymerization (Ederth, Artsimovitch *et al.* 2002). The act of scrunching DNA into the polymerase serves to trigger a rearrangement of the polymerase from an initiation complex to an elongation complex capable of processive transcription (Murakami and Darst 2003). Based on *in vitro* data, 6S RNA release in the $\gamma$-proteobacteria appears to

be similar in mechanism to DNA-dependent transcriptional initiation, but with pRNA synthesis driving a series of steps that ultimately result in the ejection of the 6S:pRNA complex from E (Figure 2.1). After a 9 nt long pRNA has been synthesized, a release hairpin rapidly forms (Panchapakesan and Unrau 2012). The two arms of the release hairpin are built from highly conserved nucleotides found in the top strand of the large open bubble (LOB) and the top strand residues from the conserved downstream region of the 6S RNA (Figure 2.1, state S3). The release hairpin helps to destabilize the ribonucleoprotein (RNP) complex resulting ultimately in the ejection of the 6S:pRNA after the pRNA is extended by a further 4 nt (Figure 2.1, state S4). Consistent with this model, swapping the downstream top and bottom strands to prevent the formation of the release hairpin, slows the 6S RNA release rate and RNA release is only achieved after the synthesis of a much longer pRNA *in vitro* (Panchapakesan and Unrau 2012). Interestingly, other bacteria like *Bacillus subtilis* seem to change their secondary structure in the bottom strand of the LOB during release (Panchapakesan and Unrau 2012, Steuten and Wagner 2012) implying that just as in DNA dependent scrunching, the mechanism of 6S RNA release is intrinsically flexible and presumably highly dependent on 6S RNA sequence.

To further explore the process of 6S RNA release, we selected a range of *E. coli* release-defective 6S RNA variants using an unbiased *in vitro* selection approach (Shephard, Dobson *et al.* 2010). The types of mutants isolated suggest the existence of several discrete stages in the release process that are consistent with the proposed scrunching model of 6S RNA release (Panchapakesan and Unrau 2012). In the present chapter, we focus on the R9-33 isolate due to its ability to remain bound to the $E\sigma^{70}$ in conditions that normally induce very rapid ($t_{1/2}$ ~30 sec) release of a truncated 6S RNA control construct called T1 (Shephard, Dobson *et al.* 2010). We uncovered in the R9-33 mutant a surprisingly complex set of interactions between individual point mutations. These interactions triggered two dominant types of release defects determined by the fraction of 6S RNA released and the rate at which this could occur. Some mutants failed to release almost completely, while others released with a broad range of rates, while others released only partially from $E\sigma^{70}$. The nearly complete release defect of R9-33 significantly interfered with *E. coli* survival on LB agar plates and delayed exponential cell growth in LB liquid culture. The fact that only a small number of point mutations in

R9-33 are required to produce large changes in 6S RNA release rate highlights how mutation can in principle adjust 6S RNA release rate over a huge range of rates allowing natural selection to fine tune transcriptional control for a wide range of bacterial types and environmental conditions.

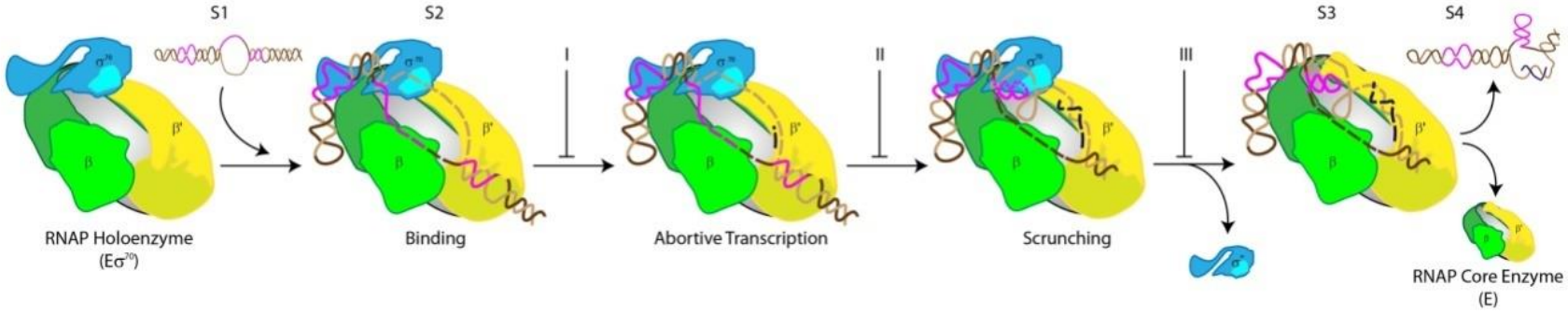



***Figure 2.1. Major proposed steps of E. coli 6S RNA release***

Highly conserved areas in the 6S RNA are shown in purple; the template strand is in light brown and non-template RNA in dark brown. A product RNA (pRNA) is shown in dark blue and triggers the formation of a release hairpin formed from the conserved '-10' region and the top strand of the downstream helix. Figure based on the scrunching-dependent release model for 6S RNA (Panchapakesan and Unrau 2012).

## 2.3. Materials and Methods

### 2.3.1. Library preparation and *in vitro* selection

The starting pool was the Round 5 population of 6S RNA mutant sequences previously selected for their ability to bind to $E\sigma^{70}$ (Shephard, Dobson *et al.* 2010). During four additional rounds of selection, 6S RNA mutant sequences were selected for their ability to bind and not be released from $E\sigma^{70}$ (Figure 2.2A) using the following protocol: DNA from the previous round of selection was transcribed and the resulting RNA gel purified. Pooled RNA at 250 nM was treated as described in the following section. One volume of native loading buffer (50% Glycerol, 0.025% Xylene Cyanol, 0.025% Bromophenol Blue) was added and complexes were resolved by native 5% (37.5:1 acrylamide:bis) PAGE run at 4ºC. The shifted band corresponding to the T1:$E\sigma^{70}$ complex was then carefully excised, and the bound RNA recovered. RNA was eluted O/N (300 mM NaCl, 4°C) and recovered by EtOH precipitation. Using this RNA, the general binding and release was repeated and samples loaded onto a native 5% PAGE. RNA bound to $E\sigma^{70}$ was again recovered before being reverse transcribed. After treatment with 100 mM KOH for 10 min at 90ºC, the resulting cDNA was neutralized with 100 mM Tris-HCl and PCR amplified for the next round of selection (Figure 2.2A).

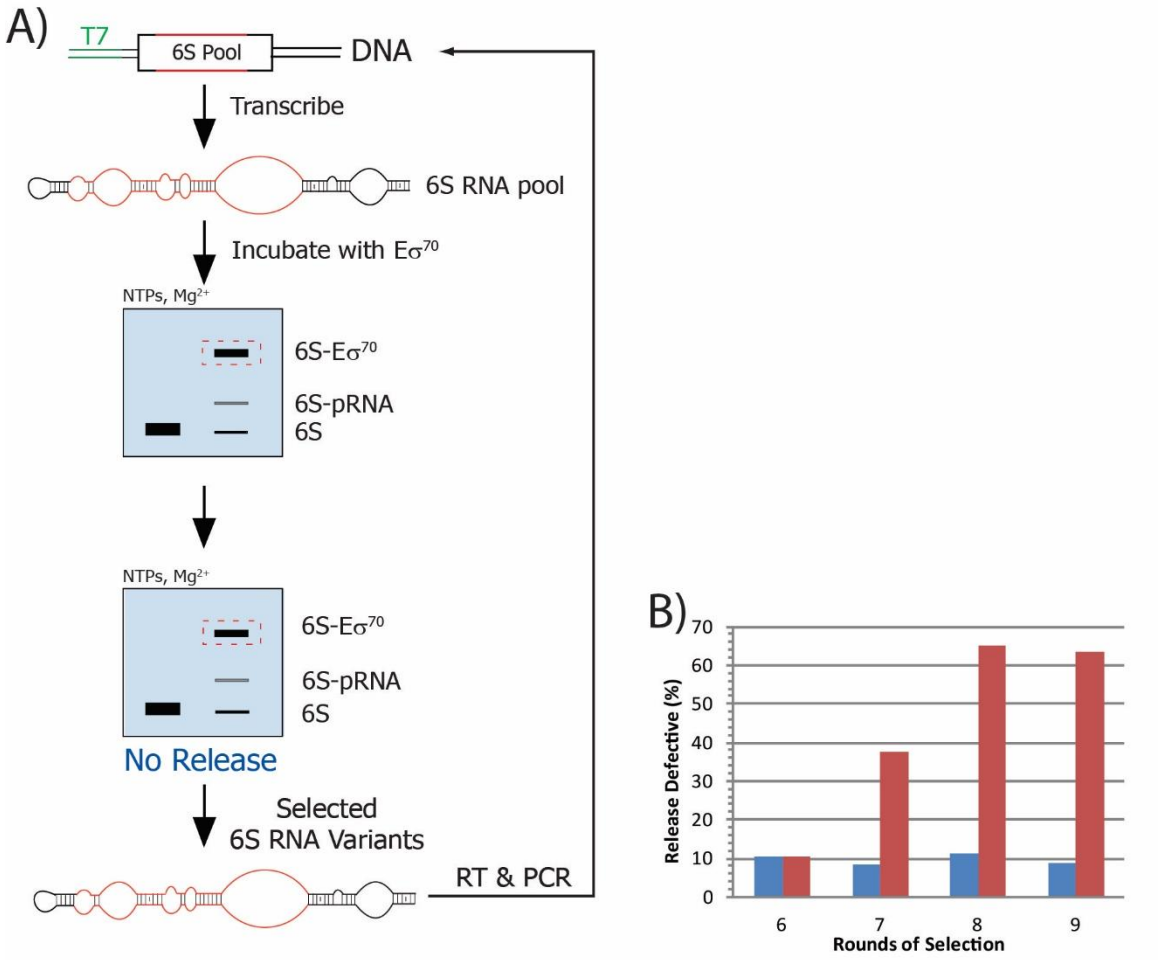



***Figure 2.2. In vitro selection scheme for release-defective 6S RNA mutants***

A) DNA pool having a diversity of $4x10^{12}$ sequences was constructed and selected for its ability to bind (and not release) $E\sigma^{70}$ for five rounds of selection. Residues 42–88 and 104–143 were initially mutagenized with a 10% frequency (T7 RNAP promoter in green, mutagenized residues in red). An 8-fold excess of the RNA pool was incubated with $E\sigma^{70}$ and then incubated under rapid release conditions (250 μM of each NTP and 4 mM $MgCl_2$, for 30 min). RNA capable of remaining in a complex was recovered from a native gel (red dotted box). These steps were repeated and mutants that remained bound to $E\sigma^{70}$ were excised from a second native gel prior to RT-PCR and transcription, ready for another round of selection. B) The bound RNA Pool fraction increased with rounds of selection. The blue bars show the fraction of T1 RNA construct that remained bound after release was induced. The red bars show the fraction of RNA pool that remained bound after each round of selection.

## 2.3.2. General 6S RNA binding and release

The same approach was used during the *in vitro* selection and testing of individual mutant constructs. *In vitro* transcribed 250 nM 6S RNA [Internally $^{32}P$ labelled

and PAGE purified (50 mM Tris-HCl at pH 7.9, 5 mM DTT, 2.5 mM spermidine, 26 mM $MgCl_2$, 0.01% Triton X-100, 8 mM GTP, 5 mM ATP, 5 mM CTP, 2 mM UTP, α-[$^{32}$P]-UTP, template DNA at ~ 100 nM, 5 U/µl T7 RNA polymerase] was mixed with RNA Buffer (20 mM HEPES pH 7.5, 5 mM), heated to 80°C for 2 min and cooled to 50°C for 5 min. Next, RNA was bound to 200 nM *E. coli* RNA polymerase holoenzyme ($E\sigma^{70}$, Epicentre) in 15 mM HEPES at pH 7.5, 90 mM KCl, 0.75 mM DTT, 75 µg/ml Heparin at 37°C for 30 min. Release from $E\sigma^{70}$ was initiated by the addition of 250 µM of each NTP and 4 mM $MgCl_2$ while incubating at 37°C. When individual constructs were tested, time points were taken between 1 and 90 min to differentiate between constructs that release slowly from those that fail to release. During *in vitro* selection, the RNA pool was incubated for 30 min with 250 µM of each NTP and 4 mM $MgCl_2$ at 37°C. Reactions were quenched by the addition of 2X native gel loading dye and resolved in a 5% native gel.

### 2.3.3. *In vitro* transcription using 6S RNA as template

T1 RNA and release-defective variants were *in vitro* synthesized as described above. A final concentration of 250 nM 'cold' RNA was bound to the $E\sigma^{70}$ as described in the previous paragraph. Release was induced with a mix of 250 µM of each NTP, 4 mM $MgCl_2$ and spiked with 25 µCi of [$\alpha$-$^{32}$P]-UTP or [$\gamma$–$^{32}$P]-ATP to radiolabel nascent RNA for 30 min. Reactions were quenched by the addition of 2X denaturing loading dye, and resolved in a 23% denaturing PAGE. To confirm that the product RNA observed was derived from $E\sigma^{70}$, time courses were performed with and without mutant 6S RNA and using either [$\alpha$-$^{32}$P]-UTP (Figure 2.10) or [$\gamma$-$^{32}$P]-ATP (Figure 2.12). Control experiments were set up to distinguish between the specific products of polymerization, contaminants from the radiolabelled NTPs or intrinsic polymerase activity. The nucleotides [$\alpha$-$^{32}$P]-UTP and [$\gamma$-$^{32}$P]-ATP were incubated in the same buffer as the transcription assay and resolved in denaturing gels and the contaminant bands were indicated in Figure 2.9 with star symbols. In addition, one reaction was set up without RNA template in it (Figure 2.11) and as expected no polymerization was seen. All *in vitro* template sequences were verified by sequencing prior to use (Winkler and Breaker 2005).

### 2.3.4. 6S RNA mutant plasmid construction

DNA from T1, R9-33 and LowBinder (Table 2.2) was PCR amplified (10 mM TRIS pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Gelatin, 200 µM each dNTP, 2.5 Units Taq per 100 µl reaction, 0.5 µM primers) with primers 85.4 and 91.1 (Table 2.2) using an annealing temperature of 50°C for 1.5 min. The primers added *Cla* I and *SgrA* I restriction sites, T7 RNA polymerase promoter, *lac* operator upstream of the mutant 6S RNA sequence, and an intrinsic terminator sequence immediately downstream of it. The three different DNA products were cloned into pEcoli-Cterm 6xHN (Clontech) vector as described below (Figure 2.3). The DNA products for T1, R9-33 and LowBinder along with pEcoli-Cterm 6xHN vector were double digested with *Cla* I and *SgrA* I (NEB). Vector was treated with CIAP (Roche) and ligation was carried out withT4 DNA ligase (Invitrogen) following supplier's recommendations. Double digested vector was ligated in absence of insert, and the product (pEcoli-Empty) was used as an additional control during the *in vivo* analysis. Plasmids were transformed into *E. coli* BL21 (DE3) (Novagen) chemically competent cells containing a chromosomal copy of T7 RNA polymerase.

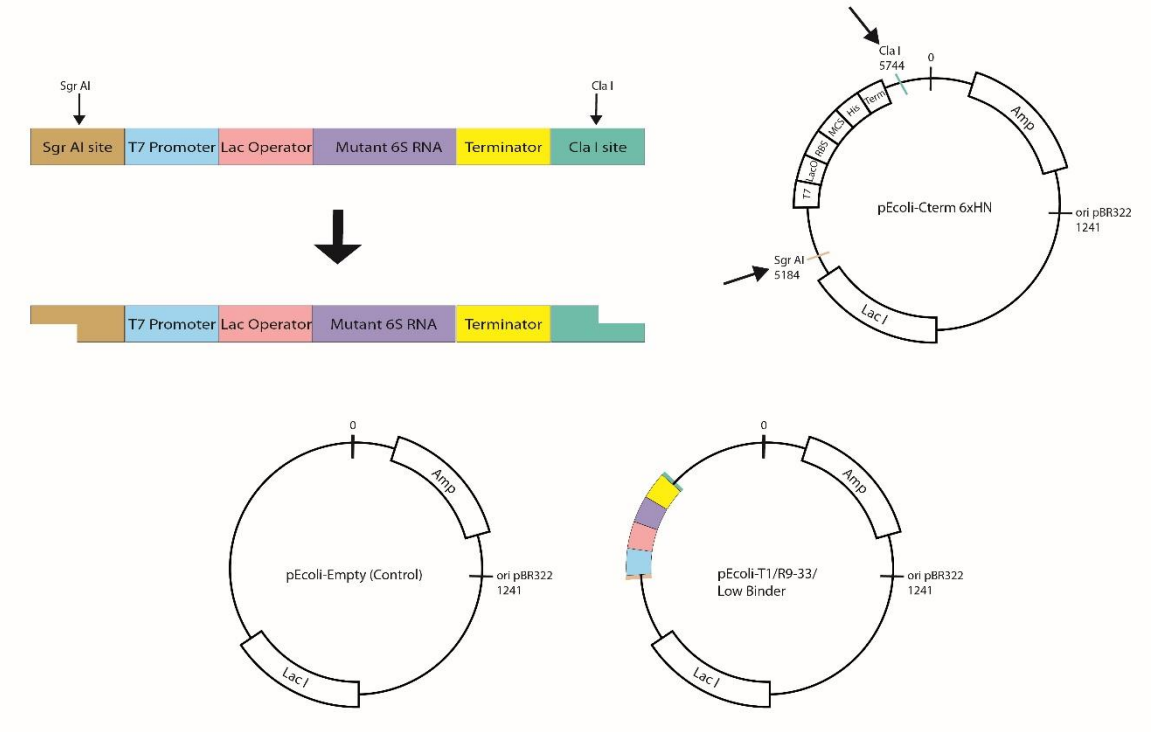



***Figure 2.3. Cloning strategy to express mutant RNA in a pEcoli-Cterm 6x HN vector***

The insert was produced by PCR reactions where the restriction enzyme sites, T7 promoter, *Lac* Operator, and terminator were introduced. Plasmid and insert where double-digested with *SgrA* I and *Cla* I. The pEcoli-Cterm vector was treated with CIAP to avoid self-ligation. Four constructs were used throughout the *in vivo* studies: Negative control does not have the insert illustrated below (no transcription should occur from it), pEcoli-T1 has a shortened version of the 6S RNA sequence that behaves like the full length 6S RNA, pEcoli-R9-33 is the selected release defective mutant sequence, pEcoli-LowBinder has the T1 sequence but 50% of it has been randomized and in consequence does not bind well to the $E\sigma^{70}$.

## 2.3.5. Cell culture

Single colonies of *E. coli* BL21 DE3 cells transformed with the different mutant constructs were picked from LB agar plates and grown O/N in 1ml LB or LB-Amp (100 μg/ml) at 37°C with constant aeration. The next morning cells were diluted 1/100, inoculated into fresh LB-Amp (100 μg/ml) in presence and absence of IPTG (5 mM) and incubated at 37°C with constant aeration. $OD_{600}$ was measured every 60 min until all cultures entered stationary phase.

### 2.3.6. Colony survival assay

Single colonies of BL21 DE3 cells transformed with one of the following: pEcoli T1, pEcoli R9-33, pEcoli LowBinder or pEcoli-Empty, were grown in LB+Amp (100 μg/ml) O/N at 37°C with constant aeration. The next day, cells were diluted 1/5 in fresh LB+Amp and incubated for another 2 hours. Serial dilutions were prepared and 50 μl of each dilution was spread on LB+Amp and LB+Amp+IPTG agar plates (done in triplicate). Plates were incubated O/N at 37°C. Colony forming units (CFU/ml) were determined for each plasmid in the presence and absence of IPTG.

### 2.3.7. Enzymes

The different digestions performed were carried out under the ionic strength and temperature conditions recommended by the manufacturer, using 1 unit (U) of restriction endonuclease per μg of DNA. When the samples needed to be digested with more than one enzyme, the first digestion reaction was terminated by heat inactivation (65°C, 10 min) or extracted with phenol chloroform solution (phenol solution: Chloroform: iso amyl alcohol).

### 2.3.8. T1 RNase digestions

T1 RNase (Fermentas) was serially diluted in 20 mM Sodium Citrate 6 M Urea and incubated with 5' end labelled RNA at 50°C for 10 min to find the optimal dilution. Reactions were quenched by adding 2X denaturating loading dye, snap frozen on liquid nitrogen and resolved in a 23% denaturing PAGE.

### 2.3.9. Alkaline hydrolysis

Five prime (5') end labelled and gel purified pRNA was incubated with 50 mM Sodium carbonate ($Na_2CO_3$) and incubated at 80°C for 10 min or until partial digestion is observed. Reaction was quenched by adding 2X denaturing loading dye, snap frozen on liquid nitrogen and resolved in a 23% denaturing PAGE. The radioactive signal from the gel was detected using a Storm 820 phosphorimager.

### 2.3.10. Transformation of competent *E. coli* cells

Chemically competent *E. coli* cells were transformed with at least 100 ng of DNA. The ligated DNA material was added to the competent cells solution and kept on ice for 10 min. Cells were heat pulsed at 43ºC for 3 min and finally kept at room temperature for approximately 10 min. The cells were transferred to a 4 ml sterile plastic tube to which an extra one ml of LB media was added. The tube was incubated for two hours at 37ºC with constant aeration to allow the expression of antibiotic resistance markers. Aliquots of 50 or 100 µl were distributed in LB-Amp plates and spread evenly with a flamed glass "hockey stick". The remaining cells were stored at 4ºC. After O/N incubation at 37°C, isolated colonies were obtained.

### 2.3.11. Determination of cellular growth rate

A single colony of transformed *E. coli* was used to inoculate 25 ml of LB-Amp in a 125 ml flask. Cells were grown O/N at 30ºC with constant aeration. The next morning, these cells were diluted to an $OD_{600}$ of ~ 0.01. Aliquots were taken every hour and the optical $OD_{600}$ density was determined. The values were plotted on a semi-logarithmic scale, and from the slope generated, the doubling time was calculated.

### 2.3.12. Isolation of plasmid DNA from transformed *E. coli* cells

A small scale bacterial plasmid preparation (Birnboim 1979) was made in order to verify that selected colonies had the right insert. Single bacterial colonies were inoculated in 1 ml LB-Amp, and grown O/N at 37°C. The next morning, the cultures were centrifuged at 12000 rpm for one min. Cells were resuspended with 100 µl of lysis buffer (50 mM glucose, 10 mM EDTA, and 25 mM Tris-HCl pH 8) and incubated at room temperature for 5 min. Next, 200 µl of freshly prepared ice cold solution of NaOH 0.2N and SDS 1% was added; the tubes were mixed by inversions and then stored in ice for 5 min. After the addition of 150 µl of 3 M Potassium acetate pH 4.8, the tubes were mixed by vortexing, and stored on ice for 5 min before being centrifuged for two min. The recovered supernatant was de-proteinized with an equal volume of phenol: chloroform: isoamyl alcohol solution (25:24:1), and mixed vigorously by vortexing. The aqueous and organic phases were separated by centrifugation for two min, and the aqueous portion

containing the nucleic acids was precipitated with cold 95% EtOH for two min while the tubes were sitting on ice. A pellet was recovered after centrifugation, dried and re-suspended in 50 µl of $ddH_2O$ containing RNase A (Sigma Company, St Louis, MO) at a concentration of 20 µg/ml. The sample was mixed briefly and incubated for 1 hour at 37°C. Next, 25 µl of 7.5 M ammonium acetate and 150 µl of 95% EtOH were added to the samples and precipitated for at least 4 hours at -20°C. The DNA was recovered by centrifugation, salt was removed with a 70% EtOH rinse; finally the pellet was dried and resuspended in 50 µl of $ddH_2O$. The presence of the insert was verified by Sanger sequencing.

### 2.3.13. Total RNA extraction

RNA was extracted under denaturing conditions (Rose, Winston *et al.* 1990). Ten millilitres of cell culture at $OD_{600}$ = 0.6 or equivalent amount was harvested and centrifuged at 5,000 g for 10 min. The supernatant was discarded; next 2 ml of LETS buffer (0.1 M LiCl, 10 mM EDTA, 10 mM Tris-HCl pH 7.4, 0.2% SDS) and 2 ml of Phenol were added. The cell slurry was vigorously vortexed for 30 seconds and stored on ice for 30 seconds, for a total 6 min. The sample was then centrifuged for 10 min at 11,000 g. The cleared supernatant was transferred to a fresh tube, an equal volume of phenol-chloroform solution (50:50) was added and vortexed vigorously prior to centrifugation for 10 min at 11,000 g. The supernatant was recovered and precipitated with 300 mM NaCl and 2.5 vol of EtOH.

### 2.3.14. Cell extract preparation

Adapted from Wassarman and Storz (Wassarman and Storz 2000). *E. coli* cells from 5 $OD_{600}$ units (or equivalent) were centrifuged at 5,000 g for 10 min. The pellet was resuspended in 200 µl of lysis buffer (20 mM Tris-HCl pH 8, 150 mM KCl, 1 mM $MgCl_2$, and 1 mM DTT) and mixed with an equivalent of 200 µl of 0.1 mM acid washed (concentrated HCl) glass beads before vortexing for 6 min alternating 30 seconds of vortex and 30 seconds of cooling on ice. An additional 200 µl of lysis buffer was added and samples were centrifuged at 16,500 g for 20 min at 4°C. The cleared extract was aliquoted, flash frozen and stored at -80°C prior to use.

### 2.3.15. Northern Blots

Aliquots of RNA (20 µg total RNA per lane for 8% denaturing gels) or cell extract (0.5 $OD_{600}$ units per lane for 5% native gels). Samples were blotted onto Hybond N+ nylon membrane (GE). Membranes were UV cross-linked using a Stratalinker (1200 µJ for 30 seconds), blocked, probed and washed according to Krieg (Krieg 1996). For all Northern blots shown, *in vitro* transcribed RNAs were used to confirm the absence of cross reactivity. Plasmids that were used as template for *in vitro* transcription were previously digested with *Cla* I, showed two bands for the controls, with the larger corresponding to the runoff product of transcription.

For hybridization, 5' end labelled DNA probes were used. The PNK reactions was carried out as suggested by manufacturer and the efficiency of [$\gamma$–$^{32}$P]-ATP incorporation was determined in some cases by running 0.5 µl of the PNK reaction on a 20% polyacrylamide/ urea gel. The efficiency of the labelling typically ranged between 80-95%, which meant our probes had high specific activity. The blotted membranes were placed in glass bottles in a Hybaid$^{TM}$ mini oven MKII at 65°C, and incubated in a minimal amount of hybridization solution with constant rotation for approximately 4 hours. The hybridization buffer was prepared with 6X SSPE (2 L 10X solution: 173 g NaCl, 27.6 g $NaH_2PO_4$ $H_2O$, 7.4 EDTA, pH 7 with NaOH), 1% SDS, 2X Denhart's solution, and 100 µg/ml of salmon sperm DNA (Rose, Winston *et al.* 1990). After pre-hybridization, approximately 10 µCi of the labeled probe were added and the membrane was hybridized for at least 18 hours. Next, the membrane was washed twice with a high stringency solution (2X SSPE and 0.1%SDS), and twice with a low stringency solution (0.2X SSPE and 0.1% SDS) for 15 min each time at room temperature. The radioactive signal from the membranes was detected using a Storm 820 phosphorimager.

## 2.4. Results

### 2.4.1. *In vitro* selection of release-defective mutants

A simple 5' and 3' truncation of the 6S RNA called T1 (Figure 2.4) was previously shown by *in vitro* selection to contain a '-35' like region (which plays a central role in

RNA binding) and a '-10' like region that strongly modulates both binding and release (Shephard, Dobson *et al.* 2010). A pool of mutants previously selected for its ability to bind $E\sigma^{70}$ (Shephard, Dobson *et al.* 2010) was taken forward by four more rounds, this time selecting for binding and no release (Rounds 6 through 9, Figure 2.2B). Selective conditions were chosen to isolate T1 RNA mutants able to remain bound to the $E\sigma^{70}$ even when incubated under rapid release conditions for periods of time 60 times longer than that required to nearly completely release T1 RNA (rapid release conditions: 250 μM of each NTP, 4 mM $MgCl_2$, for 30 min at 37°C, see methods). The fraction of RNA released from each selection round was compared to the control T1 RNA where 8 to 11% remained bound to the $E\sigma^{70}$ when incubated under the same rapid release conditions (Figure 2.2B, blue bars). After the first round of selection for release-defective mutants, only 10% of the RNA population remained bound to the $E\sigma^{70}$ after incubation in fast release conditions (Figure 2.2B, red bars). The fraction bound increased to 38% by Round 7, 65% by Round 8, and stabilized at 63% by Round 9; at this point the selection was stopped and pool isolates were characterized.

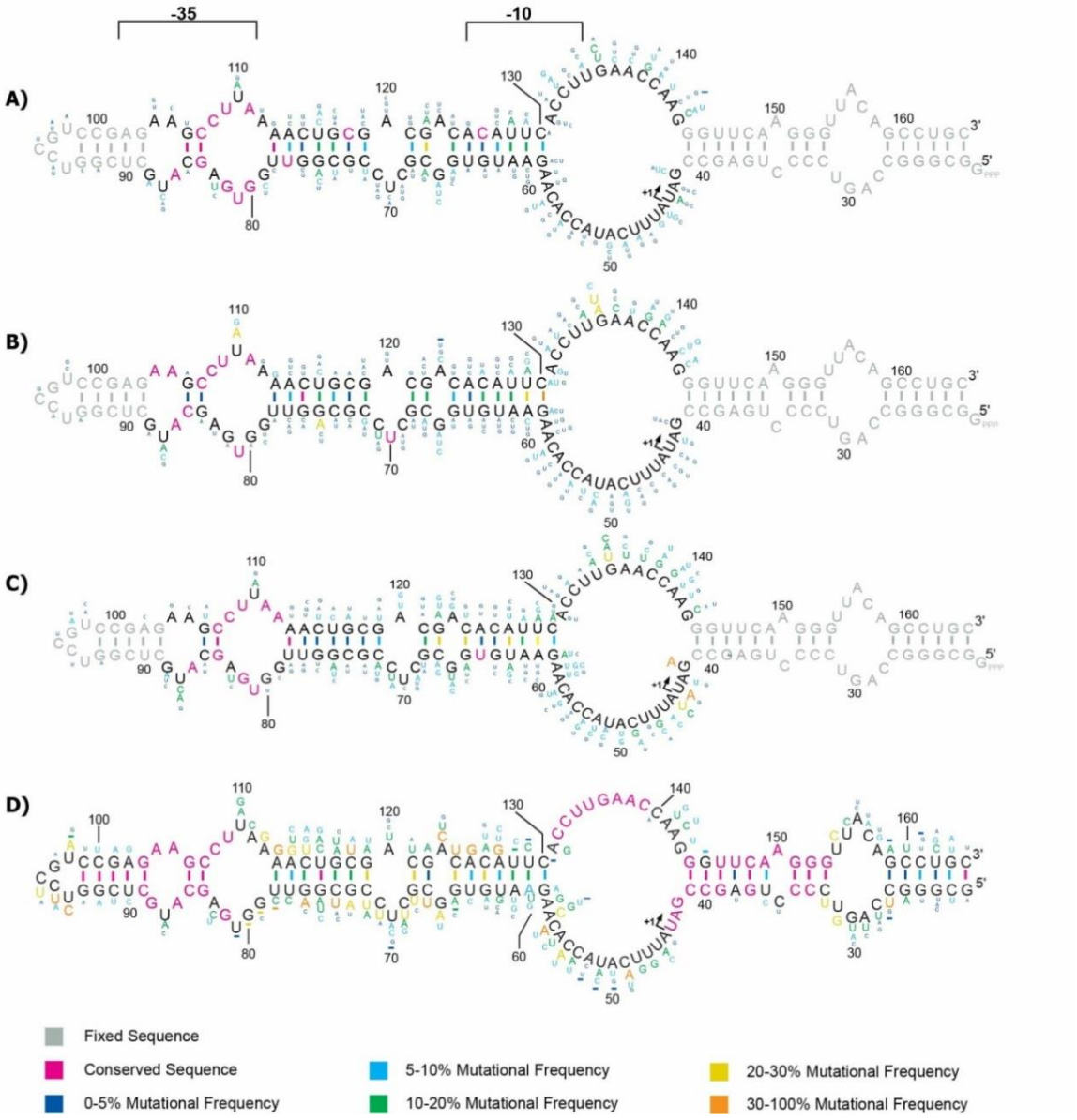



***Figure 2.4. Round 9 release-defective 6S RNA consensus compared to previous selections and γ-proteobacteria phylogeny***

Residues in black were deliberately mutagenized in the initial pool and were found to vary. Purple residues were completely conserved after the selection. Nucleotide variants and base-paired residues are color coded according to the observed percentage of mutational frequency. A) Consensus structure based on 60 variants from the end of the binding selection (Round 5). B) Consensus sequence based on 60 variants from a previous binding and release selection (Shephard, Dobson *et al.* 2010). C) Consensus sequence based on 71 variants from the binding and no release selection (Round 9) performed in this study. The '-35' sequence is conserved in 6S RNA release-defective mutants. Significant changes away from the biological consensus were found in the vicinity of the pRNA TSS (noted by arrow). D) Consensus sequence of phylogenetically conserved 6S RNAs from several γ-proteobacteria (Enterobacteriaceae, Vibrionaceae, and Pseudomonadaceae). Three main regions of highly conserved sequence are found among the γ-proteobacteria, namely the '-35', '-10' and transcriptional initiation and downstream regions.

### 2.4.2. Release-defective mutants were found during the *in vitro* selection

Seventy one isolates were cloned and sequenced from Round 9 of the release-defective selection. Aligning these release-defective sequences revealed a pattern of nucleotide conservation (Figure 2.4) that could be compared to the one observed in our previous selections for binding and release competent 6S RNA mutants (Shephard, Dobson *et al.* 2010) (Figure 2.4B) or to the phylogenetically conserved pattern found in nature (Figure 2.4D). We focused on positions with mutational frequencies significantly different from the 10% frequency initially built into the starting pool (Table 2-1), with the expectation that regions of high sequence conservation or regions of hypermutability would be correlated with release defects.

When aligned, nine nucleotides located within the '-35' sequence island were found to be absolutely conserved (Figure 2.4C). The probability that any one such location was conserved by chance alone was only ~0.06% (i.e. $0.9^{71}$, where 0.9 corresponds to probability of finding the wild type sequence in the starting pool). Notably, the pattern of absolute sequence conservation was shifted by 2 residues towards the LOB of the 6S RNA relative to the pattern observed in both our binding and release selection, and the pattern of sequence conservation found in nature (Figure 2.4 B&D). The shift in conservation pattern was accompanied by the presence of absolute base pairing between residues 78 and 113. This base pairing pattern was not found in the binding and release selection, but was found in the selection for binding (Shephard, Dobson *et al.* 2010), suggesting that pairing position 78 to 113 helps to stabilize the

6S:E$\sigma^{70}$ complex. Residues 83, 87 and 110, located in the '-35' region of high sequence conservation, remained highly variable in all *in vitro* selections as well as in the natural phylogeny, suggesting that these residues do not make specific contacts with E$\sigma^{70}$ upon binding.

According to our *in vitro* selection results, the top strand of the 6S RNA LOB could be divided into two distinct regions based on its sequence conservation. Region 131 to 134 corresponds to residues that can form the base of the release hairpin's left arm, and are highly conserved in the natural phylogeny (Figure 2.4D). During the release-defective selection the region 131-134 exhibited a relatively high level of sequence conservation, having an overall mutational frequency of 5.6% per residue relative to the original frequency of 10% found in the unselected pool population Table 2-1). This data suggests that the region cannot be easily changed and is consistent with previous findings that indicate the region has a complex effect on release rate (Shephard, Dobson *et al.* 2010). In contrast, the region from positions 135 to 141, which forms the upper part of the predicted release hairpin's left arm (Panchapakesan and Unrau 2012), had a significantly higher mutational frequency of 26.2% in the release-defective selection. It is particularly striking that, without exception, the dominant mutation at each position in this region is a transversion (Figure 2.4C), which would be predicted to weaken the formation of the release hairpin and could therefore be predicted to delay or preclude 6S RNA release.

The selection for release defects produced strong deviations from both the wild type consensus sequence and previous selections for binding and release (Shephard, Dobson *et al.* 2010) in the vicinity of the transcription-starting site (TSS, residue U44). Mutations G42A, U44A (the TSS), A45U and to a lesser extent U47G and C49G, had high statistical significance in our selection and were observed significantly more often than expected by chance alone (Table 2-1). These findings agree with previous data on the importance of the TSS region for the efficient pRNA production. However our *in vitro* selection for release defective variants indicates that E$\sigma^{70}$ has a strong preference for initiating with an A in contrast to initiation in wild type sequences where E$\sigma^{70}$ shows no preferences for a specific nucleotide (Cabrera-Ostertag, Cavanagh *et al.* 2013). The remaining nucleotides in the bottom strand (positions 50 to 61) of the LOB were found to

vary at a frequency close to the original frequency of the pool. Based on these statistics the TSS region was implicated in 6S RNA regulation, presumably by controlling pRNA synthesis as previously speculated by others (Wassarman and Saecker 2006).

### 2.4.3. Several classes of mutant 6S RNA release defects were found

The scrunching model predicts that release-defective 6S RNA mutants could result from failures at major steps in the release process (Figure 2.1). Sequences from 71 clones from Round 9 were sorted using a pair-wise alignment approach (Figure 2.5).

Only weak clusters of sequence similarity were found, and 13 clones that spanned the resulting distance tree were tested *in vitro* for their binding and release properties (Figure 2.6). Constructs R9-1 and R9-8 were most closely related to the T1 construct and consistently were the fastest mutants to release from the $E\sigma^{70}$. Even so R9-1 and R9-8 were 15 and 60 times slower to release than the T1 construct (Figure 2.7, Figure 2.8). The remaining eleven isolates released even slower, or almost completely failed to release after 90 min of incubation under rapid release conditions (Figure 2.8). While the sequence of the short pRNAs produced by each of the 13 clones was not explicitly determined, the product bands observed for each correlated well with the template TSS sequence. To detect the products of $E\sigma^{70}$ activity, newly synthesized RNAs from release-defective 6S RNAs were labeled with either [$\gamma$-$^{32}$P]-ATP or [$\alpha$-$^{32}$P]-UTP. The new RNAs were compared to the 13 nt pRNA from T1 RNA and distinguished from contaminant bands found to originate from the radioisotope source vials (Figure 2.9, see Methods). The products of synthesis where classified as either short pRNAs (typically 2-8 nt) or pRNAs (13 nt and longer). Based on the patterns of pRNA synthesis, the release-defective RNAs were grouped into three major classes:

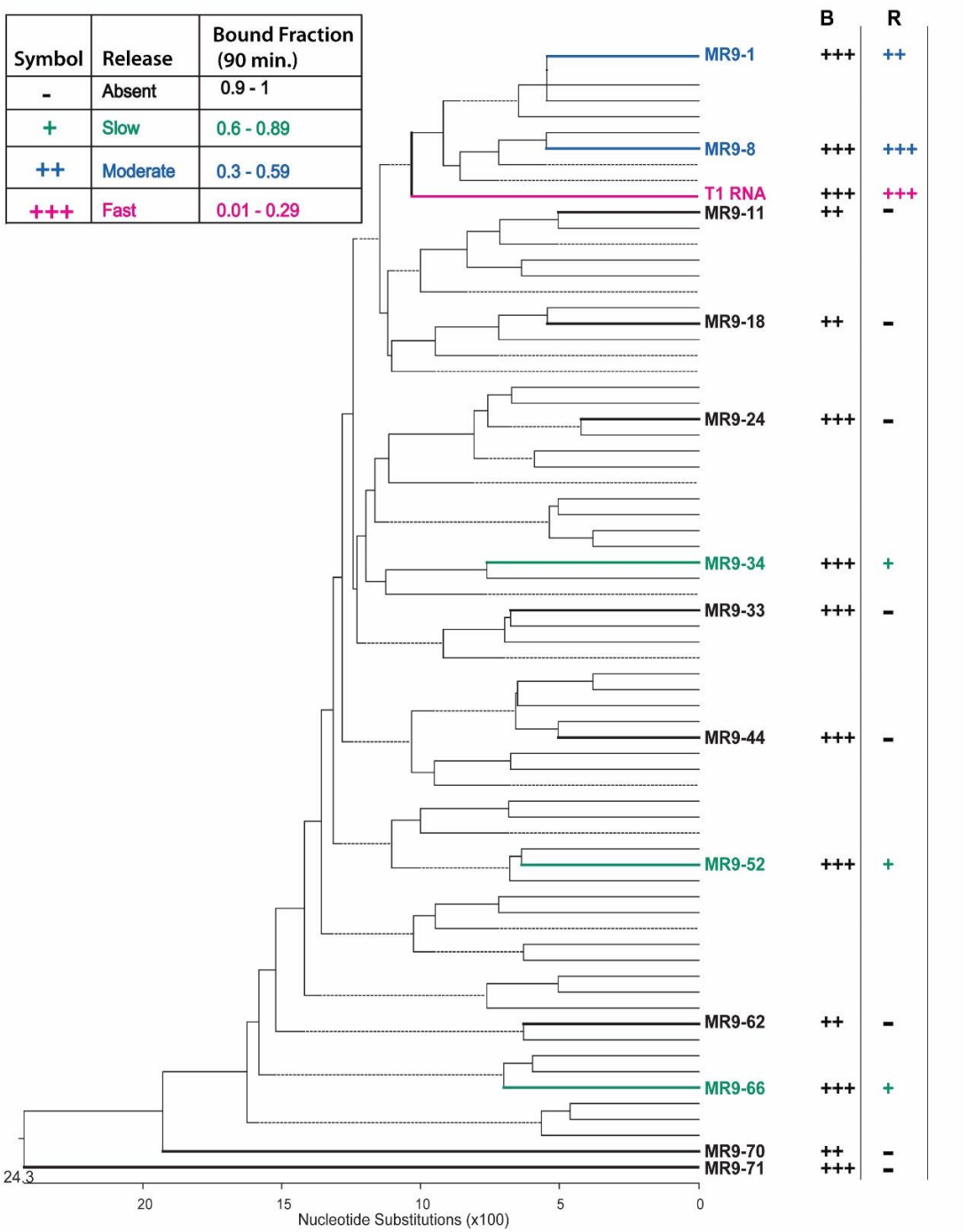



***Figure 2.5. Sequence conservation among release-defective 6S RNA variants***

Mutant 6S RNA sequences were compared to the wild type sequence (T1 RNA in pink) using MegAlign software (DNASTAR). The 6S RNA variants were as expected, capable of efficient binding as summarized in the first column (B): +++, at least 75% of the mutant RNA bound to the $E\sigma^{70}$ or ++, between 75-60 % able to bind. Release rates are summarized in the R column (and keyed to the table on the left) and color coded as Slow (green), moderate (blue), and Fast (pink).

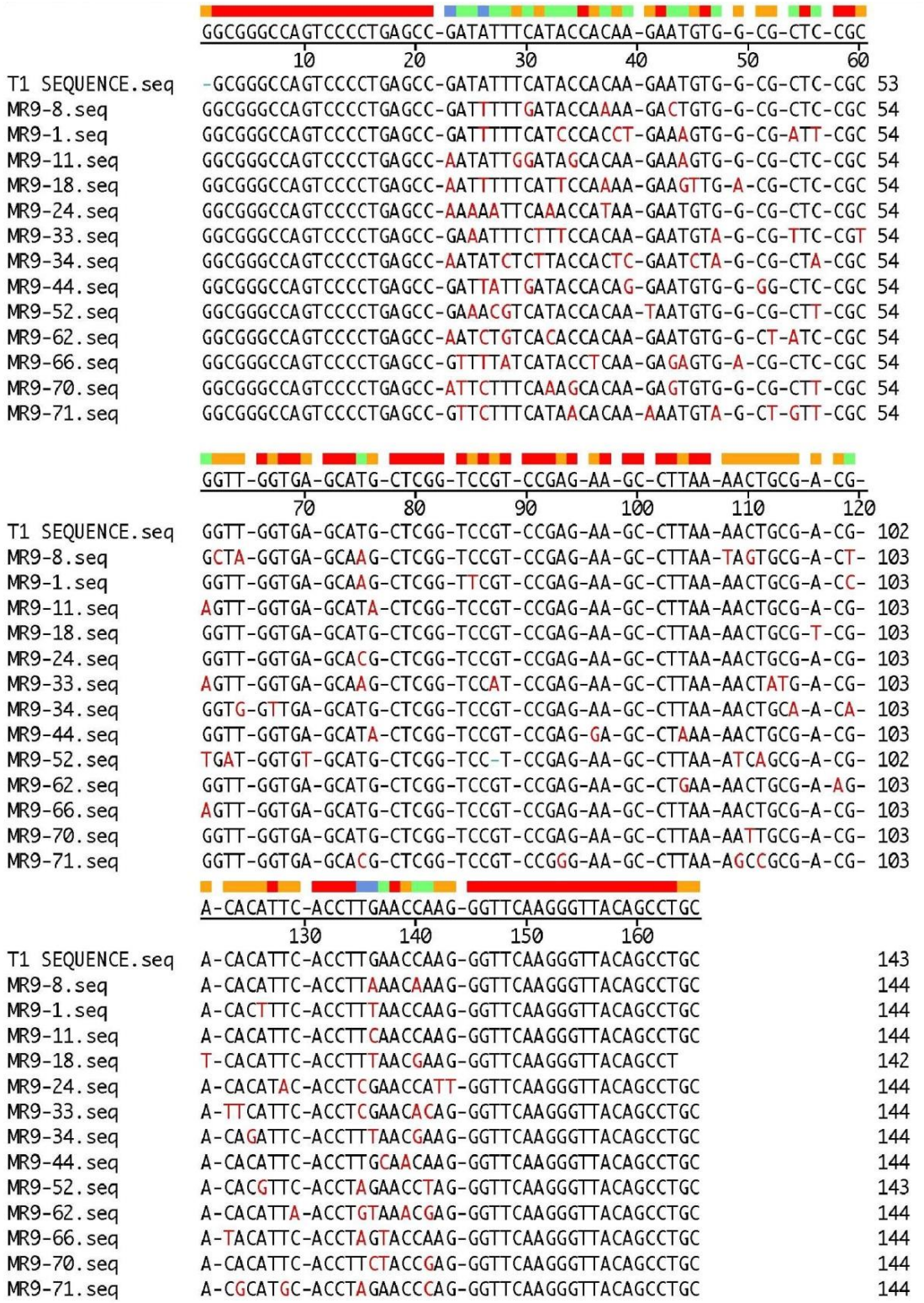



***Figure 2.6. Alignment for Release-defective isolates DNA sequences***

Mutant 6S RNA sequences compared to the wild type sequence T1. Highlighted in red are the residues that differ from the T1 sequence. The spaces in the sequences noted by a black dash indicate the areas of the sequence were the structure of the 6S RNA changes (Figure 2.6 in previous page). Position 24 represents the start of the TSS.

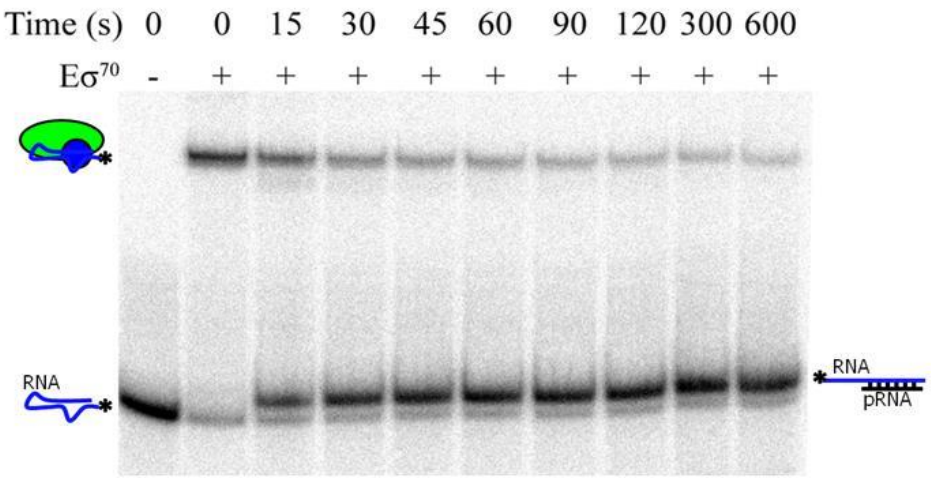
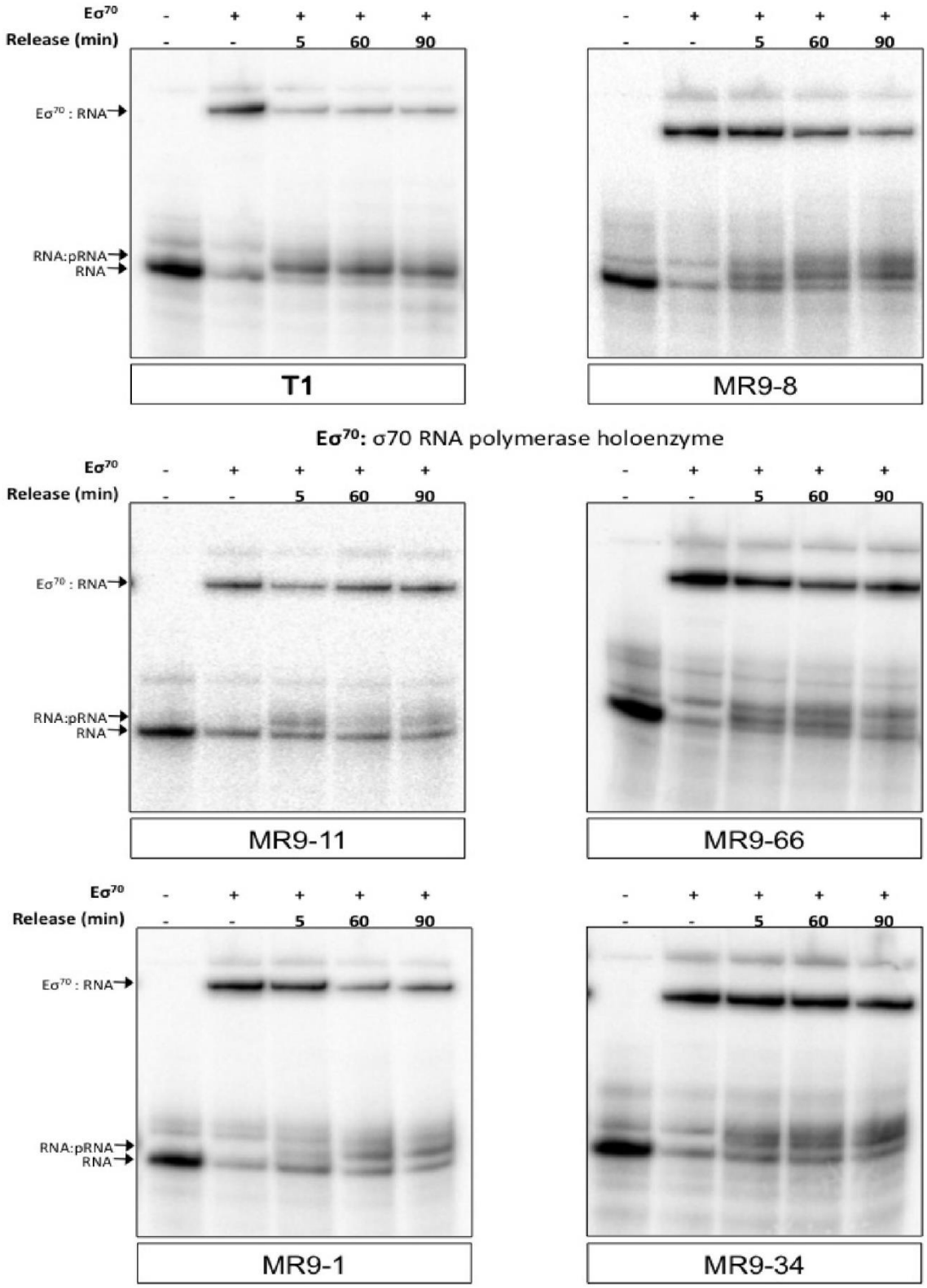
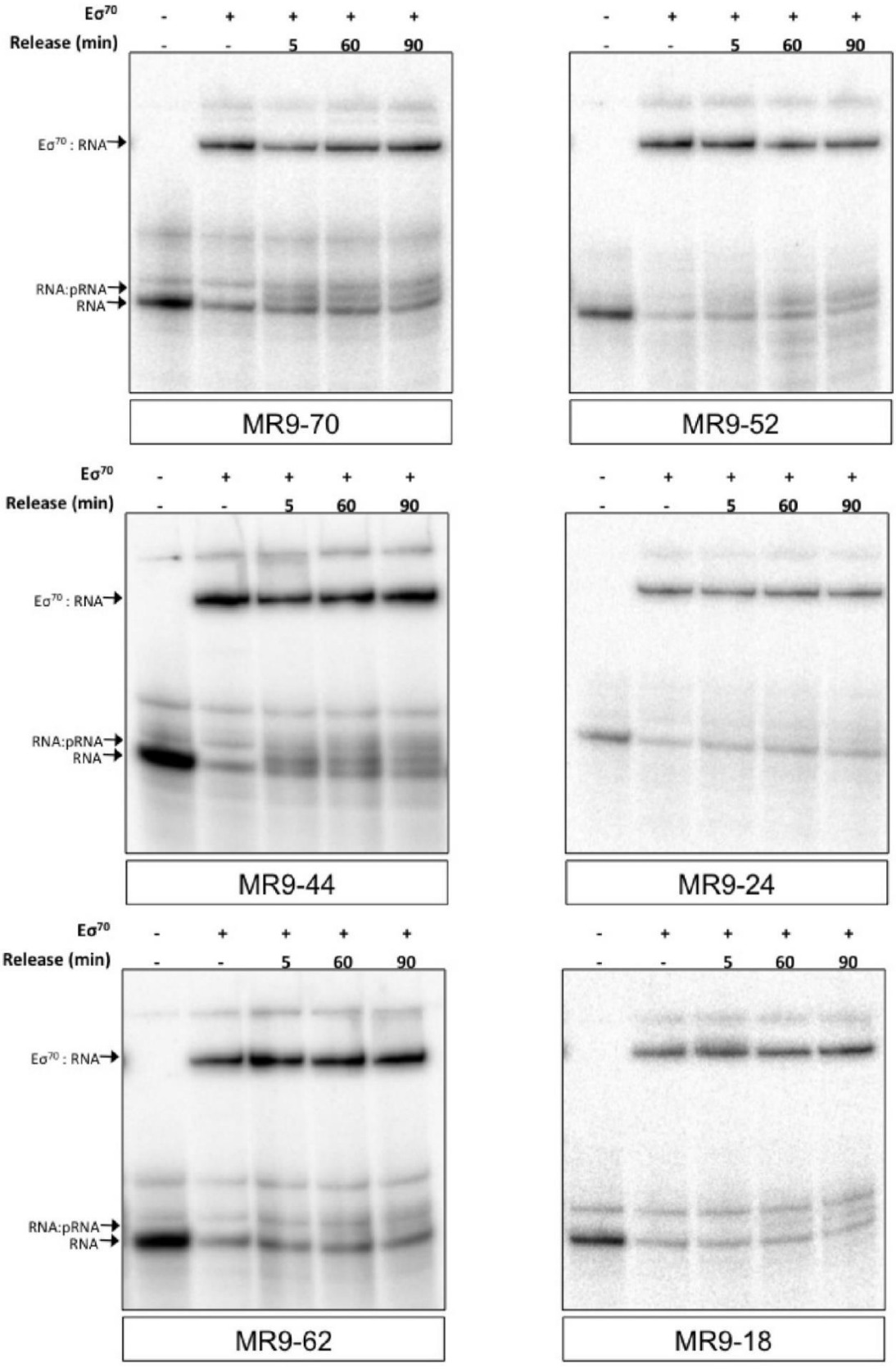



***Figure 2.7. Rapid T1 RNA release kinetics***

Binding and release capabilities were tested for the T1 RNA construct. Body labelled RNA (lane 1) was incubated with $E\sigma^{70}$ (lane 2) and release was induced by the addition of 250 μM of each NTP and 4 mM $MgCl_2$. Time points were taken between 15 and 600 seconds to assess the rate at which the RNA is being released from the $E\sigma^{70}$ (remaining lanes). After 30 s there is no substantial change in the fraction of T1 RNA that remains bound to $E\sigma^{70}$.

The most common class I (R9-70, 52, 44, 62, 18 & 33) produced very low amounts of full length pRNA, and synthesized either apparently normal (R9-33, 18, 62, 70) or low amounts of short pRNA (R9-52 and 44) relative to the T1 RNA reference construct (Figure 2.9). Constructs R9-18 & 62 have a similar sequence in the TSS region (5'..$A_{42}A_{43}$**$U_{44}$**$(C_{45}/U_{45})U_{46}$.., predicted TSS in bold) and share a similar short pRNA pattern, suggesting that both synthesize products with sequence **A**U, **A**UU, etc. (Figure 2.9). Construct R9-33 has a unique template sequence (5'..$G_{42}A_{43}$**$A_{44}$**$U_{45}U_{46}$..) and produced a labeled dinucleotide only with [$\alpha$-$^{32}$P]-UTP present consistent with the predicted $A_{44}$ TSS. R9-33 was characterized in greater detail because after 90 min of incubation under rapid release conditions on average 85% of initially bound RNA remained attached to the $E\sigma^{70}$ (Figure 2.8), the highest of any mutant tested.


Eσ70
Release (min)
- + + + +
- - 5 60 90
Eσ70 : RNA
RNA:pRNA
RNA
T1
MR9-8
Eσ70: σ70 RNA polymerase holoenzyme
MR9-11
MR9-66
MR9-1
MR9-34


Eσ70
Release (min)
- + + + +
- - 5 60 90
Eσ70 : RNA
RNA:pRNA
RNA
MR9-70
MR9-52
MR9-44
MR9-24
MR9-62
MR9-18

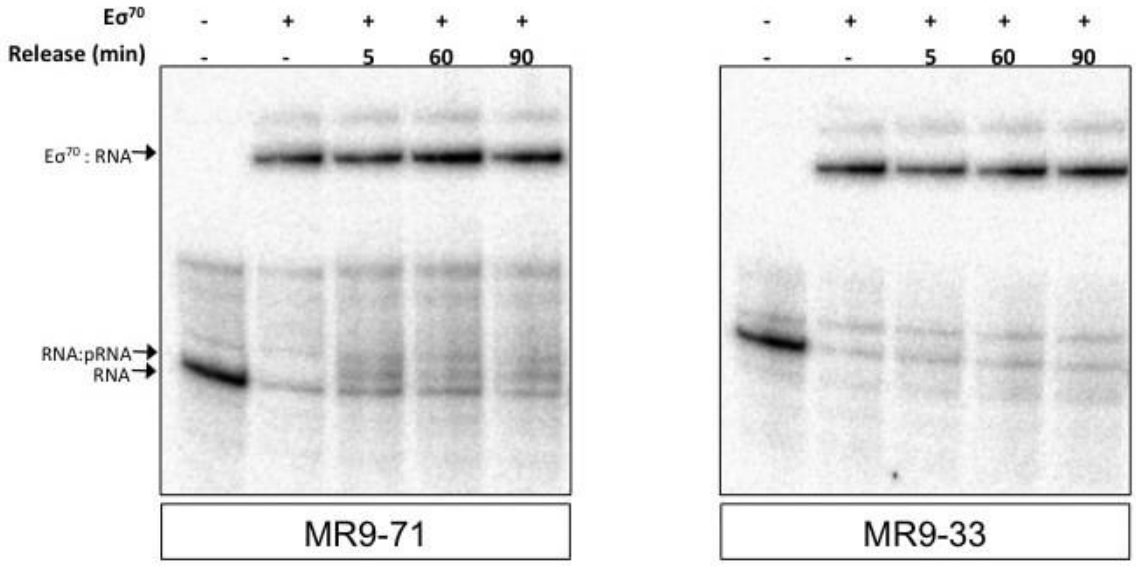



***Figure 2.8. R9-33 mutant 6S RNA fails to release from Eσ$^{70}$***

Binding and release capabilities were tested for different constructs and compared to the T1 RNA. Body labeled RNA (left lane, both panels) was incubated with Eσ$^{70}$ (second lane of each panel) and rapid release was induced by the addition $Mg^{+2}$ and NTPs. Time points were taken at 5, 60 and 90 min. The mobility difference between mutant RNA and RNA:pRNA complex is indicated with black arrows.

Class II (R9-24), was capable of producing very long pRNAs that labeled strongly with [$\alpha$-$^{32}$P]-UTP but not with [$\gamma$–$^{32}$P]-ATP (Figure 2.9, Figure 2.10, Figure 2.12) while exhibiting a very slow release rate (Figure 2.8). The sequence for R9-24 in the TSS region is 5'..$A_{42}A_{43}\mathbf{A_{44}}A_{45}A_{46}U_{47}U_{48}$.., which suggests that this construct can act as a template for the production of oligo (U) containing pRNA. However, during nucleotide feeding experiments (Figure 2.11), pRNA production was not observed when only UTP was used, suggesting that the long polymer produced by R9-24 is either not entirely composed of poly (U) or that additional nucleotides are required to stabilize the polymerization of UTP. We speculate that R9-24 remains bound to the Eσ$^{70}$, due to a template slippage type mechanism. Further work is required to characterize the pRNA synthesis mechanism from this unusual sequence.

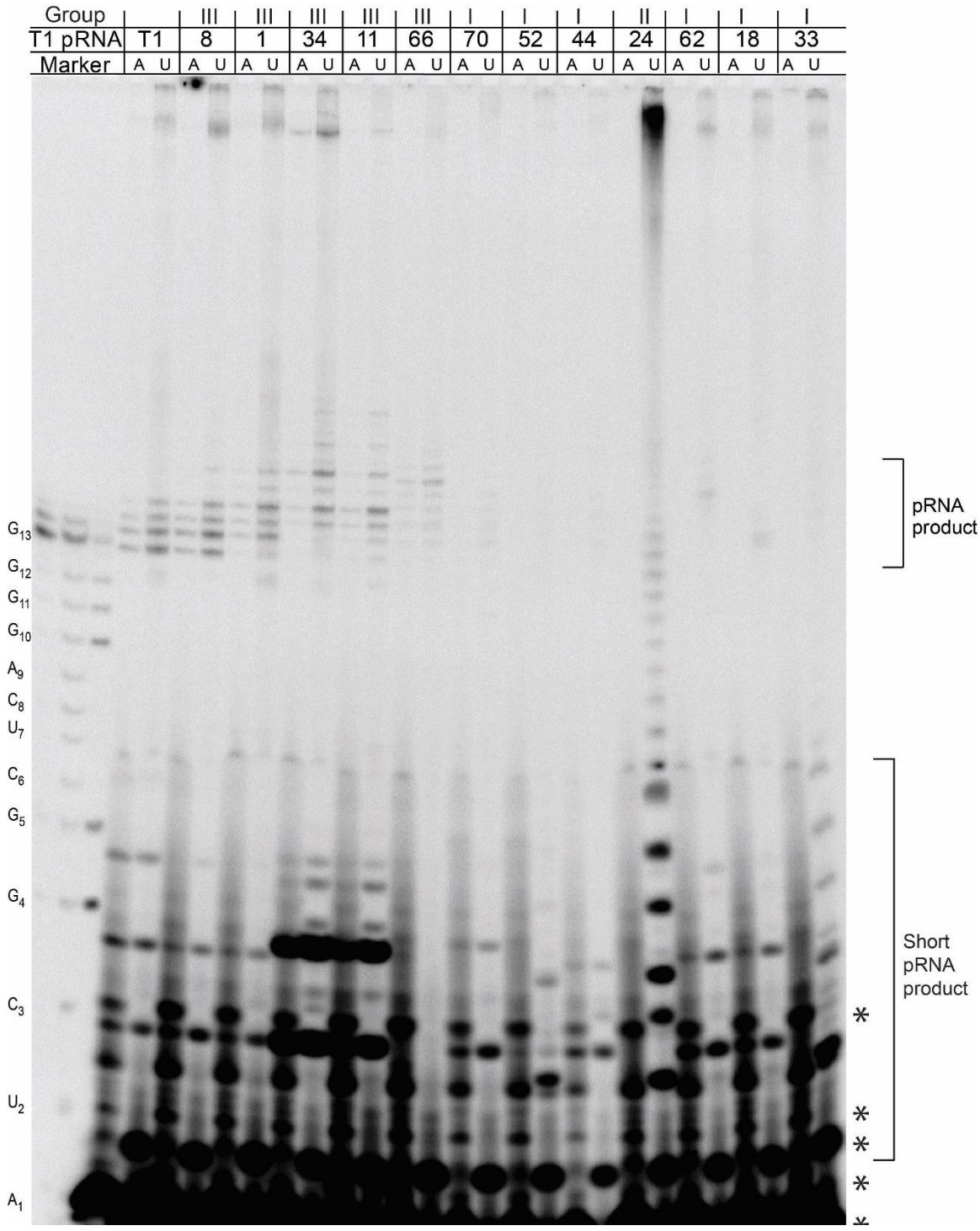



***Figure 2.9. In vitro transcription using 6S RNA release-defective mutants as template***

RNA (cold) was bound to $E\sigma^{70}$ and transcription was induced for 30 min as described in the Methods section. Control T1 RNA (truncated 6S RNA) and 12 mutant release defective RNAs were used as substrates. The newly synthesized RNAs were labelled using either [$\gamma$-$^{32}$P]-ATP or [$\alpha$-$^{32}$P]-UTP. Radiolabeled short pRNA was resolved using denaturing 23% PAGE. The three marker lanes show from left to right are gel purified 13-nt long T1 pRNA, its hydrolysis ladder and a T1 RNase ladder. The letter A stands for [$\gamma^{32}$P]-ATP label, while U for the [$\alpha^{32}$P]-UTP label used in this assay. Note that the hydrolysis and T1 RNase ladders have faster mobilities than the uncleaved pRNA bands due to the presence of an additional 3` phosphate. On the right side of the gel, indicated with a star (*) symbol are the bands that come from contamination in the [$\gamma^{32}$P] isotopes used as resolved in Figure 2.10 and Figure 2.11 (next page).

Class III (R9-8, 1, 34, 11 & 66) produced pRNAs that were as long as or longer than the T1 pRNA, and released from $E\sigma^{70}$ albeit 10 to 100 times slower than the T1 control (Figure 2.7, Figure 2.9). Within this class, R9-1 & 8 had an identical template sequence of 5'...$G_{42}A_{43}U_{44}U_{45}U_{46}U_{47}U_{48}$.. and both produced short pRNAs (2-8 nt long) with an intensity and band pattern identical to the short pRNAs produced by T1 RNA that happens to share a similar template sequence (5'..G44A43U44A45U46U47U48..). On the other hand, R9-11 & 34 [5'..$A_{42}A_{43}$**$U_{44}$**$A_{45}U_{46}(U_{47}/C_{47})$..] produced copious amounts of short RNA. These constructs, just like the T1 RNA conserve the $A_{45}$ residue, which may be of significance. The efficient production of short RNA therefore seems to be strongly modulated by the sequence immediately surrounding the TSS (Cabrera-Ostertag, Cavanagh *et al.* 2013). While this class was not studied in further detail, its phenotype of long pRNA production and slow release rate seems very similar to that of a construct where the downstream top and bottom strands were swapped to prevent the formation of the release hairpin (Panchapakesan and Unrau 2012).

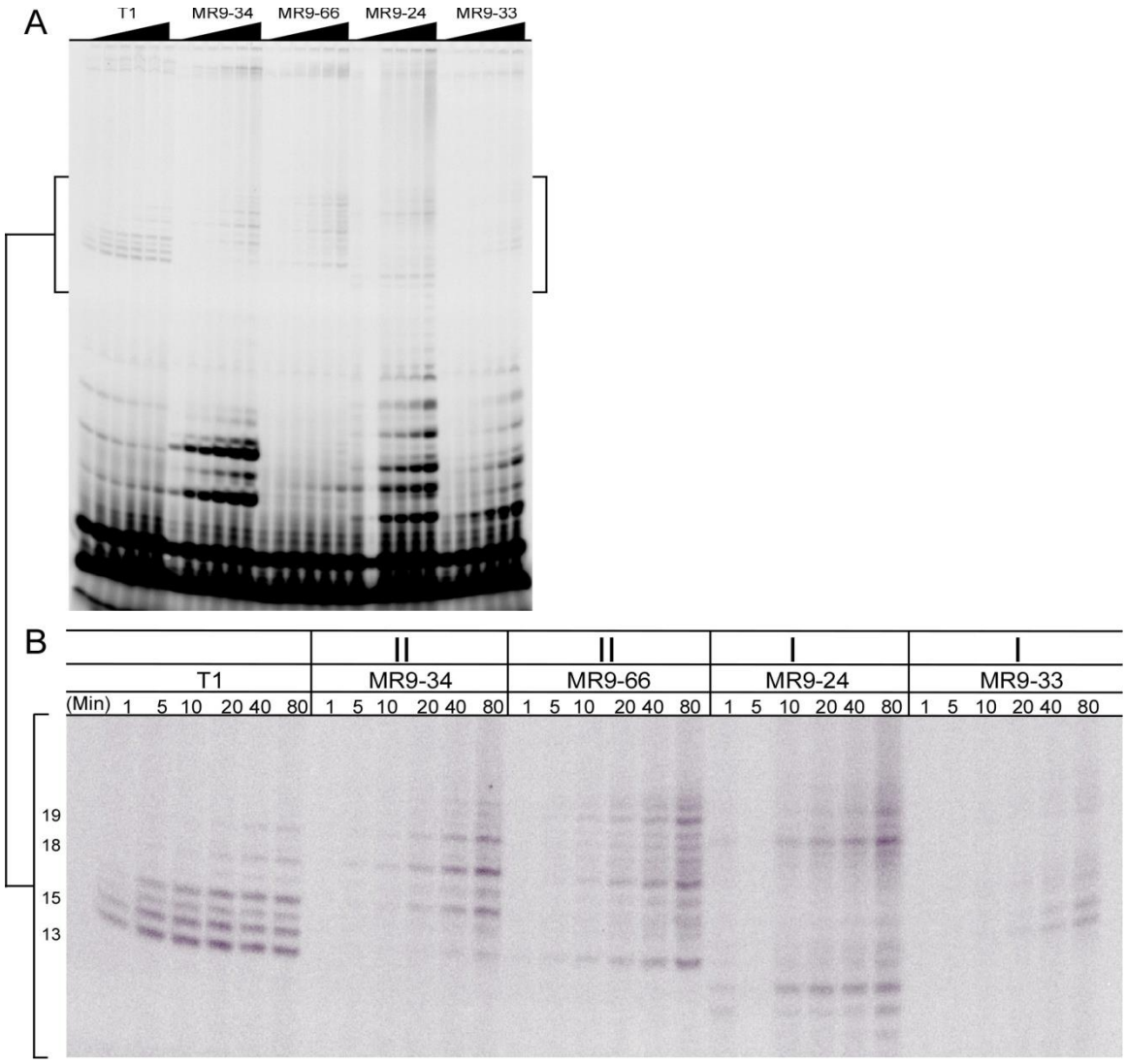



***Figure 2.10. Release-defective mutants produce pRNA slowly and with a broad range of sizes***

pRNA production for release defective mutants was compared to T1 RNA. Newly synthesized pRNA was labelled with [$\alpha$-$^{32}$P]-UTP. Samples were incubated in 250 µM of each NTP and 4 mM $MgCl_2$ for 1, 5, 10, 20, 40, and 80 min and analyzed by 23% denaturing PAGE. A) Constructs MR9-34, 24 and 33 show a rapid accumulation of shorter abortive transcription products in contrast to MR9-66, which shows much less evidence for short pRNAs normally produced during 6S RNA dependent release. B) MR9-66, in spite of producing only trace amounts of abortive initiation product does synthesize pRNA at similar levels to R9-23 and R9-34, while R9-33 produces very little pRNA of any length.

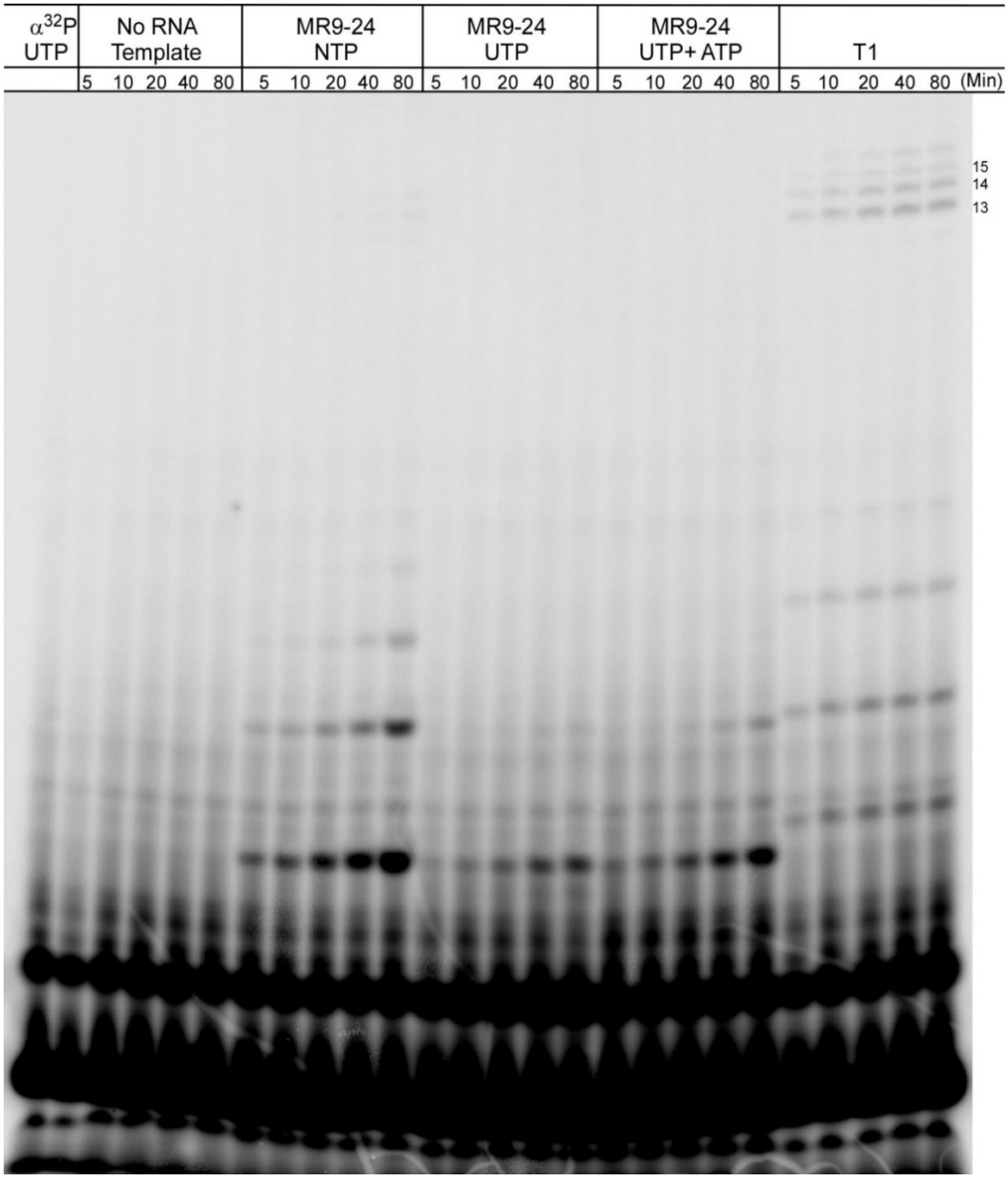



***Figure 2.11. Time course for production of short RNAs and pRNA for R9-24 mutant RNA in presence of different combinations of NTPs***

In the case of MR9-24 the long pRNAs were not observed in the presence of only UTP, other NTPs are required for polymerization to occur. The shortest and most abundant product transcribed from the T1 RNA was 13 nt long. Newly synthesized pRNA was labelled with [$\alpha$-$^{32}$P]-UTP. Samples were incubated in 250 $\mu$M of each NTP and 4 mM $MgCl_2$ for 1, 5, 10, 20, 40, 80 min, and then analyzed by 23% denaturing PAGE. The first lane on the left is [$\alpha$-$^{32}$P]-UTP alone under the same buffer conditions and the next 5 lanes are a "No RNA template" control, where an extension reaction was set up in absence of template RNA. These two controls helped us resolve the origin of the lower bands in the gel and be confident that no polymerization occurs in absence of template RNA.

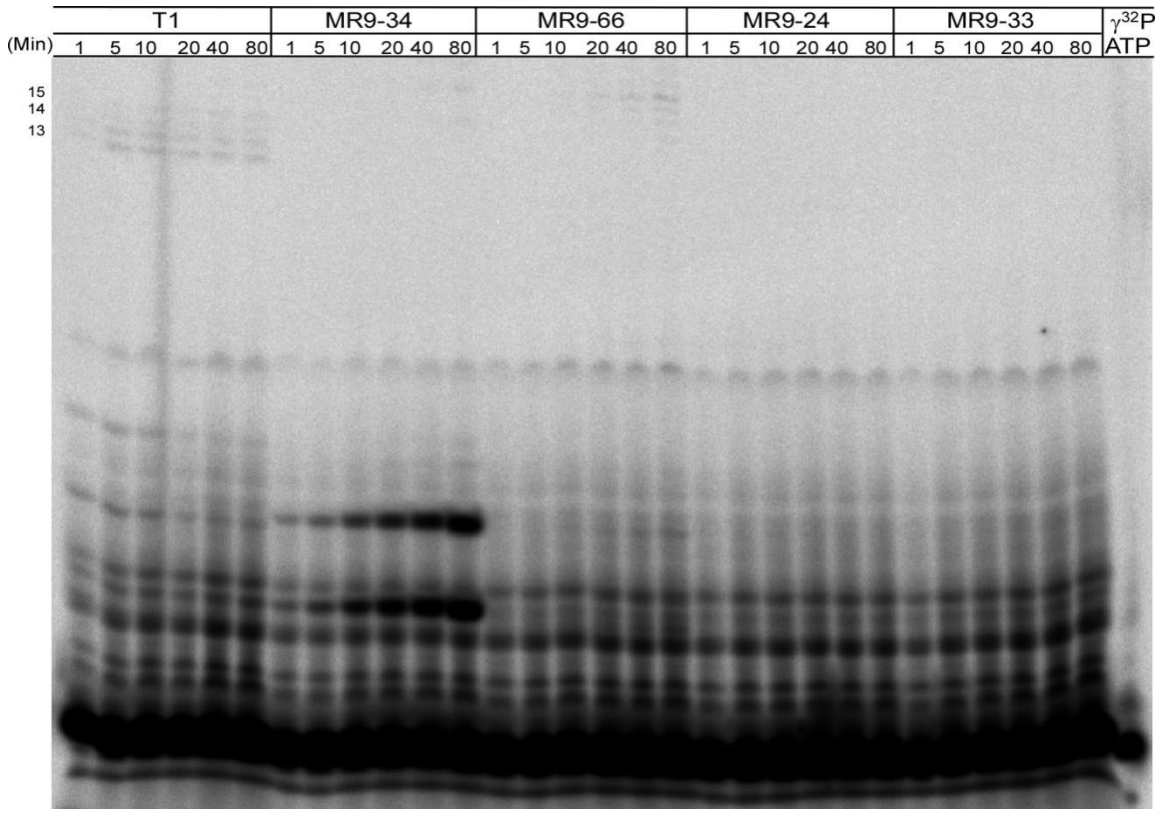



***Figure 2.12. Time course for production of short RNAs and pRNA for release-defective RNAs***

Newly synthesized pRNA transcribed from mutant RNAs was labelled with [$\gamma$–$^{32}$P]-ATP. The pRNA transcribed from the T1 RNA was 13 nt long and was the only one labelled with [$\alpha$-$^{32}$P]-UTP. Samples were incubated in 250 μM of each NTP and 4 mM $MgCl_2$. for 1, 5, 10, 20, 40 and 80 min and analyzed by 23% denaturing PAGE. The right most lane is [$\gamma$–$^{32}$P]-ATP alone under the same buffer conditions.

## 2.4.4. Complex, synergistic effects between point mutations in the R9-33 RNA

Since R9-33 was the most release resistant variant found in this screen, we decided to explore which of its mutated residues most heavily influenced release from E$\sigma^{70}$. Systematically adding the substitutions present in R9-33 back into the T1 RNA scaffold, produced two significant phenotypes: a slowdown in the release rate or a failure to release from the E$\sigma^{70}$. Often failure to release was not complete and thus combinations of the two phenotypes were commonly observed. While a slowdown in release can be explained by a single rate limiting step in the process of 6S RNA release,

a partial release defect implies a more complex failure mechanism, suggestive of a multi-step process.

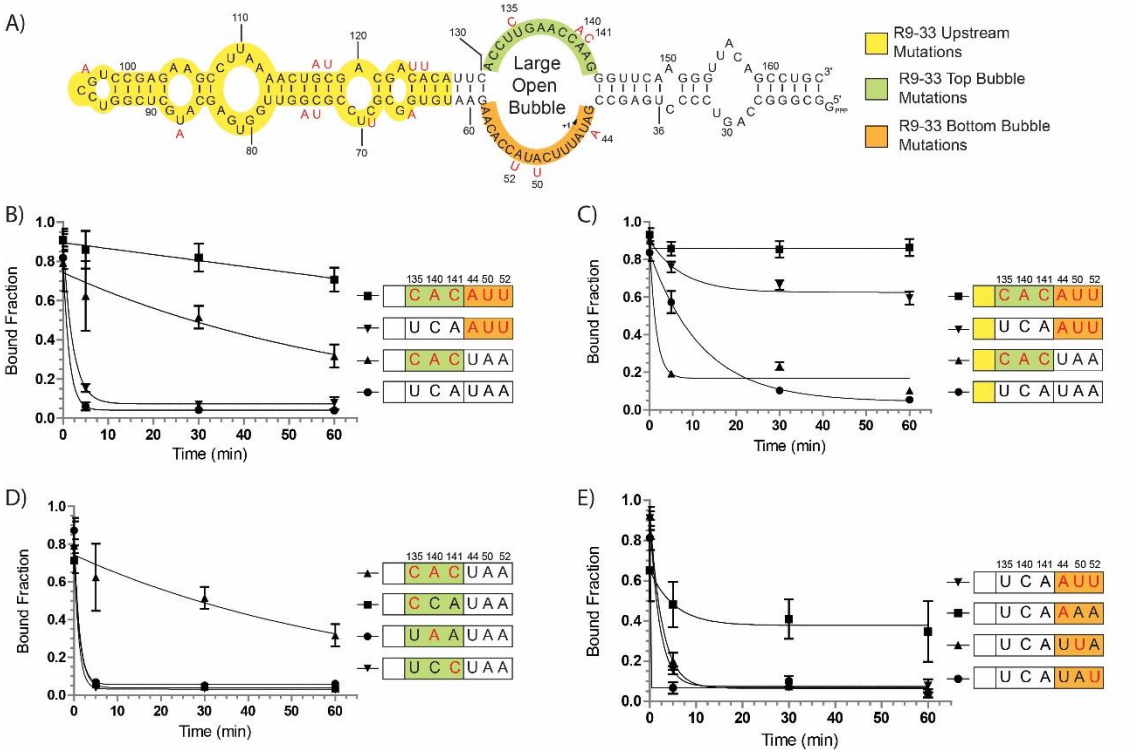



***Figure 2.13. Substitutions around the large open bubble (LOB) of R9-33 act in a highly synergistic manner and are responsible for blocking release from Eσ$^{70}$***

A) T1 RNA sequence is noted in black while substitutions found in R9-33 are in red letters. The molecule was divided into three distinct zones: A yellow upstream region that contains 10 mutations. A green top strand in the LOB region contains three substitutions at positions 135, 140, and 141. An orange bottom strand in the LOB contains three substitutions at positions 44, 50, and 52. B) Release kinetics for T1 RNA and three other constructs with triple mutations (in red letters). C) Release kinetics for R9-33 constructs with substitutions that were sequentially removed according to the color code described in panel A. D) Release kinetics for constructs with point mutations (in red letters) in the top strand of the 6S RNA LOB. E) Release kinetics for constructs with point mutations (in red letters) in the bottom strand of the 6S RNA LOB. Error bars correspond to standard deviations about the mean of at least three independent experiments.

The R9-33 sequence differs from the T1 RNA by a total of 16 nucleotide substitutions (Figure 2.13 A). Ten of these mutations were found in upstream regions of the 6S RNA that are weakly conserved either in nature or in selections for binding and release (Shephard, Dobson *et al.* 2010). Eight of these mutations were found between the '-35' region and the LOB (Figure 2.13A residues in red, yellow region). Interestingly,

these mutations substantially preserve the predicted secondary structure in this region: Four of these eight mutations replace two wild type G:C pairs with new A:U pairs. C69U is not predicted to be in a base pairing region, while G75A, repairs a G:U wobble pair and C74U creates an adjacent wobble pair. Only A125U would be predicted to destabilize the predicted upstream secondary structure and this mutation creates a potential U:U mismatch. Further upstream, only two mutations were found: The U87A mutation is located at a position that both in nature and in selections for binding and release, or binding and no release is highly variable (Figure 2.4). Further, G97A mutation is located in a region known to be dispensable for 6S RNA release (Shephard, Dobson *et al.* 2010). The remaining six mutations were split evenly between the top strand of the LOB (green region), and the bottom strand of the LOB (orange region). The top strand mutations U135C and C140A would be predicted to interfere with changes in RNA secondary structure that occur during pRNA-induced hairpin formation, while the bottom strand U44A mutation is located at the predicted TSS and could easily influence 6S release kinetics.

Top and bottom strand LOB mutations together were required for a strong release defect to manifest and this phenotype was enhanced in a complex way by the inclusion of the upstream mutations. The triple substitutions in either the top or bottom strands of the LOB were found to slow down release (Figure 2.13A inverted triangles and triangles) 25 and 2 (Table 2-3) fold respectively compared to the T1 sequence (circles). However, when all six substitutions around the LOB were combined, we observed an almost complete failure to release (Figure 2.13B, squares). Like the top and bottom strand mutations, the presence of only the 10 upstream mutations (Figure 2.13A yellow region) was found to slow release rate by 10 fold (compare Figure 2.13C and B circles). Interestingly, these upstream mutations could behave in a synergistic or antagonistic manner with mutations found in the LOB. When the 10 upstream mutations were combined with bottom strand LOB mutations, ~30% of the RNA released at a rate 8 times slower than the T1 reference sequence. The remaining ~60% of the mutant RNA released at a substantially slower rate (compare Figure 2.13C and B inverted triangles). Conversely, the combination of upstream mutations and top strand LOB mutations resulted in release at a rate similar to that found for the top strand LOB mutations by themselves, but in this case ~15% of the mutant RNA population released

at a considerably slower rate (compare Figure 2.13C and B triangles). When the upstream mutations were combined with mutations in the top and bottom strands of the LOB to recreate the full R9-33 construct, the full release defect was observed as expected. It is clear from these observations that complex interactions between these three distinct regions of the 6S RNA molecule serve to control major events that occur during 6S RNA release.

While the complete set of point mutations in the top and bottom strands could trigger major changes in release rates, the individual point mutations generally did not trigger major changes in release rate or result in a partial failure to release. Point mutations U135C (Figure 2.13D, squares), C140A (circles) and A141C (inverted triangles) did not by themselves change release rates significantly. However, when the three mutations were combined as discussed previously (Figure 2.13D, triangles), release was slowed 25 fold compared to T1 suggesting that a release defect on the top strand demands synergistic effects between the three residues.

In contrast to the top strand mutations, where all three substitutions were required for a release defect to manifest, mutation some bottom strand residues could suppress a release defect induced by mutating the TSS. By itself, the U44A mutation, resulted in a strong release defect where ~40% of the mutant RNA remained bound (Figure 2.13E squares). This was surprising as the triple mutant on the bottom strand (inverted triangles) released only two fold slower than the T1 construct. To explore this further we constructed an A50U mutant (triangles) which released only 2.5 fold slower than T1, while the A52U mutant (circles) released with a rate very similar to that of the T1 construct. The marginal effect that A50U and A52U had on release rate makes it hard to understand how these two mutations, when combined with the U44A mutation, can nearly completely rescue the release defect induced by U44A on its own.

### 2.4.5. Effects of *in vivo* expression of mutant 6S RNAs

From our *in vivo* expression studies, I concluded that mutant 6S RNAs inhibit bacterial growth and decrease cell viability in high nutrient conditions. *In vitro* studies provided us with a set of 6S RNA variants that can bind and sequester $E\sigma^{70}$ just like the endogenous

6S RNA, but with varying release rates. The best release-defective mutant R9-33, together with the T1 RNA, and a mutagenized 6S RNA with low affinity to $E\sigma^{70}$, were cloned into a modified pEcoli-Cterm 6xHN (Clontech) vector driven by a T7 promoter (Figure 2.3). This type of strongly inducible system used for expressing our mutants allowed us to draw conclusions from early time points where strong expression of plasmid-derived RNA was verified by Northern blot. The RNA transcribed *in vivo* from pEcoli-R9-33, and pEcoli-T1 DNA (due to the presence of the *lac* O and terminator), was longer than the R9-33 and T1 RNA (143 nt), but *in vitro* they showed the same binding and release capabilities as their shorter counterparts shown in Figure 2.8. When tested *in vitro* for binding and release, the RNA transcribed from pEcoli-LowBinder DNA produced a slower mobility band, relative to bound T1 RNA and did not show any evidence of pRNA dependent release. The same LowBinder RNA could be displaced from $E\sigma^{70}$ by T1 RNA during *in vitro* competitions assays (data not shown). For all of these constructs, induction of the chromosomally expressed T7 RNA polymerase by IPTG served to decouple (for the lifetime of T7 RNA polymerase), the transcription of mutant RNAs from the activity of bacterial $E\sigma^{70}$.

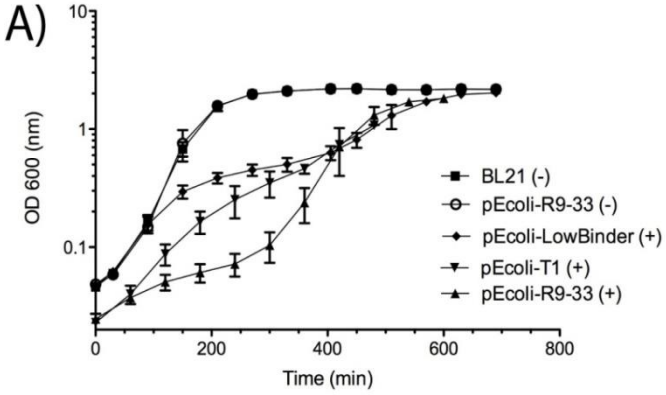
### 2.4.6. Delay in cellular growth correlates with severity of mutant phenotype

Expression of pEcoli-LowBinder, pEcoli-T1 and pEcoli-R9-33 produced a variety of growth patterns that can be correlated well with the *in vitro* behaviour of each construct. Freshly transformed cells were grown O/N in LB+Amp, diluted 200 fold the next day and induced with IPTG immediately after dilution. When grown in the absence of IPTG, all transformed cells grew at the same rate as untransformed cells grown in LB (Figure 2.14A squares and empty circles & data not shown). When induced with IPTG, *E. coli* transformed with the different plasmids showed clear evidence of unbalanced growth, which is reflected in changing rates of growth rate during the exponential phase. In contrast, uninduced cells showed a constant rate of cell division during exponential growth. The pEcoli-LowBinder (Figure 2.14A diamonds) and pEcoli-T1 (inverted triangles) cells grew slower than the wild type and their growth curves presented three different growth rates prior to entering stationary phase. Growth for pEcoli-R9-33 (triangles) transformed cells also showed three different slopes and growth was minimal

until ~ 300 min post induction when they reached a growth rate similar to that of uninduced or wild type cells. A fourth control construct, pEcoli-empty (self-ligated double-digested vector) grew like untransformed cells, suggesting that the expression of mutant 6S RNA was responsible for the slowdown in growth (data not shown). This data, particularly at early time points, suggests that pEcoli-R9-33 expression significantly inhibits cell growth even in high nutrient conditions.

To explore the hypothesis that cells expressing pEcoli-R9-33 plasmid-derived RNA, also had a lowered viability, aliquots of uninduced exponentially growing cells were plated onto LB+Amp agar with or without IPTG (Figure 2.14B). The percentage of cells able to form colonies (CFU) was calculated relative to cells transformed with pEcoli-Empty, which had an absolute colony forming potential of 85% (colonies formed on +IPTG plates relative to –IPTG plates). Cells transformed with pEcoli-LowBinder and grown on +IPTG agar showed a CFU ability of 82% relative to the pEcoli-Empty control, suggesting that the majority of these cells remained viable upon IPTG induction. In contrast, only 2.6% of the cells transformed with pEcoli-T1 formed colonies implying that high levels of 6S RNA expression either prevents cell division and/or potentially causes cell death. Strikingly, cells transformed with pEcoli-R9-33 exhibited survival rates of only 0.2%. The fact that both colony forming ability and the altered growth observed in liquid media correlate with the *in vitro* phenotypes of each RNA construct, suggests that inhibition of transcription by either T1 or mutant 6S RNA expression is at the root cause of these phenotypes.

A)




B)

| Construct | CFU Relative to pEcoli-Empty |
|---|---|
| pEcoli-Empty | 100.0 |
| pEcoli-Low binder | 82.2 |
| pEcoli-T1 | 2.6 |
| pEcoli-R9-33 | 0.2 |

***Figure 2.14. Bacterial over-expression of mutant 6S RNA produces growth defects and correlates with in vitro release defects***

A) Growth curves for E. coli BL21 (DE3) cells transformed with one of the following plasmid vectors; pEcoli-LowBinder, pEcoli-T1, or pEcoli-R9-33. Cells were grown in LB+Amp (100 µg/ml) and 5 mM IPTG (+ symbols) from time of inoculation. Cells transformed with pEcoli-R9-33 but grown in absence of IPTG (- symbols) were used as a control together with untransformed BL21 cells grown in LB broth. Error bars correspond to standard deviations from the averages of three independent experiments. B) Expression of mutant RNA correlates to cell survival and growth. Exponentially growing cells ($OD_{600}$= 0.7) in LB+Amp were plated onto LB+Amp agar with and without 5 mM IPTG. Percentage of CFU was calculated for each construct and expressed as percentage of survival normalized to that of the pEcoli-Empty. The control plasmid pEcoli-Empty grew like pEcoli-R9-33 – IPTG in LB-Amp liquid culture (data not shown in panel A).

## 2.4.7. T1 and R9-33 RNAs are initially strongly expressed *in vivo*

In order to track mutant RNA expression *in vivo,* total RNA was extracted from cells grown in LB+Amp in the presence or absence of IPTG. Equal amounts of total RNA were loaded into denaturing polyacrylamide gels and analyzed. Plasmid-derived RNA and endogenous 6S RNA (184 nt) could be visualized by SYBR Green staining and observed band patterns confirmed by Northern analysis (Figure 2.15A). T1 and R9-33 RNA was synthesized *in vitro* using plasmids digested with *Cla I* as template (pcT1 and pcR9-33), and used to control for probe specificity and the size of the plasmid-derived RNA. The two bands observed in the plasmid cut lanes (pcT1 and pcR9-33) correspond to the RNA produced by runoff transcription (top band, 226 nt long that includes the

operator and terminator sequences in the vector, Figure 2.3), and the termination at the appropriate site (bottom band). Levels of endogenous 6S RNA was low at early time points for all the cells. In the case of cells transformed with pEcoli-R9-33, plasmid-derived RNA production was the highest 120 min after induction and decayed steadily to very low levels after 400 min of growth. This decay is in agreement with the time when cells entered their fastest growth (Figure 2.15A). The ratio of R9-33 to 6S RNA as quantified by SYBR Green staining ranges from 10 to 1 at 120 min to an almost equal ratio at 400 min after induction suggesting that the initial high concentrations of R9-33 are responsible for growth inhibition at early times. Expression of pEcoli-T1 RNA peaked at ~340 min and gradually decayed until it could not be detected at 1,560 min (stationary phase) as monitored by a Northern probe able to detect both pEcoli-T1 and endogenous 6S RNA simultaneously (T1* probe, Figure 2.15B).

To verify that plasmid-derived RNA binds to cellular $E\sigma^{70}$, whole cell extracts from +ITPG cultures grown for 180 min were prepared and tested by native gel Northern blot. Plasmid-derived R9-33 RNA forms a complex (Figure 2.16, right lane) with a similar native gel mobility to that of a complex prepared from commercial $E\sigma^{70}$ (Epicentre) and *in vitro* transcribed R9-33 RNA (Figure 2.16, left lane). The slow mobility band on the Native lane (Figure 2.16) was only 24% of the total signal which suggests that the strong expression of plasmid-derived R9-33 RNA triggered by IPTG induction effectively inhibited all bacterial $E\sigma^{70}$ transcription.

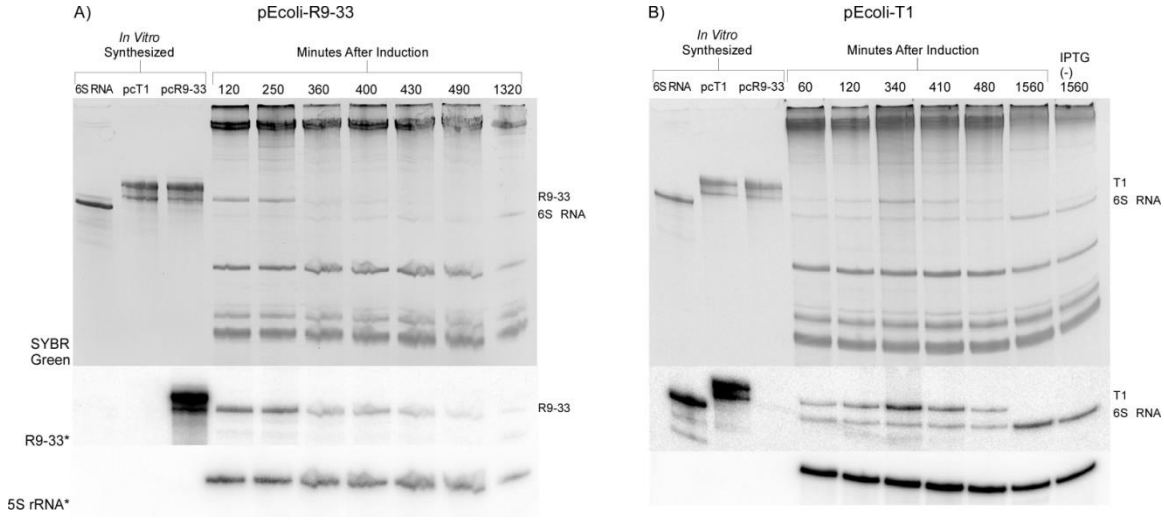



***Figure 2.15. Northern analysis of in vivo mutant 6S RNA expression***

A) Total RNA was extracted from *E. coli* pEcoli-R9-33 cells under denaturing conditions. Five $OD_{600}$ units were taken starting at 60 min of incubation in LB+Amp with IPTG and at intervals until 1,320 min (22 hr). An 8% denaturing PAGE imaged by SYBR Green is shown on the top panels. On the left, *in vitro* synthesized 6S RNA was loaded alongside *in vitro* transcribed T1 and R9-33 RNA [using plasmid cut (pc) with Cla I as DNA template] that served as a control for probe specificity and also as size reference for plasmid-derived RNA. The nucleic acid in this gel was transferred onto a membrane and probed with radiolabeled R9-33 specific probe (Middle panel, Table S2). B) Total RNA was extracted from BL21 pEcoli-T1 cells using previously described conditions. The same positive controls were loaded in this gel along with a sample extracted from BL21 pEcoli-T1 cells grown in absence of IPTG. In all cases, hybridization to 5S rRNA was used as an internal loading control (bottom panels).

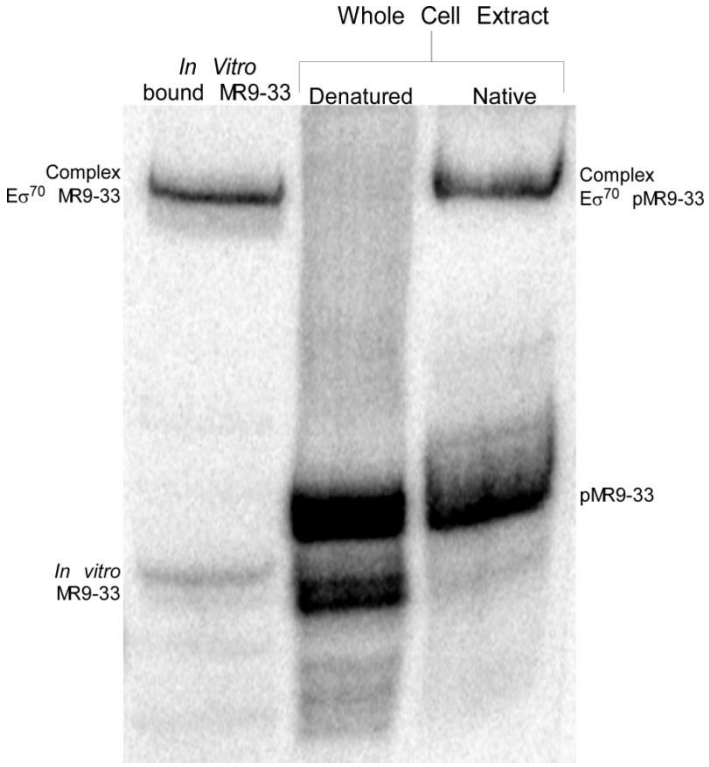



***Figure 2.16. In vivo plasmid-derived RNA forms a shifted complex consistent with RNA polymerase binding***

Native cell extracts were prepared from BL21 (DE3) pEcoli-R9-33 cells. Five $OD_{600}$ units were taken after 180 min of incubation with 5 mM IPTG. The RNA was run into a 5% native gel, blotted and then hybridized with a R9-33 RNA probe. The first lane on the left is a positive control, *in vitro* transcribed RNA bound to commercially available $E\sigma^{70}$ (see Methods). The middle lane is cell extract mixed with denaturing loading dye and heated for 3 min at 99°C. The third lane is the cell extract with native loading dye.

## 2.5. Conclusions

The release defective mutants isolated from a high-diversity RNA pool could be sorted into three classes, each with a distinct pRNA synthesis profile. The different classes of release-defective mutants appear to correspond to stalls in different stages of a previously modeled 6S RNA release process (Panchapakesan and Unrau 2012). In class I mutants the accumulation of substitutions around the '-10' and TSS regions allows the synthesis of short pRNAs while suppressing the synthesis of longer pRNAs that appear to be required for wild type 6S RNA release. Based on the 'scrunching' model shown in Figure 2.1, longer pRNAs are thought to be required after binding to transition successfully from State 2 (S2) to State 3 (S3). The class II (R9-24) mutant

further supports this idea. In contrast to the class I mutants, the class II mutant makes a very long pRNA, while also failing to release. We speculate that template slippage prevents the accumulation of mechanical force required to scrunch the RNA into the $E\sigma^{70}$ complex, and would in our interpretation also correspond to an inability to transition from the S2 to the S3 state (Figure 2.1). Slippage with shifts of -1 or -3 nucleotides have previously been shown to help stabilize the DNA:$E\sigma^{70}$ complex during transcriptional initiation (Borukhov, Sagitov *et al.* 1993, Liu and Martin 2009), so it is possible that a similar stabilization occurs in this 6S RNA mutant. The third class (III) of release-defective mutants make longer than expected pRNAs and have a phenotype consistent with a failure to leave either the intermediate scrunched state, or the S3 state formed after the release of $\sigma^{70}$ (Figure 2.13). It is hypothesized that the mutations in this group allow for stronger interactions between the RNA and the $E\sigma^{70}$, and that a longer than usual pRNA is required to create enough strain in the 6S: $E\sigma^{70}$ complex to trigger full 6S RNA release (Figure 2.17).

The presence of mutations that only partially release is consistent with a 6S RNA release model where intermediate structural states of either the 6S RNA or 6S:$E\sigma^{70}$ might become 'jammed' during pRNA synthesis. Interestingly, for the R9-33 mutant the 'jamming' was enhanced by the inclusion of upstream mutations (Figure 2.13, yellow region). Previously we have shown that the '-35' sequence region plays an essential role in initial RNA binding (Shephard, Dobson *et al.* 2010). Upstream sequence conservation showed a systematic shift towards the LOB in our selection (Figure 2.4) suggesting that mutant RNAs bind $E\sigma^{70}$ in a distinctly different fashion than wild type 6S RNA. A recent 6S RNA model docked to RNA polymerase based on the crystal structure of the bound open form DNA:$E\sigma^{70}$ (Murakami and Darst 2003, Steuten, Setny *et al.* 2013) suggests that upon initial binding, the 6S RNA interacts not only with $\sigma^{70}_{4.2}$, but also with $\sigma^{70}_{2.1}$, $\sigma^{70}_{2.3}$, $\sigma^{70}_{3.1}$ and $\sigma^{70}_{3.2}$. This model would predict that 8 out of 10 mutations (G65A , C69U, C74U, G75A, G117A, C118U, C124U and A125U) located in the upstream region of R9-33 are in the immediate vicinity of $\sigma^{70}_{2.1}$ (closer to '-10' area) and $\sigma^{70}_{3.1}$ that stretches between the '-10'and '-35' regions (Steuten, Setny *et al.* 2013). If these mutations stabilize interactions between the 6S RNA upstream region and the distal $\sigma^{70}$ domains, they might serve to inhibit changes in $\sigma^{70}$ structure that we postulate are required to trigger normal 6S RNA release. This inhibition of structural change could

therefore lead to 'jamming' of a stochastically determined subset of mutant RNAs during the process of scrunching mediated release.

Mutations within the top strand LOB cause complex release defects that are indicative of RNA structural dynamics and enzyme kinetics that are only partially understood. Our model of 6S RNA release suggests that the formation of a 9-bp release hairpin in the top strand plays an important role in normal 6S RNA release (Panchapakesan and Unrau 2012). U135C and C140A mutations, which disrupt this helix, might therefore be expected to have an effect on release rate. We however could not detect the effect of either mutation on release rate and it was only when A141C was included in addition to these two mutations that an observable release defect was observed (Figure 2.13). Each mutation would be expected to create either a bulge (U135C or C140A) or a change to in the tri-loop sequence (A141C) and might therefore not significantly change the thermodynamics of release hairpin formation individually. Combined, these mutations might have three effects: The first, as previously mentioned being thermodynamic. A second, and hard to determine consequence of these mutations is that they may alter contacts formed transiently with $E\sigma^{70}$ so as to slow the mechanism of release. While this cannot be precluded, we note that the lack of sequence conservation observed in the top strand of the TSS in our phylogeny (Figure 2.4) implies that there is no unique sequence capable of achieving such a state of affairs. The third and we think the most significant factor, concerns the register of the right arm of the release hairpin that forms during pRNA synthesis. It is notable that R9-33 accumulates a pRNA that is ~6-nt long in contrast to either the T1 control or to Class III mutants which can rapidly synthesize pRNAs ≥ 13-nt long (Figure 2.9). After synthesizing a pRNA 6-nt long (i.e. with sequence 5`-UUCGGC), R9-33 nucleotides on the top strand that were originally paired to the bottom template strand are now free to start forming the release hairpin by pairing with their reverse complements found in the top strand '-10' region. The C140A and A141C mutations seem ideally located to shift this pairing so as to make the formation of a full release hairpin nearly impossible: A141C can now pair potentially with G145, which in this new register would favour the pairing of C140A with U146, at which point further stem formation would be strongly disfavoured as C139 would not pair with U147. Based on our data, we therefore favour

a model where top strand LOB mutations found in R9-33 markedly interfere with the formation of the release hairpin and hence serve to help delay mutant 6S RNA release.

The mutations near the TSS on the bottom strand of the LOB are more challenging to explain. Not surprisingly U44A causes a marked decrease in transcriptional initiation efficiency as has been noted for many DNA TSS, and appears most likely to be a consequence of the enzyme active site favouring initiation with ribopurines (Revyakin, Liu *et al.* 2006). How adjacent mutations close to the TSS can abolish this effect remains to be explained (Figure 2.13E), but it is striking that such mutations when combined with top strand mutations are sufficient to produce a strong release defect (Figure 2.13B & C squares). These findings complement the work of Cabrera-Ostertag et al., where variation of the TSS regions was previously explored (Cabrera-Ostertag, Cavanagh *et al.* 2013).

*E. coli* cells expressing R9-33 confirmed its potent regulatory ability previously observed *in vitro,* by delaying entry into exponential growth and preventing colony growth. Although strong overexpression is not the ideal system to test 6S RNA mutants *in vivo* since it allows many potentially conflicting variables during late induction, at short times the high induction of mutant RNAs should be suggestive of the 6S RNAs ability to regulate transcription dynamic in high nutrient conditions. Our findings are in broad agreement with previous data on expression of mutant 6S RNA in *E. coli* and *B. subtilis* that prevented cells from re-entering active growth upon nutrient up shift as well as decreased cell viability (Cavanagh, Sperger *et al.* 2012). The marked cell death and inhibition of growth observed with overexpression of R9-33 and the decreasing trend in this respect seen with T1 RNA and control RNAs with even weaker binding interactions to $E\sigma^{70}$, indicates that even in high nutrient conditions, RNAs that interact with bacterial RNA polymerase can have a significant effect on growth dynamics.

The large range of release defective sequences found in this study implies that many sequences close to the 6S RNA sequence are avoided by natural selection because they do not exhibit correct pRNA induced release from $E\sigma^{70}$ and can like the R9-33 submutants characterized in this study, become ‘jammed’ during release. Interactions between the top and bottom regions of the LOB and with upstream

sequence 6S RNA sequence is strongly implicated by the study of such 'jammed' mutants and how such interactions feature in normal 6S RNA release is still an open question. Further study of different classes of 'jammed' states appears likely to improve our understanding of both 6S RNA release and nucleic acid scrunching in bacterial RNA polymerases.

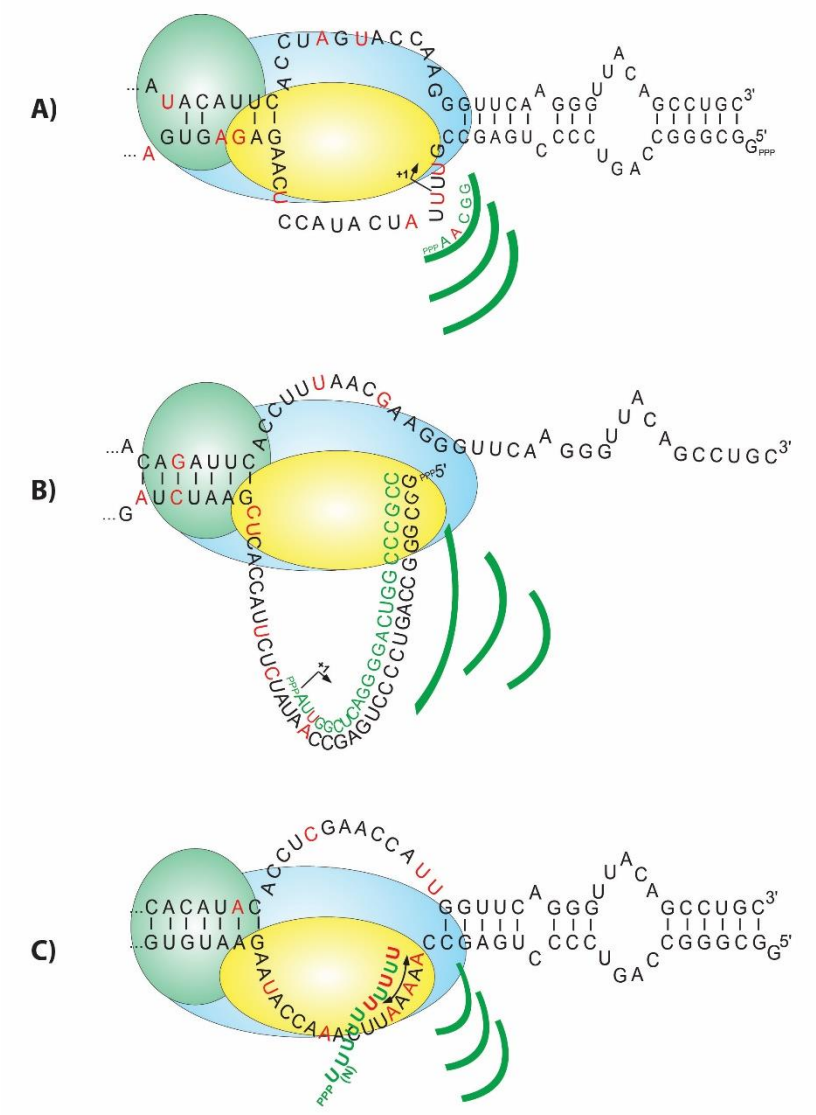



***Figure 2.17. Schematic representation of release defects occurring in RNA variants from Round 9 selection***

A) Class I mutants (e.g. MR9-66) accumulate mutations around the '-10' and TSS area allowing for synthesis of short pRNA while suppressing the synthesis of longer RNAs. B) Class III (e.g. MR9-33) mutants make longer than usual pRNA, perhaps as a consequence of stronger interactions between the RNA and $E\sigma^{70}$. C) Class II mutant (e.g. MR9-24) makes very long RNA as a consequence of template slippage.

### *Table 2-1. Statistical analysis of the T1 mutant clones sequenced*

Expected mutation frequency 0.033. Mutations at each residue were quantified and the calculated frequency of mutation for each residue was compared to the expected mutational frequency (0.033) expected from the pool design. In the last columns to the right, residues were tagged with three stars (***) as being very interesting since the observed frequencies diverged largely from the null result of 0.033. On the "6S RNA residue" column, some residues are followed by a letter "F" (Fixed) that indicates that the nucleotide was left unchanged during pool design.

| 6S RNA Residue | | | | | | | Found Frequency for each nt | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Posit** | **Wt** | **A** | **T** | **G** | **C** | **Total** | **A** | **T** | **G** | **C** | |
| 42 | G | 26 | 1 | | 1 | 28 | 0.366 | 0.014 | 0.000 | 0.014 | *** |
| 43 | A | | 4 | 1 | | 5 | 0.000 | 0.056 | 0.014 | 0.000 | |
| 44 | U | 26 | | 2 | | 28 | 0.366 | 0.000 | 0.028 | 0.000 | *** |
| 45 | A | | 19 | | 8 | 27 | 0.000 | 0.268 | 0.000 | 0.113 | *** |
| 46 | U | 6 | | | 4 | 10 | 0.085 | 0.000 | 0.000 | 0.056 | * |
| 47 | U | 2 | | 8 | 6 | 16 | 0.028 | 0.000 | 0.113 | 0.085 | ** |
| 48 | U | | | 3 | 1 | 4 | 0.000 | 0.000 | 0.042 | 0.014 | |
| 49 | C | 4 | 1 | 10 | | 15 | 0.056 | 0.014 | 0.141 | 0.000 | ** |
| 50 | A | | 5 | 6 | | 11 | 0.000 | 0.070 | 0.085 | 0.000 | * |
| 51 | U | 6 | | 1 | 4 | 11 | 0.085 | 0.000 | 0.014 | 0.056 | |
| 52 | A | | 6 | 1 | 6 | 13 | 0.000 | 0.085 | 0.014 | 0.085 | |
| 53 | C | 4 | 3 | 4 | | 11 | 0.056 | 0.042 | 0.056 | 0.000 | |
| 54 | C | 1 | 2 | 2 | | 5 | 0.014 | 0.028 | 0.028 | 0.000 | |
| 55 | A | | 2 | 4 | 3 | 9 | 0.000 | 0.028 | 0.056 | 0.042 | |
| 56 | C | 6 | 5 | 1 | | 12 | 0.085 | 0.070 | 0.014 | 0.000 | |
| 57 | A | | 4 | 1 | 5 | 10 | 0.000 | 0.056 | 0.014 | 0.070 | |
| 58 | A | | 4 | 4 | 4 | 12 | 0.000 | 0.056 | 0.056 | 0.056 | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 59 | G | 5 | 7 | | 3 | 15 | 0.070 | 0.099 | 0.000 | 0.042 | |
| 60 | A | | 3 | 3 | 1 | 7 | 0.000 | 0.042 | 0.042 | 0.014 | |
| 61 | A | | 3 | 3 | 1 | 7 | 0.000 | 0.042 | 0.042 | 0.014 | |
| 62 | T | 8 | | 5 | 3 | 16 | 0.113 | 0.000 | 0.070 | 0.042 | ** |
| 63 | G | | 2 | | 1 | 3 | 0.000 | 0.028 | 0.000 | 0.014 | * |
| 64 | T | completely conserved | | | | 0 | 0.000 | 0.000 | 0.000 | 0.000 | ** |
| 65 | G | 8 | 3 | | 2 | 13 | 0.113 | 0.042 | 0.000 | 0.028 | |

| 6S RNA Residue | | | | | | | Found Frequency for each nt | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Posit** | **Wt** | **A** | **T** | **G** | **C** | **Total** | **A** | **T** | **G** | **C** | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 66 | G | 6 | 9 | | 6 | 21 | 0.085 | 0.127 | 0.000 | 0.085 | ** |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 67 | C | 4 | | 5 | | 9 | 0.056 | 0.000 | 0.070 | 0.000 | * |
| 68 | G | 3 | 4 | | | 7 | 0.042 | 0.056 | 0.000 | 0.000 | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 69 | C | 5 | 3 | 1 | | 9 | 0.070 | 0.042 | 0.014 | 0.000 | * |
| 70 | T | | | | 3 | 3 | 0.000 | 0.000 | 0.000 | 0.042 | * |
| 71 | C | 5 | 13 | | | 18 | 0.070 | 0.183 | 0.000 | 0.000 | ** |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 72 | C | 7 | 4 | 2 | | 13 | 0.099 | 0.056 | 0.028 | 0.000 | * |
| 73 | G | 2 | 1 | | | 3 | 0.028 | 0.014 | 0.000 | 0.000 | |
| 74 | C | 2 | 3 | | | 5 | 0.028 | 0.042 | 0.000 | 0.000 | |
| 75 | G | 11 | 6 | | 2 | 19 | 0.155 | 0.085 | 0.000 | 0.028 | ** |
| 76 | G | | 2 | | 2 | 4 | 0.000 | 0.028 | 0.000 | 0.028 | |
| 77 | U | 1 | | 1 | | 2 | 0.014 | 0.000 | 0.014 | 0.000 | * |
| 78 | U | 1 | | 1 | | 2 | 0.014 | 0.000 | 0.014 | 0.000 | * |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 79 | G | | | | 1 | 1 | 0.000 | 0.000 | 0.000 | 0.014 | ** |
| 80 | G | | 2 | | | 2 | 0.000 | 0.028 | 0.000 | 0.000 | * |
| 81 | U | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | ** |
| 82 | G | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | ** |
| 83 | A | | 8 | | 3 | 11 | 0.000 | 0.113 | 0.000 | 0.042 | * |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 84 | G | completely conserved | | | | 0 | 0.000 | 0.000 | 0.000 | 0.000 | ** |
| 85 | C | 3 | 1 | | | 4 | 0.042 | 0.014 | 0.000 | 0.000 | * |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 86 | A | completely conserved | | | | 0 | 0.000 | 0.000 | 0.000 | 0.000 | ** |
| 87 | U | 11 | | 2 | 8 | 21 | 0.155 | 0.000 | 0.028 | 0.113 | *** |
| 88 | G | 5 | 1 | | 1 | 7 | 0.070 | 0.014 | 0.000 | 0.014 | |

| 6S RNA Residue | | | | | | | Found Frequency for each nt | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Posit** | **Wt** | **A** | **T** | **G** | **C** | **Total** | **A** | **T** | **G** | **C** | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 89 | CF | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 90 | TF | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 91 | CF | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 92 | GF | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 93 | GF | 1 | | | | | 0.014 | 0.000 | 0.000 | 0.000 | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 94 | UF | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 95 | CF | | 3 | | | | 0.000 | 0.042 | 0.000 | 0.000 | * |
| 96 | CF | | 2 | | | | 0.000 | 0.028 | 0.000 | 0.000 | * |
| 97 | GF | 4 | 2 | | | | 0.056 | 0.028 | 0.000 | 0.000 | * |
| 98 | UF | 1 | | | 2 | | 0.014 | 0.000 | 0.000 | 0.028 | * |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 99 | CF | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 100 | CF | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 101 | GF | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 102 | AF | | | 3 | | | 0.000 | 0.000 | 0.042 | 0.000 | * |
| 103 | GF | completely conserved | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 104 | AF | | | 3 | | | 0.000 | 0.000 | 0.042 | 0.000 | * |
| 105 | AF | | | | 1 | | 0.000 | 0.000 | 0.000 | 0.014 | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 106 | G | 1 | 4 | | | 5 | 0.014 | 0.056 | 0.000 | 0.000 | * |
| 107 | C | completely conserved | | | | 0 | 0.000 | 0.000 | 0.000 | 0.000 | ** |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |

| 6S RNA Residue | | | | | | | Found Frequency for each nt | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Posit** | **Wt** | **A** | **T** | **G** | **C** | **Total** | **A** | **T** | **G** | **C** | |
| 108 | C | completely conserved | | | | 0 | 0.000 | 0.000 | 0.000 | 0.000 | ** |
| 109 | U | completely conserved | | | | 0 | 0.000 | 0.000 | 0.000 | 0.000 | |
| 110 | U | 13 | | 1 | | 14 | 0.183 | 0.000 | 0.014 | 0.000 | ** |
| 111 | A | completely conserved | | | | 0 | 0.000 | 0.000 | 0.000 | 0.000 | ** |
| 112 | A | completely conserved | | | | 0 | 0.000 | 0.000 | 0.000 | 0.000 | ** |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 113 | A | | 1 | 1 | | 2 | 0.000 | 0.014 | 0.014 | 0.000 | |
| 114 | A | | 1 | 3 | 2 | 6 | 0.000 | 0.014 | 0.042 | 0.028 | |
| 115 | C | 1 | 4 | 1 | | 6 | 0.014 | 0.056 | 0.014 | 0.000 | |
| 116 | U | 3 | | | 4 | 7 | 0.042 | 0.000 | 0.000 | 0.056 | * |
| 117 | G | 4 | 3 | | | 7 | 0.056 | 0.042 | 0.000 | 0.000 | |
| 118 | C | | 1 | | | 1 | 0.000 | 0.014 | 0.000 | 0.000 | ** |
| 119 | G | 3 | 5 | 1 | | 9 | 0.042 | 0.070 | 0.014 | 0.000 | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 120 | A | | 6 | 6 | | 12 | 0.000 | 0.085 | 0.085 | 0.000 | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 121 | C | 2 | 2 | 3 | | 7 | 0.028 | 0.028 | 0.042 | 0.000 | |
| 122 | G | 7 | 6 | 4 | | 17 | 0.099 | 0.085 | 0.056 | 0.000 | ** |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 123 | A | | 2 | 4 | 3 | 9 | 0.000 | 0.028 | 0.056 | 0.042 | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |
| 124 | C | | 4 | 3 | | 7 | 0.000 | 0.056 | 0.042 | 0.000 | |
| 125 | A | | 1 | 2 | | 3 | 0.000 | 0.014 | 0.028 | 0.000 | |
| 126 | C | | 2 | 2 | | 4 | 0.000 | 0.028 | 0.028 | 0.000 | |
| 127 | A | | 5 | 1 | 1 | 7 | 0.000 | 0.070 | 0.014 | 0.014 | |
| 128 | U | 6 | | 2 | | 8 | 0.085 | 0.000 | 0.028 | 0.000 | |
| 129 | U | 13 | | 5 | 1 | 19 | 0.183 | 0.000 | 0.070 | 0.014 | *** |
| 130 | C | 7 | 2 | 2 | | 11 | 0.099 | 0.028 | 0.028 | 0.000 | |
| | | | | | | | 0.000 | 0.000 | 0.000 | 0.000 | |

| 6S RNA Residue | | | | | | | Found Frequency for each nt | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Posit** | **Wt** | **A** | **T** | **G** | **C** | **Total** | **A** | **T** | **G** | **C** | |
| 131 | A | | 1 | 1 | | 2 | 0.000 | 0.014 | 0.014 | 0.000 | * |
| 132 | C | 3 | 1 | 3 | | 7 | 0.042 | 0.014 | 0.042 | 0.000 | |
| 133 | C | 4 | | 1 | | 5 | 0.056 | 0.000 | 0.014 | 0.000 | * |
| 134 | U | 2 | | | | 2 | 0.028 | 0.000 | 0.000 | 0.000 | * |
| 135 | U | 10 | | 2 | 4 | 16 | 0.141 | 0.000 | 0.028 | 0.056 | ** |
| 136 | G | 11 | 14 | | 8 | 33 | 0.155 | 0.197 | 0.000 | 0.113 | *** |
| 137 | A | | 8 | 3 | 4 | 15 | 0.000 | 0.113 | 0.042 | 0.056 | ** |
| 138 | A | | 7 | 3 | 6 | 16 | 0.000 | 0.099 | 0.042 | 0.085 | ** |
| 139 | C | 4 | 6 | 8 | | 18 | 0.056 | 0.085 | 0.113 | 0.000 | ** |
| 140 | C | 5 | 4 | 10 | | 19 | 0.070 | 0.056 | 0.141 | 0.000 | ** |
| 141 | A | | 6 | 5 | 2 | 13 | 0.000 | 0.085 | 0.070 | 0.028 | |
| 142 | A | | 5 | 2 | | 7 | 0.000 | 0.070 | 0.028 | 0.000 | |
| 143 | G | 1 | 1 | | 8 | 10 | 0.014 | 0.014 | 0.000 | 0.113 | |

***Table 2-2. DNA oligonucleotides used during PCR reactions, mutagenesis, and Northern blot hybridization***

| Primer Sequence | Modification Introduced | Primer Number | Template construct |
|---|---|---|---|
| GCA GGC TGT AAC CCT TGA ACC ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG | T1 Control | 21.48 42.12 | |
| GCA GGC TGT AAC CCT TGA ACC CTG TGT TCG AGG TGA ATG | A135G, G140T, T141G | 39.34 | T1 |
| GCA GGC TGT AAC CCT TGA ACC CTT GGT TCA AGG TGA ATG | G135A, T140G, G141T | 39.35 | R9-33 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG AAA TTT CTT TCC ACA AG | T44A, A50T, A52T | 59.5 | T1 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG ATA TTT CAT ACC ACA AG | A44T, T50A, T52A | 59.6 | R9-33 |
| 39.34 + 59.5 | | | |
| 39.35 + 59.6 | | | |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG AAA TTT CAT ACC ACA AG | T44A | 59.7 | T1 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG ATA TTT CTT ACC ACA AG | A50T | 59.8 | T1 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG ATA TTT CAT TCC ACA AG | A52T | 59.9 | T1 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG ATA TTT CTT TCC ACA AG | A44T | 59.10 | R9-33 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG AAA TTT CAT TCC ACA AG | T50A | 59.11 | R9-33 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG AAA TTT CTT ACC ACA AG | T52A | 59.12 | R9-33 |
| GCA GGC TGT AAC CCT TGA ACC CTT GGT TCG AGG TGA ATG | T135C | 39.36 | T1 |
| GCA GGC TGT AAC CCT TGA ACC CTT TGT TCA AGG TGA ATG | C140A | 39.37 | T1 |
| GCA GGC TGT AAC CCT TGA ACC CTG GGT TCA AGG TGA ATG | A141C | 39.38 | T1 |
| GCA GGC TGT AAC CCT TGA ACC CTT TGT TCG AGG TGA ATG | T135C + C140A | 39.39 | T1 |
| GCA GGC TGT AAC CCT TGA ACC CTG GGT TCG AGG TGA ATG | T135C + A141C | 39.40 | T1 |

| | | | |
|---|---|---|---|
| GCA GGC TGT AAC CCT CGA ACC CTT GGT TCG AGG TGA ATG | T135C + A149G | 39.41 | T1 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCC GAG CCG ATA TTT CAT ACC ACA AG + 39.41 | T135C+A149G +T36C | 59.13 | T1 |
| GCA GGC TGT AAC CCT TGA ACC CTT GGT TCG AGG TGA ATG | A149 G | 39.42 | T1 |
| 39.42 + 59.13 | A149G+ U36C | | T1 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG AAA TTT CTT ACC ACA AG | U44 A + A50 U | 59.14 | T1 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG ATA TTT CAT ACC ACA AG | A44 U + U50 A | 59.15 | R9-33 |
| GCA GGC TGT AAC CCT TGA ACC CTG GGT TCA AGG TGA ATG | T140G + G141T | 39.43 | T1+6/ R9-33 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG AAA TTT CAT ACC ACA AG | T50A + T52A | 59.16 | T1+6/ R9-33 |
| ttc taa tac gac tca cta tag GCG GGC CAG TCC CCT GAG CCG ATA TTT CTT ACC ACA AG | A44T + T52A | 59.17 | T1+6/ R9-33 |
| GCA GGC TGT AAC CCT TGA ACC CTT TGT TCG AGG TGA ATG | G141T | 39.44 | T1+6/ R9-33 |
| GCA GGC TGT AAC CCT TGA ACC CTG GGT TCG AGG TGA ATG | T140G | 39.45 | T1+6/ R9-33 |
| GCA GGC TGT AAC CCT TGA ACC CTT GGT TCA AGG TGA ATG | C135T | 39.46 | T1+6/ R9-33 |
| GCA GGC TGT AAC CCT TGA ACC CTG GGT TCA AGG TGA ATG | G135A + T140G | 39.47 | T1+6/ R9-33 |
| GCA GGC TGT AAC CCT TGA ACC CTT TGT TCA AGG TGA ATG | G135A + G141T | 39.48 | T1+6/ R9-33 |
| GCA GGC TGT AAC CCT CGA ACC CTT GGT TCA AGG TGA ATG | G135A + A149G | 39.49 | T1+6/ R9-33 |
| GCA GGC TGT AAC CCT CGA ACC CTT GGT TCG AGG TGA ATG | A149G | 39.50 | T1+6/ R9-33 |
| GAT GCC TGG CAG TTC CCT ACT CT | Nothern probe 5S rRNA | 23.43 | n/a |
| CTT GGT TCA AGG TGA ATG TG | Northern probe for T1/6S | 20.117 | n/a |
| AAC CCT GTG TTC GAG GT | Northern probe for R9-33 | 17.94 | n/a |
| CAC CTG TGG CGC CGG TGA AAT TAA TAC GAC TCA CTA TAG GGG AAT TGT GAG CGG ATA ACA ATT CCG CGG GCC AGT CCC CTG AGC C | Adds T7 promoter, lacO, SgrAI site | 85.4 | n/a |

| | | | |
|---|---|---|---|
| GAC AGC TTA TCA TCG ATT TTC AGC AAA AAA CCC CTC AAG ACC CGT TTA GAG GCC CCA AGG GGT TAT GCT AGC AGG CTG TAA CCC TTG AAC C | adds T7 terminator, Cla I site. | 91.1 | n/a |
| GAG CAT GAA AAG GCT GCA GTT TAT CTC ACG GTC GCT TTC GTT CTT TAG TAC ACC CAA TAC ctc gga cgg acc gag cat gct cac caa ccg cg | LowBinder | 92.1 | n/a |

***Table 2-3. Point mutations around the large bubble of T1 RNA act in a complex and highly synergistic way***

| Construct Name | 10 Upstream Substitutions | Position 135, 140, 141 | Position 44, 50, 52 | Release time (min)* | Release from Eσ$^{70}$ # | pRNA |
|---|---|---|---|---|---|---|
| T1 | No | UCA | UAA | 0.8 | Yes | Yes +++ |
| T1+6 | No | CAC | AUU | 30 | No | Yes + |
| T1+U44A +A50U+A52U | No | UCA | AUU | 1.5 | Yes | Yes +++ |
| T1+U135C +C140A +A141C | No | CAC | UAA | 21 | Yes | Yes + |
| T1+U135C | No | CCA | UAA | 1 | Yes | Yes ++ |
| T1+C140A | No | UAA | UAA | 0.9 | Yes | Yes ++ |
| T1+A141C | No | UCC | UAA | 0.4 | Yes | Yes + |
| T1+U44A | No | ACA | AAA | 4.2 | No | Yes + |
| T1+A50U | No | ACA | UUA | 2 | Yes | Yes ++ |
| T1+A52U | No | ACA | UAU | 0.3 | Yes | Yes ++ |
| R9-33 | Yes | CAC | AUU | 0.2 | No | No |
| R9_33+C135U+A140C + C141A | Yes | UCA | AUU | 6 | No | No |
| R9_33+A44U+U50A+ U52A | Yes | CAC | UAA | 1.2 | Yes | Yes ++ |
| R9_33-6 mutations in LOB | Yes | UCA | UAA | 8.2 | Yes | Yes + |

The six mutations found in the LOB of R9-33 were added to the T1 construct (44, 50, 52 bottom of LOB, 135, 150 & 141 top of LOB). In red, the nucleotide sequence for control T1 RNA.

*Time in min required for 50% of the mutant RNA bound to the Eσ$^{70}$, to be released.

# Full release is defined by at least 80% of the RNA recovering a fast mobility regardless of the presence of pRNA:6S RNA complex.

Expressing a release defective 6S RNA in *E.coli* temporarily slowed down cellular growth. However, cellular division recovered six hours post-IPTG induction, reaching almost wild type doubling rate. These observations in liquid media also coincided with the Northern blot analysis, which showed that the cellular growth rate increased as the expression level of the release defective 6S RNA decayed. Combining these results leads me to question what the fate of the 6S RNA molecules is when they are no longer needed to modulate $E\sigma^{70}$ activity.

Chapter 3 attempts to explore whether 6S RNA could have additional roles throughout cellular growth in *E. coli* and the mechanisms by which 6S RNA is degraded. From a technical point of view answering these questions was more difficult than anticipated. Although we did not determine definitive answers to these questions, preliminary evidence suggests that 6S RNA interacts with an RNase as observed in cell extracts. The work described in Chapter 3 has the potential to be a starting point for further research that help us delineate the life cycle of 6S RNA molecules in bacteria.

Chapter 3.

# Exploring additional regulatory roles for 6S RNA and potential degradation pathways in *E. coli*

## 3.1. Introduction

The work described in chapter 2 leads to the finding that the expression of plasmid-derived mutant 6S RNA in *E. coli* causes a transcriptional shutdown, ultimately stalling the cell division process. The slow down in the cellular growth rate is temporary however. The current working hypothesis is that attempting to bypass such an inhibition and resume exponential growth causes bacteria to induce the production of protein factors that bind to the excess mutant RNA and tags it for destruction. As a result, the core polymerase (E) is again free to bind to different sigma factors in order to resume transcription and cellular growth. Such factors would presumably also be at play during the normal 6S transcriptional regulation process, but are only clearly evident by the expression of release defective 6S RNA mutants.

The findings described in the previous chapter show that overexpressing MR9-33 and a series of milder release defective mutants in *E. coli* resulted in a delayed entry into the exponential phase and a decrease in cell survival that correlated with the severity of the *in vitro* phenotype for each of the constructs (Figure 2.16). These findings suggest that 6S RNA release rates can be fine-tuned by simple sequence modifications and that those changes to the RNA can have widespread effects on cell division rates, survival capacities and global transcriptional rates. Consequently, the release rate for the 6S RNA sequence found in nature has probably been selected because it has strong enough binding to the $E\sigma^{70}$ while also releasing fast enough that it can aid in controlling the transcriptional dynamics essential to normal cellular growth and viability (Panchapakesan and Unrau 2012, Oviedo Ovando, Shephard *et al.* 2014). Research on

6S RNA homologs in other γ-proteobacteria is only in the initial stages (Cavanagh, Sperger *et al.* 2012, Cavanagh and Wassarman 2013) and although a large amount of research has been done on 6S RNA and its action mechanisms, several interrogants pertinent to the binding release dynamics in *E. coli* still remain unidentified *(Panchapakesan and Unrau 2012, Oviedo Ovando, Shephard et al. 2014)*.

As mentioned in Chapter 1 and 2, we have a good understanding of the role that 6S RNA plays during the stationary phase of bacterial growth (Wassarman and Storz 2000, Barrick, Sudarsan *et al.* 2005, Wassarman and Saecker 2006). However, the mechanisms by which 6S RNA is turned over are still unclear. Certain proteins, such as Hfq, S1 and RNase III, have been found to interact *in vitro* with 6S RNA (Windbichler, von Pelchrzim *et al.* 2008) and might provide clues as to what the turnover mechanisms for 6S RNA could be. Learning more about the proteins that interact with 6S RNA might help to elucidate 6S RNA turnover as well as other functions of this interesting molecule. In the following paragraphs, different proteins have been found to interact with 6S RNA during affinity chromatography (Windbichler, von Pelchrzim *et al.* 2008) will be discussed in further detail in order to understand if an interaction between 6S RNA and either an RNA or a protein could explain the observed return to exponential growth (Oviedo Ovando, Shephard *et al.* 2014).

Most small RNAs in bacteria regulate the translation of mRNA by changing their secondary structure and altering their accessibility to the ribosome (Gottesman 2004). RNA molecules fold into a wide variety of structures, some of which will allow the molecules to carry out different specific functions. It is possible that 6S RNA executes additional activities in the cell following some of these principles and it is assumed that proteins assist RNAs in folding and also partner with them to form RNA-protein particles (Clancy 2014). Exploring RNA and protein interactions, Windbichler and colleagues recovered several proteins that interacted with streptotag 6S RNA during affinity chromatography. The proteins found were Hfq, Protein S1, RNA polymerase beta subunit, Elongation factor Tu, Polyribonucleotide nucleotidyltransferase, Poly (A) polymerase I, DNA binding protein HU-alpha (HU-2), and RNase III (Windbichler, von Pelchrzim *et al.* 2008). It is worth mentioning that since the protein extract from the logarithmic and stationary phase was used as a starting material, it would be

conceivable that some of the reported interactions do not occur *in vivo* due to compartmentalization or previous engagement with other cellular factors. In addition, given that the input material is crude protein extract, highly abundant proteins that bind to 6S RNA are not further verified as being specific interactions. The initial results described in the next section lead us to think that the additional proteins interacting with the 6S RNA were responsible for the cell resuming exponential growth after expressing plasmid-derived mutant 6S RNA sequences. Although the possibility that 6S RNA interacts with other proteins has not been discarded, the process of identifying them will be more challenging than previously expected.

Some of these proteins are not likely to be involved in "real" interactions and are just an artifact of the experimental procedure. DNA binding protein HU-alpha (HU-2), a Histone-like DNA-binding protein, is one such protein. It probably bound 6S RNA because it is a small, highly abundant basic protein (Durrenberger, Bjornsti *et al.* 1988, Jaffe, Vinella *et al.* 1997). Perhaps HU-2 was purified with 6S RNA as a consequence of a protein-protein interaction with the RNA polymerase Beta subunit, rather than the result of a specific interaction between HU-2 and 6S RNA. The ribosomal protein S1, which is present in *E. coli* and plays a role in translation regulation, is another protein that might fit in this category (Sorensen, Fricke *et al.* 1998). Although it is related to a wide variety of regulatory events in the cell, Windbichler and colleagues found that it only binds weakly to 6S RNA and no further evidence has been found of it binding to 6S RNA (Windbichler, von Pelchrzim *et al.* 2008, Duval, Korepanov *et al.* 2013). S1 has no sequence specificity and favours binding to A/U rich single-stranded RNA regions (Hajnsdorf and Boni 2012). This affinity could explain the binding to 6S RNA, given that 6S RNA LOB contains A/U rich sequence elements. EF-Tu is one of the most abundant proteins in *E. coli* and is another example of proteins that are captured simply because of their abundance and not necessarily due to specific interactions with 6S RNA. However, translation regulators should not be automatically discarded as potential candidates to interact with 6S RNA. According to Panchapakesan, the regulation of the rate of transition from SC (6S pRNA-E complex) to S4 (6S-pRNA) might ultimately be linked to translational regulation, given that transcription and translation are tightly coupled events in bacteria. The conversion from the S3 state to the S4 state (Figure 2.1) is dependent on GTP concentration and GTP hydrolysis is intimately ligated to the

EF-Tu dependent ribosomal translation-elongation (Panchapakesan and Unrau 2012). The complex that EF-Tu is part of is especially important in the transition between growth phases and in stress conditions (Venkataramanan, Jones *et al.* 2013). Given 6S RNA's role in stress conditions, it is possible that the activities of 6S RNA and EF-Tu might be interconnected. Poly A polymerase, protein S1, and Hfq have been co-purified in the past with the β subunit of the RNA polymerase. The above mentioned complex was isolated not only with 6S RNA, but also with DsrA, MicF and OxyS RNAs (Windbichler, von Pelchrzim *et al.* 2008). Consequently, it is possible that these proteins were purified as the result of protein-protein interactions. In the complex that has been previously isolated, protein S1 associates *in vitro* with RNAP and then Hfq associates with the complex via S1. The biological role of the Poly A polymerase, protein S1, and Hfq complex can have in the cell as well as the nature of the biochemical reactions that are coupled remain unclear (Sukhodolets and Garges 2003, Arluison, Mutyam *et al.* 2007).

Hfq is a protein that is worth taking a second look at due to its close participation in the degradosome breaking down RNAs in the cell. This protein has no sequence specificity and favours binding to A/U-rich single-stranded RNA regions *(Le Derout, Folichon et al. 2003)*. As previously mentioned, this relatively small protein (11 KDa monomer) participates in several cellular mechanisms. It globally affects the physiology of the cell (Sobrero and Valverde 2012) and acts on various cellular mechanisms, thereby affecting several functions through various processes such as growth rate, reduced stress tolerance, and attenuated virulence *(Le Derout, Folichon et al. 2003, Sobrero and Valverde 2012)*. Hfq is interesting as it has been identified as a minor component of the degradasome. When associated with RNase E, Hfq acts as a chaperon involved with small regulatory RNA-mediated cleavage or the gene-silencing of specific transcripts (Burger, Whiteley *et al.* 2011). The hybrids mRNA-sRNA are rapidly degraded by the Hfq-RNase E interaction. Ongoing research on the degradosome shows that much more work needs to be done in order to understand this highly dynamic structure (Burger, Whiteley *et al.* 2011). It is discouraging to note that two independent transcriptome-wide studies failed to identify 6S RNA as one of the factors binding to Hfq. One of the studies used co-immunoprecipitation with Hfq and direct detection of the bound RNAs on genomic microarrays to identify members of the

small RNA family (Zhang, Wassarman *et al.* 2003). The second study used Genomic SELEX, a method for identifying protein-binding RNAs encoded in the genome, in order to search for further regulatory RNAs in *Escherichia coli* using Hfq as bait (Lorenz, Gesell *et al.* 2010).

According to Windbichler and colleagues (2008), RNase III or Ribonuclease III – an enzyme that binds and specifically cleaves double stranded RNA – is the last protein found to bind to 6S RNA. RNase III is a 52 KDa homodimer that is strongly $Mg^{+2}$ dependent, exhibits phosphodiesterase activity (Nicholson 1999), and is primarily involved in the maturation of ribosomal RNAs precursors (Robertson, Webster *et al.* 1968, Nashimoto and Uchida 1985). RNase III cleavage is also required for the proper maturation of the 5`-end of tRNAs (Nicholson 1999). Although RNase III has not been traditionally involved in the work of the degradosome, there seems to be a considerable level of redundancy and/or cooperation between RNase E and RNase III activity. Stead and colleagues made this observation in a study that used microarrays to observe the transcriptome changes on RNase E and RNase III deficient bacteria (Stead, Marshburn *et al.* 2011). RNase III can initiate mRNA degradation by cleaving the 5' untranslated region of its own mRNA, which is subsequently degraded by an RNase E dependent pathway that down-regulates RNase III production (Bardwell, Regnier *et al.* 1989). RNase III also controls the mRNA translation of other bacterial and viral mRNAs (Nicholson 1999). This small protein, which is active as a homodimer, seems to be a suitable candidate for interaction with 6S RNA.

RNase E could also be a candidate for binding to 6S RNA in order to degrade it. This RNase is the backbone for the process in which major and minor components of the degradosome associate to form a complex. The degradosome is a large multi-protein complex that has the following major components: RNase E, PNPase (polynucleotide phosphorylase), RhlB helicase and enolase (Burger, Whiteley *et al.* 2011). However, the current understanding of the degradosome system indicates that RNase E works with Hfq, which chaperons small regulatory RNA-mediated cleavage or gene-silencing of specific transcripts. As previously mentioned, there is no evidence that Hfq interacts with 6S RNA in *E. coli* (Zhang, Wassarman *et al.* 2003, Lorenz, Gesell *et al.* 2010). Finally, some of the results presented below support the hypothesis of an

RNase interaction with the 6S RNA, although the specific nature of this interaction has not been identified.

It appears that 6S RNA could participate in at least two different routes, in addition to engaging with the well-described interaction with $E\sigma^{70}$ and playing a role in transcription regulation during stationary phase. It is possible that 6S RNA is being tagged for destruction via the degradosome or via the RNase III. 6S RNA would need to interact with the PNPase or directly interact with RNase E in order to be broken down via the degradosome. Assuming that the interactions with S1, the elongation factor Tu or Poly (A) polymerase, are "real" 6S RNA-protein interactions, it is also possible that 6S RNA could be related to translation regulation**.** It is plausible that free 6S RNA contributes to the down-regulation of genes related to adaptations to stress and survival when environmental conditions change (like the ones observed during stationary phase). The 6S RNA molecules could also be contributing to translation regulation during early growth stages by up regulating genes related to active cellular growth and cell division. Although our findings are by no means conclusive, they seem to support the idea that 6S RNA is interacting with an RNase and perhaps Superoxide Dismutase. However, such an interaction needs to be confirmed through independent experiments. The work described below provides a starting point for further understanding the additional activities carried out by the 6S RNA.

## 3.2. Materials and Methods

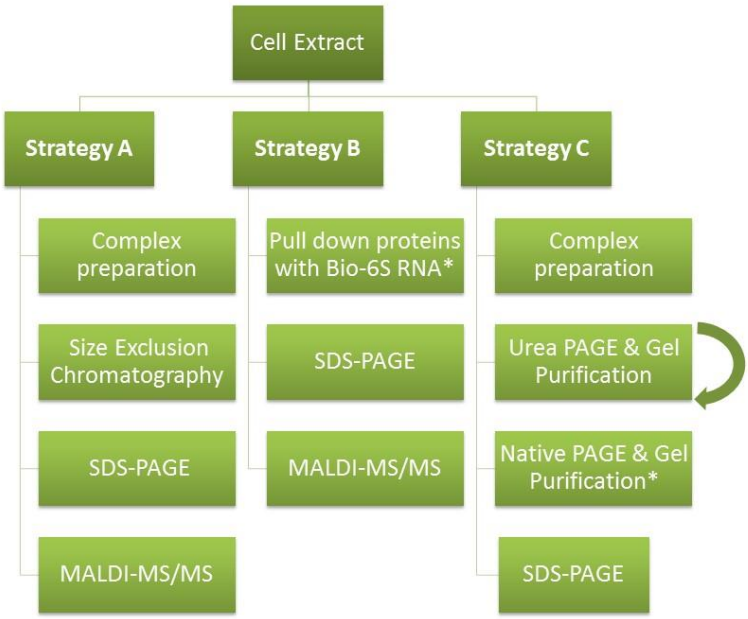



***Figure 3.1. Protein purification strategies used to attempt the isolation of proteins that interact with 6S RNA***

Summary of the purification strategies attempted during this chapter. The green arrow next to the box on the Strategy C path indicates that this step was repeated a couple of times in some repetitions of the experiment. However, that repetition did not notably contribute to improving the levels of purification

### 3.2.1. Cell extract and complex preparation

*E. coli* cells from 5 $OD_{600}$ units (or equivalent) were centrifuged at 5,000 g for 10 min. The pellet was resuspended in 200 µl of lysis buffer (20 mM Tris-HCl pH 8, 150 mM KCl, 1 mM $MgCl_2$, and 1 mM DTT) and mixed with an equivalent of 200 µl of 0.1 mm acid-washed (concentrated HCl) glass beads before vortexing for 6 min (alternating between 30 seconds of vortex and 30 seconds of cooling on ice) . An additional 200 µl of lysis buffer was added and samples were centrifuged at 16,500 g for 20 min at 4°C. The cleared extract was aliquoted, flash frozen, and stored at -80°C prior to use (Wassarman and Storz 2000). Preliminary tests showed that freshly prepared cell extract and flash frozen extract behaved in the same manner.

Radiolabeled 6S RNA was incubated with crude cell extract at a ratio of 9 pmol of RNA/ 50 µl of cell extract and a temperature of 37°C for 30 min in the presence of RNase Inhibitor (Roche). In most cases, the samples were quenched with 2X denaturing loading dye (0.025% Bromophenol Blue and 0.025% Xylene Cyanol, 50% Glycerol, 40 mM HEPES pH 7.5, 120 mM KCl and 8 mM EDTA) and heat denatured at 99°C for 3 min before loading in a denaturing gel.

### 3.2.2. *In vitro* 6S RNA transcription

An *in vitro* reaction was set up using a 6S RNA DNA template along with 40 mM Tris (pH 7.9), 5 mM DTT, 2.5 mM Spermidine, 25 mM $MgCl_2$, 0.01% Triton X-100, 8 mM GTP, 5 mM ATP, 5 mM CTP, 2 mM UTP, template DNA, and 5 U/µL T7 RNA polymerase (with or without 10.0 µCi/µL [$\alpha$-$^{32}$P]-UTP). Reactions were incubated for 1 hour at 37°C and quenched by the addition of one volume of: 90% Formamide, 50 mM EDTA (pH 8.0), 0.025% Xylene Cyanol, and 0.025% Bromophenol blue. Next, the RNA was separated in a 8% denaturing PAGE and the full length 6S RNA band was excised, eluted O/N (300 mM NaCl, 4°C), and recovered by EtOH precipitation.

### 3.2.3. Biotinylation of 3' end RNA

Freshly synthesized RNA was used as the starting material in an amount that ranged from 1 to 100 pmol. Sodium Periodate ($NaIO_4$) was added to a final concentration of 100 mM and the mixture was incubated for 20 min at room temperature in the dark with shaking. This oxidation reaction should be carried out in the minimum volume possible and the solution should be prepared fresh each time the process is conducted**.** The $NaIO_4$ is neutralized with 200 mM ($_fC$) of freshly prepared Sodium Hypophosphite ($NaPO_4H_2$) solution. The reaction is incubated for 20 min at room temperature, precipitated with 0.3 M NaCl, 2.5 vol of EtOH and centrifuge at 16000 g, for 30 min at 4 °C. The pellet is resuspended in a minimum volume (10 µl) of freshly prepared 100mM Sodium Acetate (NaOAc, pH 4). EZ Link Biotin LC Hydrazide (Thermo Scientific PI-21340, dissolved in DMSO to 50 mM) is added to a final concentration of 10 mM and incubated at 37 °C for 120 min. The newly formed hydrozone bond is stabilized by incubation at 37 °C for 30 min with 100 mM Sodium Cyanoborohydride ($NaBH_3CN$).

The $NaBH_3CN$ is dissolved in acetonitrile (anhydrous) and prepared fresh every time. The reaction is brought to a final volume of 100 µl final and precipitated with 10X Butanol. The mix is centrifuged at 16,000 g, 4 °C, for 30 min. The pellet is resuspended in 400 µl $ddH_2O$ and the unincorporated biotin is removed with three consecutive Phenol extractions and one Chloroform extraction. Finally, samples are EtOH precipitated and resuspended in the desired volume (Rio 2010).

### 3.2.4. Protein pull down using 3' Biotinylated 6S RNA

Dynabeads® Magnetic beads (Life Technologies) were used to pull down the biotinylated 6S RNA. The supplier established that 50 µg or 5 µl of Dynabeads® bind to 1 pmol of ss oligonucleotide. Based on this information, 50 µl of Dynabeads® were used to pull down 9 pmol of 3' biotinylated 6S RNA

Dynabeads® were washed a total of three times with 1 vol of 0.1 M NaOH and 0.05 M NaCl for 5 min in a rotator in order to remove the storage buffer. The next two washes were conducted with 1 vol of wash buffer (140 mM KCl, 10 mM $NaH_2PO_4$ pH 7.2, 1 mM $MgCl_2$, 0.05% Tween-20), 0.1 vol of Heparin, and 0.01 vol of 100X BSA (Bovine serum albumin). Finally, the beads were equilibrated in 1 vol of Lysis buffer (20 mM Tris-HCl pH 8, 150 mM KCl, 1 mM $MgCl_2$, and 1 mM DTT), which was used to prepare the cell extract.

Cell extract spiked with RNase inhibitor was coupled with radiolabelled 3' biotinylated 6S RNA, as described in section 3.2.1. A proportional volume of Dynabeads® was gently resuspended and the mix was incubated for 10 min with constant rotation at room temperature. Two control reactions were set up in parallel, the first without cell extract and the second without Dynabeads®. The tubes were placed in a magnetic holder for 3-5 min in order for the beads to be pulled down to the bottom of the tube. The supernatant was removed and the pellet was washed three times with 1 vol of PBS, then resuspended in a 0.1 vol of PBS. Finally, 0.1% SDS was added and the tubes were boiled for 5 min. The reaction tube was placed in the magnet once again to pull down the beads and recover the supernatant from each tube. The products of the

PBS washes were saved and ran in parallel with all the samples in a SDS-PAGE, as described in the next section.

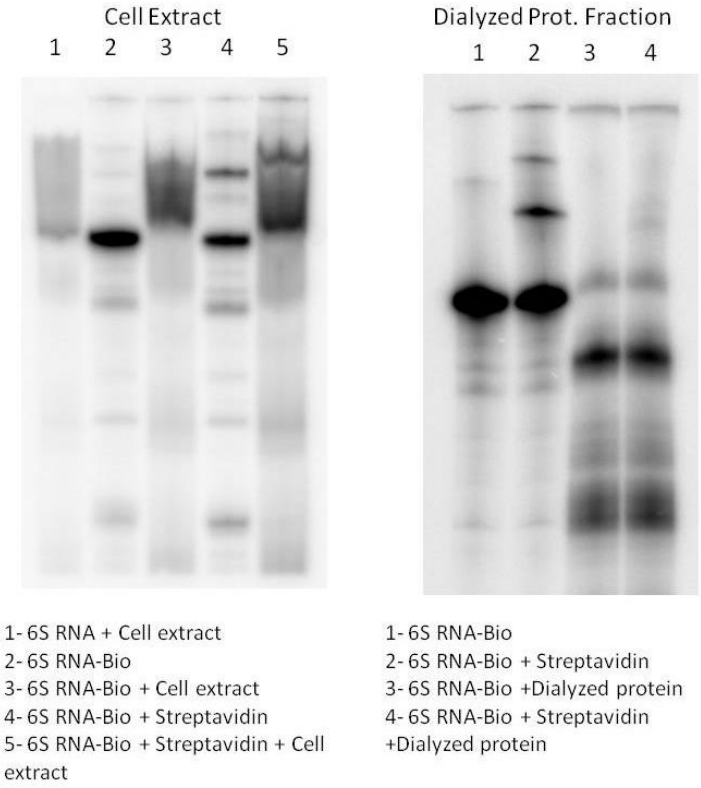



***Figure 3.2. Biotinylated 6S RNA produces a mobility shift when incubated with cell extract***

In average, the batches of Bio-6S RNA showed a percentage of Biotin incorporation of ~75%. The percentage of Biotin incorporation ranged between 70 – 90% throughout the different repetitions of the experiment. The biotinylated 6S RNA was tested with cell extract, and produces a mobility shift (left panel, lane1). The panel on the left, lane 5 shows that biotinylated 6S not only binds to cell extract but also to Streptavidin magnetic beads. The panel on the right, shows that when the biotinylated 6S RNA was tested against dialyzed protein fraction and Streptavidin magnetic beads showed a change in mobility that was different from the mobility observed in the panel on the left.

## 3.2.5. Size exclusion chromatography

A size exclusion column 45 cm long and 2.5 cm wide was packed with Superdex 200 (GE Healthcare Life Sciences). The flow rate used throughout our experiments was 1.0 ml/min with a buffer of 20 mM Tris-HCl pH 8, 150 mM KCl, 1 mM $MgCl_2$, and 1 mM DTT. Fractions were collected every 2.5 min. The column was sanitized with 0.5 M NaOH when necessary.

BSA (0.6 mg) was run as a standard and the amount of BSA in each fraction was determined by $Abs_{280nm}$. Freshly transcribed and gel-purified radiolabeled 6S RNA was the second standard to be loaded into the column and the elution profile was determined by using a scintillation counter to account for the amount of radiolabelled RNA in each fraction. Radiolabeled 6S RNA was *in vitro* bound to commercially available $E\sigma^{70}$ (Epicentre) – as was described in chapter 2 – and run through the column. Fractions were also quantified by using a scintillation counter.

### 3.2.6. SDS-PAGE

Denaturing separation was done by using SDS polyacrylamide mini gels, in accord with Laemmli's method. Samples were mixed with an equal volume of Laemmli loading buffer (4 % SDS, 20 % Glycerol, 10 % 2-mercaptoEtOH, 0,004 % Bromophenol Blue, and 0,125 M Tris), heated up to 100° C for 5 min, and cooled down on crushed ice. Samples were resolved using a Tris Glycine running buffer (10 X buffer: 250 mM Tris, 1.92 M Glycine, 1% SDS) at 4° C (Laemmli 1970).

### 3.2.7. Silver Staining and Sample Preparation for Mass Spectrometry

The samples submitted for MALDI-MS/MS analysis were cut from SDS-PAGE Silver stained gels. The gels were stained using Dodeca Silver stain Plus (Bio-Rad cat # 161-0449) as recommended by the kit**,** which can detect as little as 0.6 ng of protein. Each gel piece was submitted to Alphalyse (http://www.alphalyse.com/) in 1.7 ml microfuge tubes with $ddH_2O$ at room temperature, in accord with the Alphalyse website's recommendations.

### 3.2.8. Northern Blot

Aliquots of RNA (20 μg total RNA per lane for 8% denaturing gels) or cell extract (0.5 $OD_{600}$ units per lane for 5% native gels) were used to run Northern blots. Samples were blotted onto a Hybond N+ nylon membrane (GE). Membranes were UV cross-linked using a Stratalinker (1200 μJ for 30 seconds), and blocked, probed and washed

according to Krieg's method (Krieg 1996). *In vitro* transcribed RNAs were used to confirm the absence of cross reactivity for all the Northern blots.

5' end-labelled DNA probes were used for hybridization. The PNK (NEB) reactions were carried out as suggested by manufacturer and in some cases, the efficiency of [$\gamma$–$^{32}$P]-ATP incorporation was determined by running 0.5 µl of the PNK reaction on a 20% polyacrylamide/ urea gel. The efficiency of the labelling typically ranged from 80-95%, which meant that the probes had a high level of specific activity. The blotted membranes were placed in glass bottles, put in a Hybaid$^{TM}$ mini oven (MKII) at 65°C, and incubated in a minimal amount of hybridization solution with constant rotation for approximately four hours. The hybridization buffer was prepared with 6X SSPE (2L 10X solution: 173 g NaCl, 27.6 g $NaH_2PO_4$ $H_2O$, 7.4 g EDTA, pH 7 with NaOH), 1% SDS, 2X Denhart's solution, and 100 µg/ml of salmon sperm DNA (Rose, Winston *et al.* 1990). After pre-hybridization was complete, approximately 10 µCi of the labeled probe were added and the membrane was hybridized for at least 18 hours with the probes described in Table 2-2. Next, the membrane was washed twice with a high stringency solution (2X SSPE and 0.1%SDS), and twice again with a low stringency solution (0.2X SSPE and 0.1% SDS) for 15 min (at room temperature each time). The radioactive signal from the membranes was detected by using a Storm 820 Phosphorimager.

### 3.2.9. Proteinase K treatment

Proteinase K (NEB) was diluted to 20 mg/ml in a buffer (10 mM Tris-HCl, pH 8, 1mM EDTA, 0.5% SDS), according to the manufacturer's instructions. 2.5 µl of Proteinase K solution were added to 50 µl of cell extract and the mix was incubated at 37°C for 30 min. Complete protein degradation was verified by SDS-PAGE and followed by silver staining.

### 3.2.10. RNase A treatment

RNase A (Fermentas, Catalogue # 0531) was diluted to 10 µg/µl in 0.3 M NaCl, according to the manufacturer's recommendations. Add RNase A to a concentration of

1 µg/µl to the cell extract and incubate at 37°C for 30 min. RNA degradation was monitored by following the disappearance of radiolabelled 6S RNA in cell extract samples that were resolved in 8% Urea denaturing polyacrylamide gels.

## 3.3. Results

Working with *E. coli* that had been transformed with pEcoli-R9-33 revealed a correlation between the decay in the expression of plasmid-derived R9-33 RNA and the return to a wild type growth rate during growth in liquid media. As the growth rate increased in liquid media to eventually reach wild type values, the amount of plasmid-derived RNA decreased until it could no longer be detected in the Northern blots (Figure 2.16, Figure 2.15). These observations raised the question as to whether another factor (protein or RNA) is up-regulated independently of the type of mutant RNA being expressed and that 'sequesters' the defective plasmid-derived RNA, which in turn allows the cells to resume wild type growth. The starting assumption was that learning more about the proteins that interact with 6S RNA *in vivo* might help to elucidate 6S RNA turnover as well as other functions. However, isolating a discrete number of proteins that interact with 6S RNA proved to be more challenging than expected. Furthermore, the few proteins that were isolated and identified cannot be confirmed as 'real' RNA-Protein interactions with 6S RNA *in vivo* and they do not contribute to clarifying what the additional functions of 6S RNA are.

### 3.3.1. Factor(s) in the bacterial extract can tightly bind to radioactive 6S RNA

When crude cell extract was probed for 6S RNA or combined with radiolabelled 6S RNA, the mobility of the 6S RNA was slower than the mobility of *in vitro*-transcribed 6S RNA (Figure 2.16). The difference in mobility was even more obvious when *E. coli* pEcoli-R9-33 cells were tested against *in vitro* R9-33 (data not shown). However, we chose to focus on wild type *E. coli* as it would be easier to extrapolate the resulting findings to other γ-proteobacteria. In contrast, when radiolabelled 6S RNA was added and the complex was prepared in the manner described in Section 3.2.1 (quenched with 2X denaturing loading dye), then loaded in a native gel, it exhibited a faster mobility than

the free 6S RNA (Figure 3.3). In contrast, when the same complex was loaded in a denaturing gel, 6S RNA exhibited a lower mobility. However, in both cases, the 6S RNA recovered a mobility similar to that of free 6S RNA after phenol extraction. The fact that the change in mobility remained constant after an aggressive treatment that included a high concentration of Urea in the gel and harsh treatment of the sample (50% Formamide and heated 3 min at 99°C) suggests that the factor slowing down 6S RNA mobility – and potentially helping cells to resume wild type growth– is tightly bound to the 6S RNA. The change in mobility was also observed when cell extract was loaded in both native and denaturing gels and transferred onto a nylon membrane, on which it was

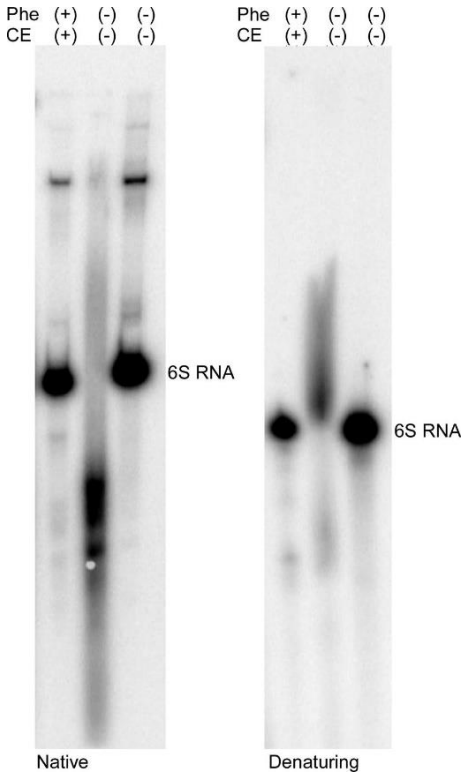



***Figure 3.3. 6S RNA incubated with crude cell extract shows different mobilities in native and denaturing gels***

Radiolabelled 6S RNA was incubated with (+) or without (-) cell extract (CE). The 6S RNA (+) cell extract was divided into two aliquots and one of them was subjected to phenol-chloroform extraction (Phe +). The same samples were loaded in an 8% denaturing gel (right panel) and a 5% native gel (left panel). In native gels, the mobility of the 6S RNA (+) was faster than the mobility of the free RNA. The sample that was phenol chloroform extracted was found to have intact 6S RNA and had the same mobility as the free 6S RNA in both types of gels. The 6S RNA (+) cell extract in the denaturing gel exhibited a slower mobility than the free RNA. In this case, the phenol-chloroform extraction made the RNA return to a mobility rate similar to that of the free 6S RNA.

probed for 6S RNA. Interestingly, a small variation in the speed of migration was detected in cell extract samples harvested at different $OD_{600}$ times (Figure 3.4A, top panel) and such difference was more evident when the samples were loaded under denaturing conditions (Figure 3.4A bottom panel). However, it was difficult to explain the differences in mobility that occurred when loading in native versus denaturing gels, given that they did not show a consistent pattern (the mobility pattern was highly reproducible).

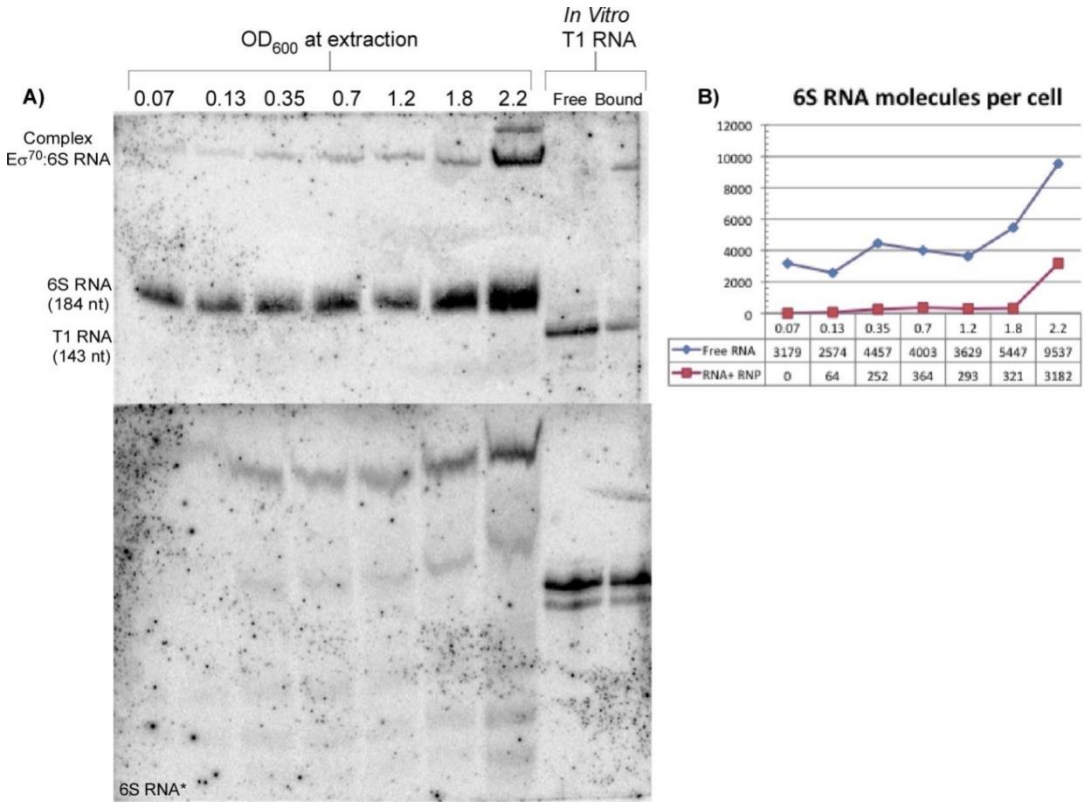



***Figure 3.4. Free 6S RNA is more abundant than the 6S-Eσ$^{70}$ complex in the E.coli cells***

Native cell extracts were prepared from *E. coli* BL21 (DE3) wild type cells. Five $OD_{600}$ units of cells were harvested at intervals between 2 and 24 hours after inoculation. A) Top panel: Equal amounts of RNA were run into a 5% native gel, blotted, and then hybridized with a 6S RNA radiolabeled probe. As part of the controls, *in vitro* synthesized T1 RNA (a truncated version of 6S RNA, 143 nt long) was incubated in the presence or absence of commercially available Eσ$^{70}$. On the bottom panel: The same samples in equal amounts were quenched with Formamide loading dye and heated for 3 min at 99°C, before being loaded into a 8% Urea-polyacrylamide gel. The gels were blotted and then hybridized with a 6S RNA radiolabeled probe. B) Plot of the number of RNA molecules, free (blue line) or in complex (red line), as a function of OD (X axis).

It is important to keep in mind that bringing nucleic acids, lipids, proteins and carbohydrates together by physically disrupting all cell membranes and compartments creates a very complex cell extract**.** This complexity might contribute to the difficulty of predicting how the sample will react to various chemicals, or to being separated by gel electrophoresis under different conditions. Another possibility that should be considered is that the observed changes in mobility are an artifact product of the interactions between the complex samples and the gel systems used to resolve them.

To further verify the previously observed change in mobility and to determine the nature of the putative factor binding 6S RNA, the samples were treated with agents that would disrupt the complex (Figure 3.5). I concluded that the putative factor binding 6S RNA is a protein because treatment of the cell extract with proteinase K or phenol, returns the mobility of the RNA to its free form (either plasmid-derived R9-33 or endogenous 6S RNA,). The "*in vitro"* lanes show a double band due to the effect of incomplete termination at the terminator site (fast mobility band). The main RNA band is the run off product transcription (slow mobility band). Because *E. coli* pEcoli-R9-33 cells also express the endogenous 6S RNA, after hybridization with probes for 6S RNA and R9-33 RNA, we observed two bands in the proteinase K lane. The Phenol and Extract lanes are shown on the right side of Figure 3.5. As expected, when the cell extract was treated with RNase A, none of the radiolabelled probes showed a radioactive signal.

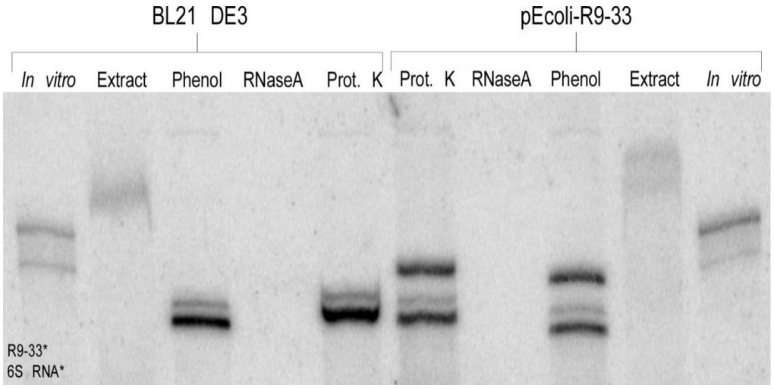



***Figure 3.5. Presence of proteins and phenol soluble substances changes 6S RNA mobility***

Removing proteins from cell extract speeds up RNA mobility. On the left, cell extract from wild type BL21 DE3 cells was subjected to phenol-chloroform extraction (Phenol lane), RNase A, or proteinase K treatment. On the right side, the cell extract from BL21pEcoliR9-33 cells was treated in the same way. Samples were quenched with Formamide loading dye and heated for 3 min at 99°C, then loaded into an 8% denaturing urea-polyacrylamide gel. The gel was blotted, hybridized with a R9-33 specific radiolabeled probe, and then hybridized again with a 6S RNA specific radiolabeled probe. As a result, the lanes on the left that contain the cell extract from BL21pEcoliR9-33 show both the plasmid-derived RNA (top bands) and endogenous 6S RNA (bottom bands). The *in vitro* lanes show double bands of plasmid-derived RNA because when linearized plasmids are used for *in vitro* transcription, incomplete termination is observed and produces the faster mobility band. The slower mobility band in the *In vitro* lanes is the product of run off transcription.

Additional experiments were designed to better understand the nature of the interaction between RNA and this hypothetical protein. One of the first tasks was to determine whether 6S RNA interacts specifically with a protein or proteins in the cell extract. A pool of random RNA sequences (94 nt long) was used in parallel to 6S RNA. The radiolabelled RNAs were combined with cell extract and treated with proteinase K, or were subject to phenol-chloroform extraction (Figure 3.6). By quantifying the intensity throughout each one of the lanes we learnt that only 6S RNA changes mobility when incubated with cell extract. While the shifted fraction constitutes 31% of the total count in the 6S RNA+Cell extract lane (Figure 3.6, third lane from the right), only 2% of the total counts are present in the Random RNA+Cell extract (Figure 3.6, lane on the far left). The above described data suggests that since we do not see a prominent size shift occurring with the random sequences, the interaction between the crude cell extract and the 6S RNA is specific. The lane 6S RNA*+Cell Extract shows an important percentage of bands with faster mobility. In light of our observation that several of these bands remained after being treated with proteinase K, it appears that they are actually

degradation products. Section 3.3.3 of this chapter will show that this is a recurring observation, which suggests that an RNase is possibly interacting or co-migrating with the 6S RNA in the cell extract.

The amount of counts present in Figure 3-5 below the full length RNA in both lanes lead us to interpret the faster mobility RNAs as degradation products, which suggests that both types of RNA are subject to degradation. The fact that we can see the presence of discrete bands in 6S RNA is suggestive of an organized and specific pattern of cleavage. The homogeneous smear observed below the full length 94 nt strongly suggests a random exonucleolytic attack to the RNA ends. Independently of what might be the cause for the degradation pattern in both samples, it is very clear that the cell extract reacts differently when coupled with 6S RNA. Future research should include the sequencing of the breakdown products that show faster mobility than 6S RNA, which would enhance our understanding the RNA cleavage pattern.

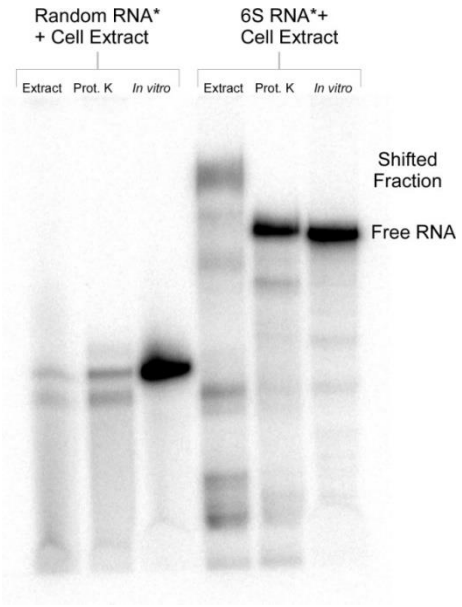



***Figure 3.6. 6S RNA-Cell extract interaction is specific***

Crude cell extract is incubated at 37°C for 30 min, with radiolabeled 6S RNA or 94 nt random sequence RNA (in the presence of RNase inhibitor). A prominent size shift (31% vs 2%) was only observed with 6S RNA. Furthermore, when samples are treated with Proteinase K (Prot. K lane) after binding to the radiolabeled RNA, they recover the fast mobility observed for the *in vitro* synthesized RNA in the absence of cell extract (*In vitro* lane). Samples were resolved in 8% denaturing gel.

It was verified that the bands with delayed mobility observed in the Northern blots, and the samples prepared by incubating the crude cell extract with radiolabeled 6S RNA (37°C for 30 min) show the same mobility in denaturing urea gel (Figure 3.7). This data lead us to assume that the 6S RNA of endogenous or *in vitro* origin, must be interacting with the same protein factor/s. In both cases when the samples were treated with proteinase K, the 6S RNA returned to a fast rate of mobility typical of naked 6S RNA (Figure 3.7, first lane on the right).

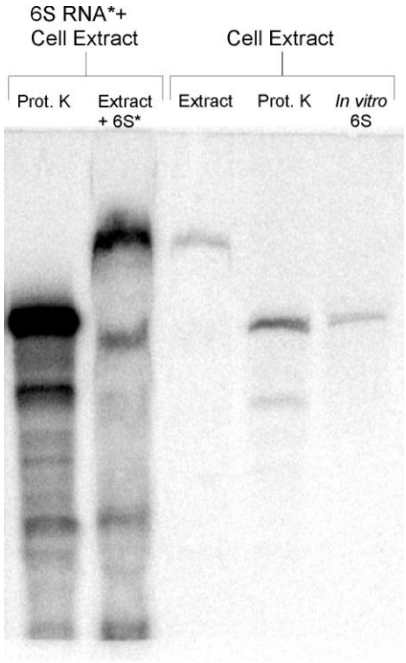



***Figure 3.7. The endogenous complex and the complex prepared in vitro have a similar mobility in an 8% denaturing gel***

The complex was prepared as described in section 3.2.1. Half of the sample was loaded directly while the other half was treated with Proteinase K. Crude cell extract was loaded in a 8% denaturing gel, before and after being treated with Proteinase K, alongside *in vitro* transcribed (cold) 6S RNA. The entire gel was blotted onto a nylon membrane and hybridized with a radiolabelled 6S RNA probe.

To better understand if the interaction between RNA and the protein is as strong as suggested by previous observations (Figure 3.3), we observed the mobility of the 6S RNA under different buffer conditions. The complex was prepared as previously described, buffer exchange columns (Amicon YM3) were used to dissolve the 6S RNA-CE complex in a couple of buffers: 1X TBE, 0.3M NaCl and the buffer for preparing the cell extract (20 mM Tris [pH 8], 150 mM KCl, 1 mM $MgCl_2$, and 1 mM DTT). The 6S

RNA showed different mobilities depending on the amount of salt present (Figure 3.8). Furthermore, when the slow mobility band was gel purified and run in a second denaturing gel, only free 6S RNA was observed (data not shown). It is apparent from these observations that the putative 6S RNA-protein complex does not survive either the elution of the first gel or the loading/running in a second gel. The present data indicates that this complex might be more transient and unstable than previously thought. Another possibility is that the changes in mobility observed are just a function of the gel running conditions, rather than a real protein-RNA interaction.

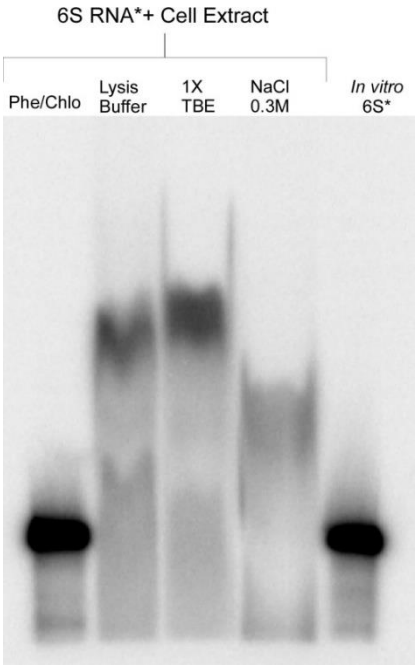



***Figure 3.8. The protein RNA complex shows changing mobility***

Crude cell extract was bound to the radiolabelled 6S RNA (6S RNA*+Cell Extract) and split into fractions that had different treatments. One fraction was subject to phenol chloroform extraction (Phe/Chlo) and the 6S RNA had the same mobility as the lane on the far right (In vitro 6S*), which contains only RNA**.** The next three lanes contained samples of the complex in different salt conditions (Lysis Buffer, 1X TBE, or NaCl 0.3M). Samples were resolved in a denaturing gel.

Efforts were also focused on trying to observe the putative complex under native gel conditions in order to determine whether the strong shift in 6S RNA mobility is a real physiological interaction or only a consequence of the stringent conditions that had been used. These harsh conditions might be forcing the generation of new interactions that would not occur under other physiological conditions. Native PAGE (no SDS or Urea in

gel or in running buffer) was used to run freshly prepared 6S RNA-complex, but the previously mentioned size shift was not observed (Figure not shown). Thus far, the mobility shift for 6S RNA has been seen only when the sample is resolved in a Urea gel and heat denatured with Formamide loading dye. The next step was therefore to elucidate if the Urea gel or the Formamide loading dye are solely responsible for the change in mobility. To that end, a native gel was loaded. The complex was prepared as previously described and run in 5% native gel with denaturing loading dye or with native loading dye (0.025% Bromophenol Blue and 0.025% Xylene Cyanol, 50% Glycerol, 40 mM HEPES pH 7.5, 120 mM KCl and 8mM EDTA). The native (Figure 3.9) gel shows that the CE+6S lane loaded with denaturing loading dye, has a faster mobility than naked 6S RNA in contrast to the slower mobility observed in denaturing gels. In contrast, when samples are resolved in the same native gel combined with native loading dye, only 3% of the 6S RNA show a faster mobility. Eighty-eight per cent remains unbound (runs side by side with 6S RNA Ctrl) and 6% has a mobility similar to the mobility shown for 6S RNA bound to purified $E\sigma^{70}$ (Figure 3.9). The 3% of the 6S RNA that showed a faster mobility in the "Native LD CE+6S" lane has the same band pattern as the "Denaturing LD CE+6S" lane, which implies that the Formamide denaturing loading dye is not responsible for the changes in 6S RNA mobility. The nature of the cell extract plus the Urea in the gel seems to be a more important determinant of RNA mobility.

The variety of conditions tested to confirm or discard the presence of the complex 6S RNA:protein leads us to claim with a degree of confidence that the interaction observed is not one that could be found in *E. coli*. The change in mobility for 6S RNA in presence of cell extract is most likely a consequence of the co-migration of 6S RNA being slowed down by the presence of a complex mix such as the crude cell extract.

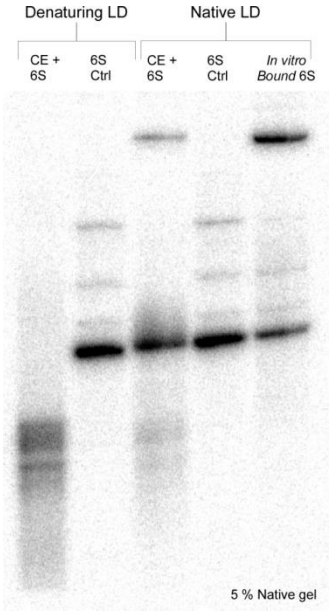



***Figure 3.9. The protein:RNA complex shows different mobility in Native gel depending on the loading dye and type of gel***

Radiolabelled 6S RNA was bound to the crude cell extract or to commercial *E. coli* RNA polymerase (*In vitro* Bound 6S lane), as described in the methods section. The samples were loaded in a 5% native after being mixed with either Denaturing loading dye (Denaturing LD) or Native loading dye (Native LD). The material on the *in vitro* bound 6S RNA lane (in absence of loading dye) was the input for the Superdex-200 column.

## 3.3.2. 6S RNA molecules are present in the cell throughout growth in liquid media

The number of 6S RNA molecules per cell was estimated by correlating an $OD_{600}$ with cell number and the intensity produced in the Northern blots by a known amount of *in vitro* transcribed 6S RNA (Figure 3.4B). The values obtained for free 6S RNA range from a few thousand molecules (3179) during the lag phase and are almost ten thousand (9537) during the late stationary phase. The number of 6S RNA-E$\sigma^{70}$ molecules went from zero in the lag phase to 321 during the early stationary phase (8 hr post inoculation) to 3182 (24 hr post inoculation). These estimations are in agreement with previous reports by Wassarman (Wassarman and Storz 2000) and Lee (Lee, Bailey *et al.* 1978). The fact that there is a large excess of endogenous 6S RNA that remains unbound in a faster mobility band (Figure 3.4A top panel, 6S RNA) at any given time throughout the growth process supports the idea that this molecule might have other potential roles during different stages of cellular growth (as suggested in the previous

chapter). Bacteria are very efficient as they can maximize the use of resources, so it is logical to assume that *E. coli* must have a function for the thousands of 6S RNA molecules that are produced throughout the life of a cell.

### 3.3.3. The attempts to purify a protein/s interacting with 6S RNA, failed to demonstrate a definitive protein partner

Given the complexity of the mix in the cell extract, a few multi-step approaches were tested to sequentially concentrate the different protein/s that might be in a complex with 6S RNA in the *E. coli* cell extract in order to try to isolate one or more of the proteins that might be interacting with 6S RNA.

#### *Strategy A. Size exclusion chromatography*

The first round of purification took advantage of the variety of MW proteins in the cell. A Superdex resin with a wide range of separation (10-600 KDa) was used to separate the proteins in the cell extract by size exclusion. In several independent experiments, BSA, 6S RNA, *in vitro* formed 6S RNA-E$\sigma^{70}$ complex, and cell extract combined with 6S RNA were run through the column in order to detect changes in mobility that might demonstrate that a small protein is binding the 6S RNA and consequently changing its elution time. The goal of these experiments was to determine exclusion times for 6S RNA and for 6S RNA in complex. Initially, 1 mg of BSA (66.5 KDa) and ~ 20 pmol of RNA (61 KDa and ~ 20 Kcpm) were loaded into the column. As expected, bigger molecules moved through the resin faster (Figure 3.10). Further analysis made it clear that there is a relatively large separation for the elution time of BSA and 6S RNA, given their very similar size. An explanation for this is that BSA is generating dimers or some level of homodimers that increase the total size of the BSA and affect its elution time.

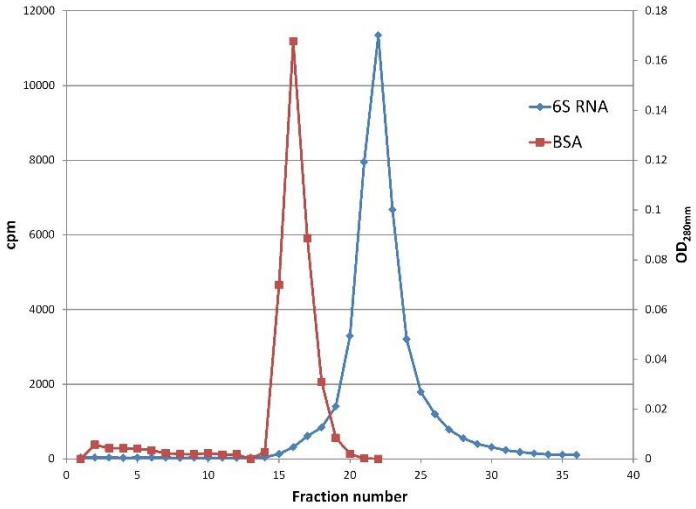



***Figure 3.10. Bovine Serum Albumin (BSA, 66.5 KDa) and radiolabelled 6S RNA (61 KDa) elution profile***

Data for the elution of BSA and RNA samples are shown. BSA fractions were collected and Absorbance was determined at 280 nm (left Y axis), while radioactive 6S RNA fractions were quantified using a scintillation counter (right Y axis). In the X axis the fraction number is indicated.

As mentioned in the previous paragraph, radiolabelled 6S RNA was bound to commercially available $E\sigma^{70}$ and run through the column (Figure 3.11). When the input sample was quantified in a polyacrylamide gel, and the area under each peak was integrated, the same proportions of RNA were observed in both experiments for free 6S RNA and the 6S RNA+$E\sigma^{70}$ complex. Several of the fractions collected were loaded onto a native gel and the same band pattern was observed. The predominant bands correspond to free 6S RNA and 6S RNA bound to $E\sigma^{70}$. These findings confirm that that the current experimental conditions favour the proper folding of 6S RNA. We assume that 6S RNA folds properly because it can interact with biologically active $E\sigma^{70}$. The approach described in this section should allow us to detect any other complexes that are formed between 6S RNA and other proteins in the cell extract under physiological conditions

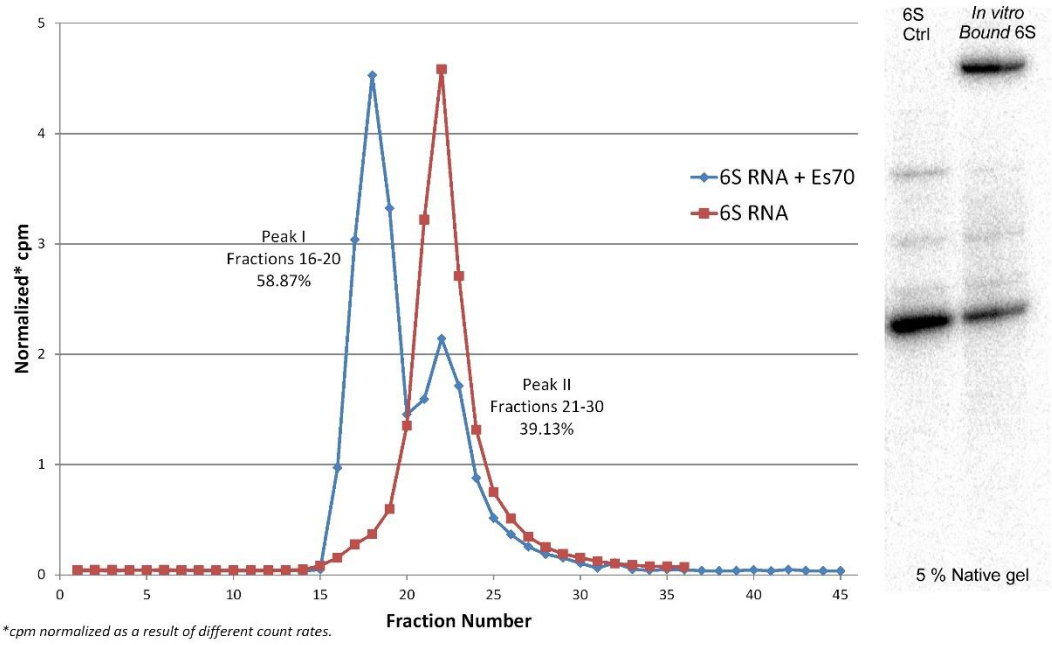



***Figure 3.11. Radiolabelled 6S RNA binds only RNA Polymerase Holoenzyme***

Elution profiles for 6S RNA (red) and *in vitro* formed 6S RNA+E$\sigma^{70}$ complex (blue). The areas under the blue curve were quantified and the percentages for each peak have been noted along with the fraction numbers that are considered to be forming each peak. The X axis indicates the fraction number. Given the differences of cpm observed when running the two samples, data was normalized to have them fit the same scale and facilitate a better comparison. The mean was subtracted from each value and then divided by the standard deviation. The panel on the right shows the input sample resolved in 5% native gel (*In vitro* bound 6S RNA). The main bands on the "*In vitro* bound" lane were quantified and it was determined that the top band has 59% of the counts, while the bottom band (which corresponds to the free 6S RNA) has 41% of the total counts. The fainter bands in between were not taken into account as we know they are 6S RNA molecules that bind to each other (Shephard, Dobson *et al.* 2010).

After establishing the retention time for the control samples loaded under native conditions, the radiolabelled 6S RNA was combined to the cell extract in order to observe its elution profile and determine the presence of additional 6S RNA protein complexes (Figure 3.12). The elution profile for the 6S RNA+CE sample showed two additional peaks around fractions 17 and 33 (Figure 3.12 red line) and the elution pattern was more similar to the band pattern observed when the input sample was loaded into a native gel (Figure 3.12B). Three main bands can be observed in the native gel and the percentage of counts in each one is similar to the percentages observed after quantifying the area under the green curve (Figure 3.12). This is to be expected, given that the buffer conditions in both systems are similar. A comparison with the elution

profiles used as standards (Figure 3.10 and Figure 3.11) revealed that the first peak was 6S RNA bound to $E\sigma^{70}$ (Figure 3.11), which corresponds to the top band on the native gel (Figure 3.12B). The second most prominent peak corresponds to free 6S RNA and is the most intense band observed in the native gel (Figure 3.12 B panel, Figure 3.6). The last peak, which emerged after fraction 30, showed to be degradation products. The putative 6S RNA-protein complex would be expected to elute between the fractions that correspond to 6S RNA and the 6S RNA-$E\sigma^{70}$ complex, because we estimate from previous size shift gels that a small protein is interacting with the 6S RNA. However, our size exclusion experiments do not show such a peak.

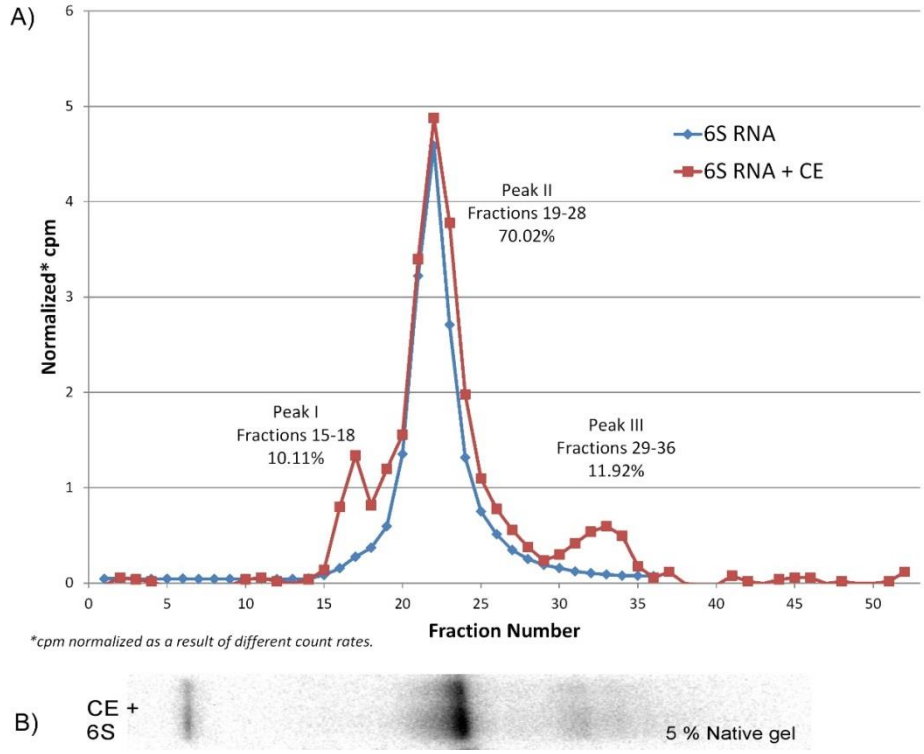



***Figure 3.12. The majority of radiolabelled 6S RNA eluted out in the same fraction independently of the presence of cell extract in the input mix***

A) Elution profiles for 6S RNA (blue) and 6S RNA incubated with cell extract (6S RNA+CE, in red). Normalized data for the counts present in each aliquot are described in the Y axis, while the X axis indicates the fraction number. B) Input sample for panel A) Superdex-200 column chromatography showing the slowest band that is 6S RNA+ $E\sigma^{70}$ that corresponds to Peak I, the middle band that corresponds to free 6S RNA and Peak II of the A panel and finally a faint fast mobility band that corresponds to Peak III of panel A.

The fractions from peak I (15 to 18), peak II (19-28), and peak III (29-36) produced during the run of 6S RNA+CE were concentrated using Centricon YM-3 centrifugal devices (Millipore), according to the manufacturer's instructions and loaded in a 15% SDS-PAGE. The fractions collected from peak II were divided into two lanes in the gel in order to improve the visualization these proteins. Peak I was determined to contain the 6S RNA-E$\sigma^{70}$ complex, which includes large proteins. Free 6S RNA (61 KDa) eluted in peak II ($II_A$ and $II_B$) and showed the large number of proteins that co-migrated with it. The abundance of proteins co-migrating with 6S RNA may be in part responsible for the slower mobility observed for the RNA under denaturing conditions in previous experiments (Figure 3.5, Figure 3.7). Peak III was the only peak that size exclusion data suggested might be a 6S RNA break down product. When the concentrated sample from peak III was run in a denaturing gel degradation of the 6S RNA was confirmed as very few counts were observed at the bottom of the gel. In an SDS-PAGE, Peak III showed a wide size range of proteins, although there was a smaller amount of proteins eluted in this peak than in Peak I and II. The larger proteins that also eluted in the other peaks are probably highly abundant cell proteins, but they were not considered interesting as they did not demonstrate enrichment. The smaller bands seen in the peak III lane demanded our attention because these bands have not been seen in previous experiments or showed evidence of enrichment.

Based on our previous runs with control samples in the size exclusion column, the identity of the two first peaks was clear. However, there was no certainty as to what the identity of the Peak III. Two of the bands of the Peak III fraction were cut and sent to be identified by Maldi-MS/MS (Figure 3.13, black arrows). The mass spectrometric peptide mapping and sequencing analysis identified the bottom band as peptidyl-prolyl *cis*-trans isomerase A (20 KDa, Score: 103, Sequence coverage 27%). The top band was identified as superoxide dismutase [Fe] (17.6 KDa, Score: 94, Sequence coverage 37%). Peptidyl-prolyl *cis-trans* isomerase A is a protein found in both prokaryotes and eukaryotes that interconverts *cis* and *trans*-isomers of the amino acid proline. Most amino acids are stabilized in the *trans*-conformation of the peptide bond. In the case of proline – and due to its cyclic structure with a side chain **–** the peptide bonds are better stabilized in the *cis* position (Fischer and Schmid 1990). Superoxide dismutase are metalloenzymes that play a central role in protecting organisms against super-oxide

radicals ($\bullet O_2^-$). The superoxide radical is the first intermediate of oxygen reduction. It can act as an oxidant or a reductant and is converted by dismutases to oxygen and peroxide after being deprotonated on a metal ion (Lavelle, McAdam *et al.* 1977, McAdam, Fox *et al.* 1977). A review of the literature did not help to identify a potential connection between any of the identified proteins and 6S RNA.

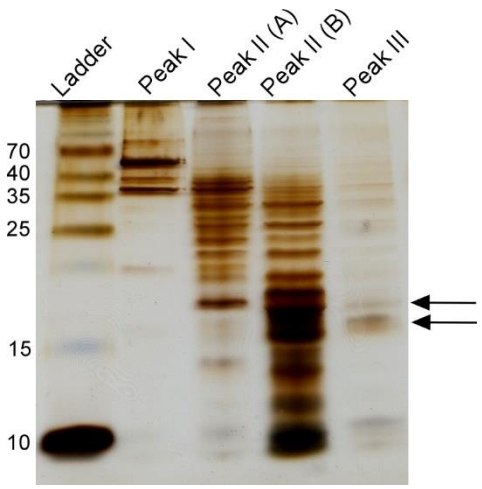



***Figure 3.13. Proteins eluted in each fraction according to their MW***

The concentrated proteins from Peaks I to III were loaded in a 15% SDS-PAGE and visualized by Silver Stain. All peaks show an abundance of proteins that are consistent with separation based on the size of the proteins. Given the large volume that remained after Peak II concentration, the fractions collected were divided into two lanes: Peaks $II_A$ (Fractions 19-23) and $II_B$ (Fractions 24-28). Since Peak I and II were clearly identified, small proteins from peak III were selected to be identified by mass spectrometry (Maldi-MS/MS). The number on the left indicates the size of each of the bands in the protein ladder.

When the 6S RNA+Cell extract (CE) complex was prepared (Section 3.2.1), small aliquots of the input material were loaded in 8% denaturing and native gels, while the remaining sample was loaded into the Superdex-200 column. In the denaturing gel, 78% of the 6S RNA signal shows a slower mobility (data not shown, but similar to the profile shown in Figure 3.7, second lane from the left). However, when the area under each curve of the elution profile for 6S RNA+Cell extract was quantified, the percentages for each peak did not match the proportions observed in the denaturing gel for the input same sample. The major peak had 70% of its counts eluted is the fraction where naked 6S RNA run in during control runs. (Figure 3.11). The previously described results confirm that the slower mobility band observed in the denaturing gels is an artifact due to the Formamide and Urea interacting with the 6S RNA+CE (as has been suggested for

earlier gel experiments) (Figure 3.3). The discrepancy between the percentage of counts between the input and the output suggests that under the current working conditions, 6S RNA is merely co-migrating with the proteins in the cell extract and not interacting specifically to produce a larger complex with faster mobility. When some of the eluted aliquots from the 6S RNA+ CE input were loaded in both native and denaturing gels (Figure 3 13), only one slower mobility band was observed in the native gel and that band corresponded to the 6S RNA bound to $E\sigma^{70}$. The peak after fraction 30 cannot be visualized in the gels due to the lower sensitivity because the gels have a higher detection limit than the scintillation counter. None of these gels showed further evidence of having captured 6S RNA bound in a complex with any protein other than $E\sigma^{70}$. Finally, the aliquots collected from the run of crude cell extract through the Superdex 200 column were reacted with radiolabelled 6S RNA (as described in the Methods section) and then loaded into a denaturing gel. Interestingly, the fraction that produced a change in mobility, along with the three fractions that eluted right after it, produced the strongest degradation of the 6S RNA. This suggests that an RNase is being eluted in the same/close fractions that trigger a change in the mobility of 6S RNA.

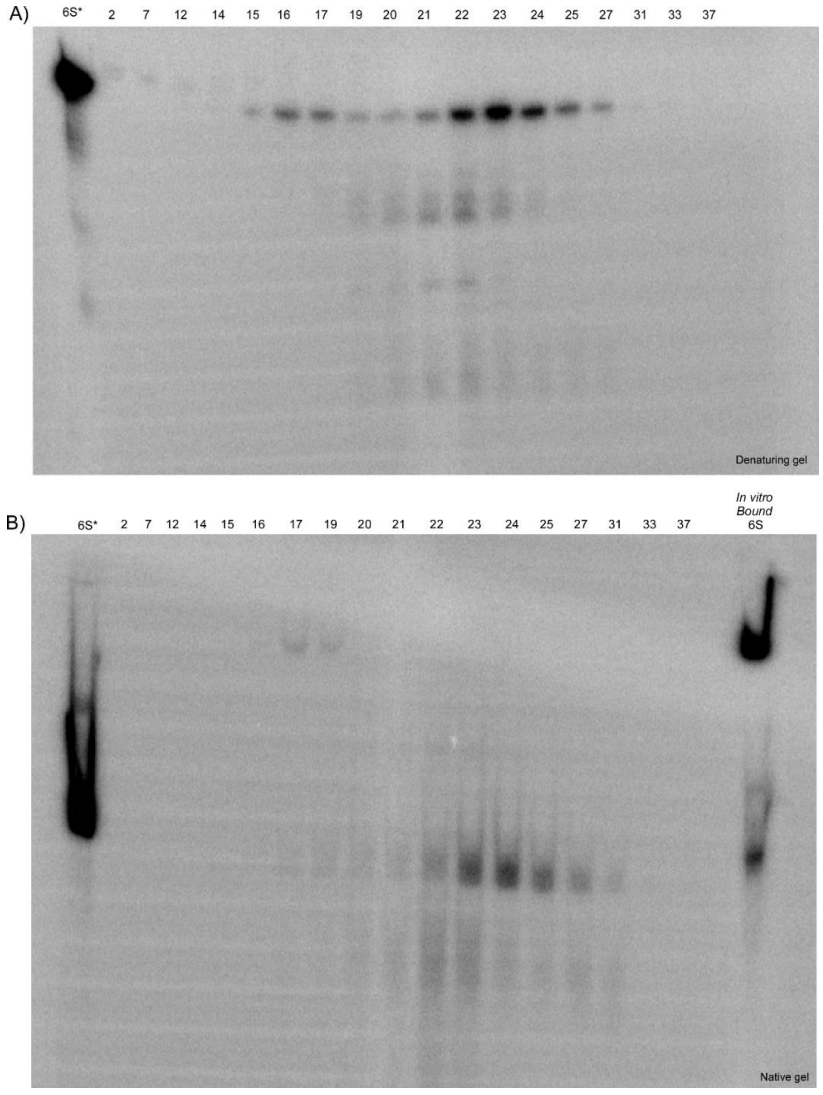



***Figure 3.14. Radiolabelled 6S RNA binds only RNA Polymerase Holoenzyme***

Cell extract fractions product of size exclusion chromatography were loaded into both, a denaturing (A) and native (B) gel. Fractions to be loaded were selected based on scintillation counting results in order to try to capture the three peaks that are seen in Figure 3.12A. The first lane on the left is a control that shows 6S RNA bound to commercially available $E\sigma^{70}$. The fractions collected from the size exclusion column are numbered from 2 to 37. The first lane on the right is free 6S RNA (6S*).

## *Strategy B. Protein Purification using Biotinylated 6S RNA yielded unrelated or highly abundant proteins*

Streptavidin magnetic beads are superparamagnetic particles covalently coupled to a highly pure form of streptavidin and can be used to capture biotinylated substrates with high specificity. To take advantage of this tool, biotinylated 6S RNA was produced

(section 3.2.3) in order to specifically capture 6S RNA-interacting proteins by incubating it with the crude cell extract or with the protein fractions obtained during size exclusion experiments. Based on the pmol of RNA added to the 6S RNA-Dynabeads experiments, and assuming a 1:1 RNA-protein interaction, it would have been expected to recover 9-100 ng of 11 KDa protein (depending on the scale of the experiment). Such an amount of protein should be easily visualized with silver staining that has a detection limit of 0.25 ng. The complexity of the cell extract mix was very evident when the samples from the protein pull-down experiment were resolved in a SDS-PAGE. The first and sixth lanes (Figure 3.15, from left to right) of the gel had a protein ladder. The second lane shows the crude cell extract which was used as input for this experiment. The next lane shows 6S RNA and how it stains differently in comparison to proteins. The fourth lane shows only the beads only that were treated exactly like the real sample, which allowed us to identify the streptavidin that was released onto the solution from the magnetic beads after boiling. The fifth and seventh lanes are the output of the pull-down experiment that concentrated the amount of proteins in the cell extract ~ 17X. The eighth lane in the gel shows the profile for the RNase Inhibitor on its own. The SN lane shows a small fraction of the supernatant recovered after the first pull-down of the magnetic beads. The SN lane looks very similar to the lane with the cell extract (C.E.), which highlights the advantages of using magnetic beads to specifically pull down proteins of interest. The product of the first wash of the beads already shows that a very small amount of free protein was left behind (WI). Subsequent washes did not show evidence of any protein.

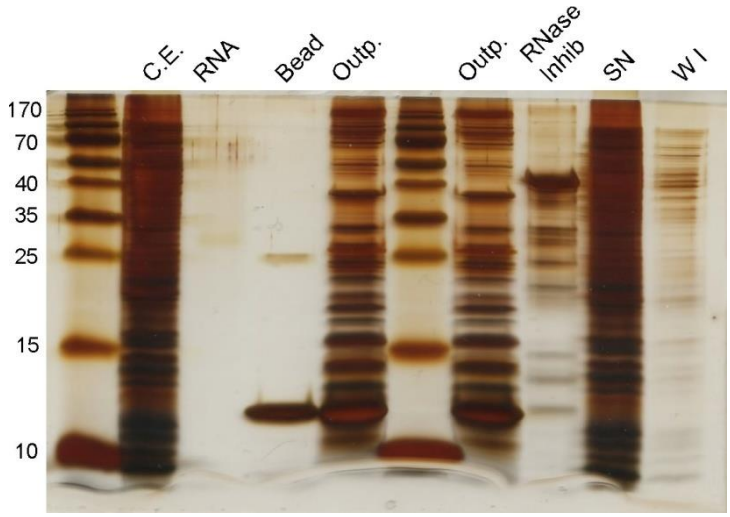



***Figure 3.15. Too many proteins were recovered when the 6S RNA-Dynabeads experiment was carried out using crude cell extract as input sample***

Lane 1 and 6 (from left to right) contain protein ladders. The second lane (C.E.) uses the crude cell extract as an input (No Biotin, No RNA control). The next lane (RNA) shows biotinylated 6S RNA that stains in a different hue of brown in comparison to proteins. The fourth lane (Bead) shows only the beads only that were treated exactly like the real sample, which allowed us to identify the streptavidin that was released onto the solution from the magnetic beads after boiling. The two bands seen in this lane are Streptavidin, which is a monomer or a dimer. The output lane (Outp.) shows the proteins that remained bound to the Dynabeads after 3 washes with PBS. The eighth lane contains a sample of the RNase inhibitor we used to reduce degradation while incubating the 6S RNA with the cell extract. The next lane (SN) shows the aqueous phase that contains all the proteins that did not bind to the beads. The lane on the far right is the first wash with PBS.

The bands in the output lanes were compared to those in the neighbouring control lanes in order to find bands that showed enrichment or were completely absent. It was difficult to achieve good resolution given the small size of the mini gel SDS-PAGE used. Nonetheless, six bands ranging in size from 10-40 KDa (purified using crude cell extract as input) were sent for Maldi-MS/MS analysis (Figure 3.15). One band was Streptavidin, which was used as internal control. Four other bands were 30S or 50S ribosomal protein fragments indicating that further steps of purification are required to eliminate highly abundant proteins. The sixth band was the alpha-subunit of RNA polymerase, which proved that 6S RNA is specifically binding to the RNA polymerase components, as expected. When the 6S RNA-Dynabeads approach was used with protein fraction as the input (instead of cell extract), the number of bands in the output sample was not reduced considerably and there was no obvious enrichment of any band

in particular. The expected 90 ng of protein product of the previously described enrichment were not recovered. These samples were not sent out for Maldi-MS/MS analysis (Figure 3.16).

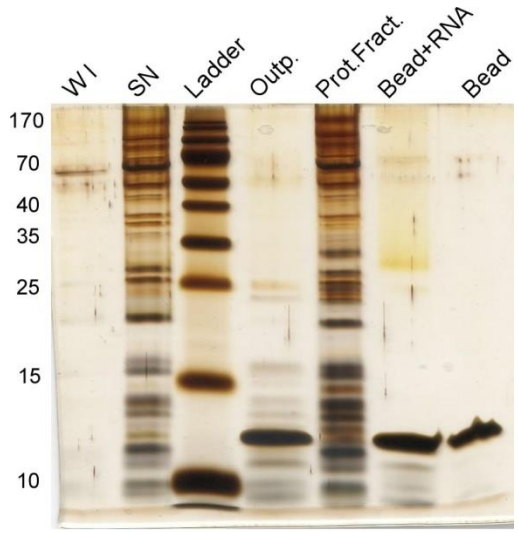



***Figure 3.16. Fewer proteins were observed when the 6S RNA-Dynabeads experiment was carried out using protein fraction as input sample***

However, no specific enrichment was observed for any protein band. The lane on the left is the first wash (WI). The second lane (SN) is the aqueous phase, which contains all of the proteins that did not bind to the Bio-6S RNA and in consequence, to the beads. The third lane is protein ladder. A discrete number of proteins are observed in the output lane (Outp), some of which come from the Dynabeads (Below 15 KDa). Others come from the RNA (around 70 KDa) and the remainder seem to be proteins that are highly abundant in the input and survived all the washes. The next lane shows the input sample Protein Fraction (Prot.Fract.). The lane after that (Bead+RNA) has Dynabeads and Bio-6S RNA, which were treated just like the output sample but did not see any cell extract. The last lane on the right had the same treatment as the previous one, but only with Dynabeads

### *Strategy C. Purification using combination of gels did not recover a 6S RNA-protein complex*

Another attempt was made to recover the proteins causing the mobility shift in successive gel purification steps that included native and denaturing conditions (Figure 3.17). The 6S RNA-CE complex was prepared and loaded in a Urea denaturing gel under the conditions that had previously shown slower mobility for 6S RNA. The slow mobility band was excised and gel purified. The recovered sample was separated in a SDS-PAGE gel, and silver stained, and in consequence the resulting sample was cleaner than the input cell extract. There were still too many bands that co-purified with the slower mobility RNA, as was observed during the size exclusion experiments. One

of the drawbacks of gel purification is that this method can pick up proteins that are merely co-migrating with the 6S RNA, but are not specifically interacting with it.

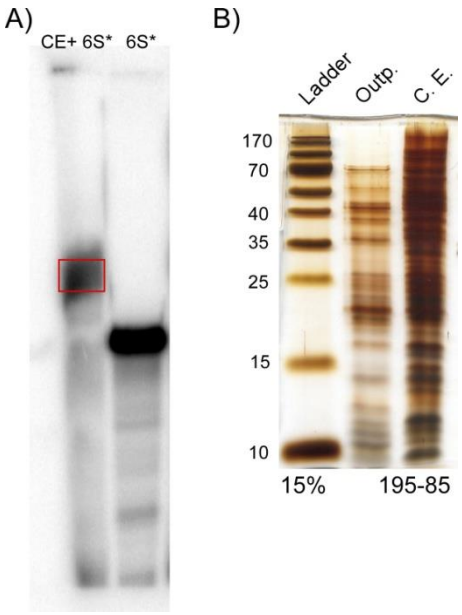



***Figure 3.17. Purification from a Urea-polyacrylamide gel did not improve the enrichment in individual bands***

The shifted complex was excised (red box) from a denaturing gel (panel A) and eluted out of the gel. The recovered material was loaded in a 15% SDS-PAGE (panel B). Although the sample looks cleaner when compared to the neighbour lane with crude cell extract (C. E.), a wide range of highly abundant proteins are still visible, while there is no evidence that any particular bands have been enriched.

A second urea polyacrylamide gel was also used in an attempt to remove further unrelated proteins. The product of the size shift in the first gel is recovered in a low salt buffer (1X TBE) and run in a second urea polyacrylamide gel. The 6S RNA is still visualized in the second gel, but the slower mobility band has disappeared. It seems like the putative 6S RNA-protein complex does not survive either the elution out of the first gel, or the loading/running in a second gel. These observations led to the testing of the integrity of the putative complex under the different salt conditions that have been previously described (Figure 3.8).

The same sample was loaded in a thin gel (Figure 3.9) and in a thicker native gel of equal percentage, with the purpose of gel purifying material from it. However, the resolution decreased considerably (Figure 3.18). The bands in the red box (Figure 3.18A) were cut from lanes with Cell extract alone or Cell extract + 6S RNA. Material

was eluted out of the gel, concentrated (using Amicon YM3), and run in 15% SDS-PAGE. Very little sample was recovered with a size less than 25 KDa. There is no observable difference between the proteins observed in the + and – 6S RNA lanes (Figure 3.18B), which suggests that particular proteins have not been enriched.

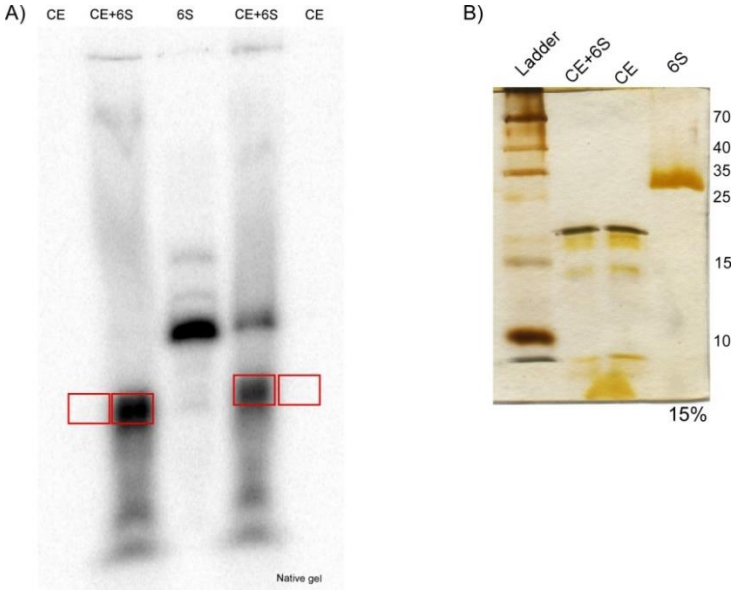



***Figure 3.18. Small amount of protein was gel purified from Native gel***

A) The sample loaded in Fig 3.13 was also loaded in the native gel shown above. However the gel resolution was not as good as the resolution in Figure 3.13 due to the gel being thicker. Samples were cut as indicated by the red boxes, eluted and concentrated. B) Gel purified samples were run in SDS-PAGE. Only bands below 25 KDa can be observed in both lanes. However no obvious enrichment or band pattern differences are seen between cell extract alone (CE) and cell extract+6S RNA (CE+6S)

## 3.4. Conclusion

Our work gave us a glimpse of the versatility and complexity that *E. coli* displays in order to adapt to different environments and stress conditions to ensure bacterial survival. The results described in this chapter did not allow us to verify our hypothesis of a protein factor that binds to the excess mutant RNA and tags it for destruction so that cells can resume exponential growth. Furthermore, we cannot conclusively say that the proteins identified are actually interacting with 6S RNA *in vivo.* Some of the proteins that were identified through Maldi-MS/MS were 30S or 50S ribosomal proteins. These findings provide evidence that further purification steps are required to eliminate the highly abundant proteins that might be preventing 6S RNA from specifically binding to

proteins that are less abundant in the cell extract. Cell extract could be treated with DNase and even RNase in order to clean it up and reduce the chance that non-specific interactions will occur. Combining strategies A and B could be useful for improving the chances of picking up less abundant proteins that interact with 6S RNA. Adding at least one other size exclusion column with a narrower size range of protein separation could help to separate the cell extract into fractions with proteins in specific size ranges. Each one of those fractions could then be reacted with the biotinylated 6S RNA bound to Streptavidin beads. This approach worked well and improving the quality of the input material could increase the chances of detecting low abundance proteins interacting with the Bio-6S RNA.

The results that spearheaded the research described above might have been misinterpreted. Consequently, the change in mobility observed for 6S RNA when it is run with the total cell extract under denaturing conditions might not be dependent on a protein/that is/are interacting with 6S RNA. Rather, it might be dependent on the buffer and running condition of the gel system at use. Nonetheless, we still contend that it is possible to find proteins that interact specifically with 6S RNA. Bacteria are highly efficient organisms and there must be a good reason why thousands of free 6S RNA molecules are present in the early stages of cellular growth in liquid media.

The possibility that 6S RNA is interacting with an RNase is plausible based on observations made during our size exclusion experiments. These observations showed a strong degradation of 6S RNA in specific fractions (Figure not shown). Gel experiments also confirm that a small percentage of the counts in the 6S+CE lane (Figure 3.6) are indeed degradation products because, some of the bands with faster mobilities remain, even after treatment with Proteinase K and the slow mobility band returns to normal 6S RNA mobility. RNase E or RNase III are candidates to be the RNase at the heart of the complex degrading 6S RNA. Unlike RNase III, which targets double-stranded RNA, RNAse E is a single-strand specific endoribonuclease (Lalaouna, Simoneau-Roy *et al.* 2013). Although RNase III and RNAse E seem to target different substrates and participate in different paths, recent microarray studies have also detected a certain redundancy in their activities (Stead, Marshburn *et al.* 2011). It is discouraging to note that transcriptome-wide studies failed to identify an 6S RNA binding

to Hfq, which is often paired with RNase E in the degradosome (Zhang, Wassarman *et al.* 2003, Lorenz, Gesell *et al.* 2010). Nonetheless, this might not be sufficient to discard an interaction with RNase E. Another protein factor could possibly be working as a chaperone between 6S RNA and RNase E. The second possibility is that RNase III is pairing with 6S RNA in a Hfq-independent manner via the A/U rich areas of the 6S RNA and thereby degrading it (Lalaouna, Simoneau-Roy *et al.* 2013). The literature provides several examples of RNase III playing a role in rRNA maturation, translation regulation, and mRNA degradation. However, none of these studies talk about RNase III degrading non-coding RNAs (Huntzinger, Boisset *et al.* 2005, Darfeuille, Unoson *et al.* 2007, Lim, Ahn *et al.* 2014).

Another recently developed approach to tagging and following RNA could also be used to improve the chances of identifying low abundance complexes between 6S RNA and other proteins. RNA Mango, which has been developed in the Unrau lab, is an RNA aptamer with a G quadruplex at its core that binds to a fluorophore (TO1), increasing its fluorescence by up to 1100 fold. The RNA aptamer sequence has already been incorporated into the 6S RNA sequence and *in vitro* expression has shown that it behaves indistinguishably from wild type 6S RNA. In addition, the strength of the binding interaction has been already tested by capturing 6S RNA Mango tagged onto streptavidin beads by using TO1-Biotin as a handle (Dolgosheina, Jeng *et al.* 2014). Taking advantage of this newly developed tool and incorporating it into the purification scheme might facilitate the capture of 6S RNA-Protein complexes that remain undetected.

In the previous two chapters I have focused on generating a better understanding 6S RNA, a nc-RNA that has a widespread effect during transcription regulation in bacteria. The next chapter moves away from the focus on prokaryotic cells to focus on eukaryotic cells, specifically *Leishmania sp.* RNA and a potential connection to pathogenesis remains my main interest. In Chapter 4, I explore the possibility that Leishmania-derived exosomes are not only shuttling proteins into potential host cells, but also RNAs.

Section 1.2.3 of Chapter 1 (page 42) provides background information on the biology of Trypanosomatids in general and Leishamania in particular. It also discusses previous work that describes exosomes and supports our interest in exosomes as vehicles that aid the parasite during pathogenesis. Finally, this section will introduce a new type of small regulatory RNA known as tRFs. Although several questions remain as to whether these are real regulatory RNAs, they are worth studying as an increasing body of new evidence suggests they might have a regulatory role. However, how this role is performed has not yet been determined.

Chapter 4.

# Small RNAs Derived From tRNAs and rRNAs Are Highly Enriched in Exosomes From Both Old and New World *Leishmania* Providing Evidence For Conserved Exosomal RNA Packaging

Ulrike Lambertz[1]*, Mariana Oviedo Ovando[2]*, Elton J R Vasconcelos[3], Peter J Unrau[2], Peter J Myler[3,4] and Neil E Reiner[1§]

[1]Departments of Medicine, Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada

[2]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada

[3]Seattle Biomedical Research Institute, Seattle, WA, USA

[4]Departments of Global Health and Biomedical Informatics & Medical Education, University of Washington, WA, USA

*These authors contributed equally to this work

A modified version of this chapter as been accepted at the BMC Genomics Journal 16 (1) P151.

The work described in the present chapter was done in collaboration with Dr. Neil Reiner's laboratory at the University of British Columbia. Mariana Oviedo-Ovando was in charge of optimizing RNA extraction protocols from *Leishmania* sp. and exosomes, RNA characterization, development and optimization of RNA library construction

protocol, as well as Northern blots. UL is a PhD student in Dr. Neil Reiner's laboratory at the University of British Columbia. UL was in charge of cell culture and exosome preparation, Nanosight analysis, Vesicle delivery to macrophages experiments, Agilent Bioanalyzer experiments, manual data inspection in genome browser Artemis to identify the most abundant clusters and the novel transcripts. UL also supervised all data analysis performed by EJRV. Once protocols were established and optimized, UL collaborated with RNA extractions, RNA characterization and library construction. EJRV is a Post-Doctoral fellow at Dr Peter Myler's laboratory at the University of Washington. EJRV was in charge of initial data clean up (read orientation, adapter trimming, reads collapsing), alignment of reads with reference genomes, clustering-detailed analysis and characterization of tRNA & rRNA fragments, open reading frame analysis for novel transcripts, BLAST searches.

RNA characterization as well as library construction were in part based on the book chapter reproduced as Appendix A, by Ebhardt, H. A., Oviedo Ovando, M. E., and Unrau, P. J. (Ebhardt, Ovando *et al.* 2012). Supplemental data for the present chapter can be found in Appendix B.

## 4.1. Abstract

*Leishmania sp.* use exosomes to communicate with their mammalian hosts and these secreted vesicles appear to contribute to pathogenesis by delivering protein virulence factors to macrophages. In other eukaryotes, exosomes were found to carry RNA cargo, such as mRNAs and small non-coding RNAs capable of altering recipient cell phenotype. Whether *Leishmania sp.* exosomes also contain RNAs which they are able to deliver to bystander cells is not known. Here, we show that *Leishmania sp.* exosomes indeed contain RNAs and we compare them and contrast the RNA content of exosomes released by *Leishmania donovani* and *Leishmania braziliensis*.

We purified RNA from exosomes collected from axenic amastigote culture supernatant and found that when compared with total *Leishmania sp.* RNA, exosomes mainly contained short RNA sequences. Exosomes with intact membranes were capable of protecting their RNA cargo from degradation by RNase. Moreover, exosome

RNA cargo was delivered to host cell cytoplasm *in vitro*. Sequencing of exosomal RNA indicated that the majority of cargo sequences were derived from non-coding RNA species such as rRNA and tRNA. In depth analysis revealed the presence of tRNA-derived small RNAs, a novel RNA type with suspected regulatory functions. Northern blotting confirmed the specific and selective enrichment of tRNA-derived small RNAs in exosomes. We also identified a number of novel transcripts, which appeared to be specifically enriched in exosomes compared to total cell RNA. In addition, we observed the presence of sequences mapping to siRNA-coding regions in *L. braziliensis*, but not in *L. donovani* exosomes.

The high-throughput sequence data as well as Northern Blot results show that *Leishmania sp.* exosomes are selectively and specifically enriched in small RNAs derived almost exclusively from non-coding RNAs. These exosomes are competent to deliver their cargo of novel, potential small regulatory RNAs to macrophages where they may influence parasite-host cell interactions. The remarkably high degree of congruence in exosomal RNA content between *L. donovani* and *L. braziliensis*, argues for the presence of a conserved mechanism for exosomal RNA packaging in *Leishmania sp*. These findings open up a new avenue of research on non-canonical, small RNA pathways in trypanosomatids, which may elucidate pathogenesis and identify novel therapeutic approaches.

## 4.2. Introduction

Protozoan parasites of the genus *Leishmania sp.* are highly endemic to tropical and sub-tropical regions of the world. They are transmitted to humans and other mammals by sandfly vectors that inject the flagellated, promastigote life cycle stage of *Leishmania sp.* into the dermis of the host while taking a blood meal. After inoculation, promastigotes are engulfed by host mononuclear phagocytes either directly or indirectly as cargo of apoptotic neutrophils (van Zandbergen, Klinger *et al.* 2004). Following their ingestion by host cells, promastigotes take up residence in the phagolysosome, where they transform into amastigotes and undergo cell proliferation. Depending on the infecting *Leishmania* species, disease manifestations and symptoms can vary widely from mild self-healing cutaneous lesions to lethal visceral disease. The two species that

are the focus of the present study, *Leishmania donovani* and *Leishmania braziliensis*, cause visceral and mucocutaneous Leishmaniasis, respectively. While the former is naturally the more serious threat as it can lead to death if left untreated, the latter can have an extremely high impact on the affected individual due to debilitating and disfiguring destruction of critical soft tissue structures.

The current paucity of effective and well tolerated drug treatments, along with the lack of highly efficacious, well standardized and widely available vaccination strategies, can be attributed at least in part to the gap of knowledge about the intricate interplay between *Leishmania sp.* and host macrophages. Macrophages are key players in both innate and adaptive immune responses. Their primary function is to engulf and digest prey, whether pathogen or debris from cellular turnover, which makes their intracellular environment very nutrient-rich. *Leishmania sp.* exploits these macrophage characteristics in a very sophisticated manner. *Leishmania sp.* allows the phagocyte to ingest it, and then uses the cell as its safe nursery, where it scavenges nutrients and replicates while remaining unrecognized by other immune cells. The mechanisms by which *Leishmania sp.* manages to survive within these potent immune cells are starting to be elucidated. One key strategy employed by *Leishmania sp.* appears to be the prevention of macrophage activation, a step that is crucial to induce macrophage digestion and killing functions (Nandan, Lo *et al.* 1999, Junghae and Raynes 2002, Nandan, Yi *et al.* 2002). At the same time, *Leishmania sp.* are resistant to the harsh conditions of the acidifying phagolysosome (McConville, De Souza *et al.* 2007).

In principle, there are two categories of molecules, surface associated and secreted, made available by *Leishmania sp.* to communicate with the host and turn on and off macrophage cellular functions. Regarding secreted molecules, our group has recently discovered that *Leishmania sp.* use a non-classical secretion mechanism to export a majority of their secreted proteins, which involves the release of small vesicles called exosomes (Silverman, Chan *et al.* 2008, Silverman, Clos *et al.* 2010).

Exosomes are 50-100 nanometre-sized membrane vesicles secreted by a variety of single- as well as multi-cellular eukaryotic organisms. They are distinct from membrane microvesicles, which are produced by blebbing, since their release occurs

through fusion of multivesicular bodies from the endocytic/exocytic pathway with the plasma membrane of the cell (Kowal, Tkach *et al.* 2014). Extracellular vesicles such as microvesicles and exosomes had long been considered to be simply cellular garbage bags. Only recently has the release of specific cargo within vesicles, as well as their uptake and effects on recipient cells, been appreciated to represent important biological events. Extracellular vesicle release has also been documented in the context of infection, where the vesicles were shown to contain both host and pathogen-derived antigens and virulence factors (reviewed in (Silverman and Reiner 2011)). Extracellular vesicles containing pathogen-derived factors may be released either by infected cells, as has been shown following infection with Eppstein-Barr virus, mycobacteria, toxoplasma or plasmodia (Giri, Kruh *et al.* 2010, Pegtel, Cosmopoulos *et al.* 2010, Mantel, Hoang *et al.* 2013, Pope and Lasser 2013), or released by the pathogen directly, e.g. mycobacteria, cryptococci, *Trypanosoma sp.* and *Leishmania sp.* (Panepinto, Komperda *et al.* 2009, Silverman, Clos *et al.* 2010, Prados-Rosales, Baena *et al.* 2011, Bayer-Santos, guilar-Bonavides *et al.* 2013, Garcia-Silva, das Neves *et al.* 2014).

Importantly, in our studies, *L. donovani* exosomes and exosomal proteins were detected in the cytosolic compartment of infected macrophages (Silverman, Clos *et al.* 2010). Moreover, we showed that *L. donovani* exosomes can modulate mononuclear cell phenotypes *in vitro*, rendering them anti-inflammatory by specifically inhibiting cytokine production. Studies with C57Bl/6 and Balb/c mice provided evidence that treatment with exosomes from *L. donovani* as well as *L. major* prior to infection exacerbated disease *in vivo* (Silverman, Clos *et al.* 2010). These findings have fundamentally transformed our understanding of how *Leishmania sp.* are able to communicate with the host. Two other studies have since supported a role for exosomes in *Leishmania sp.* pathogenesis. In the first study, the authors showed that the metalloprotease GP63 delivered by *L. donovani* exosomes cleaved the nuclease Dicer 1 in murine hepatocytes, resulting in down-regulation of microRNA-122 expression, lowering of serum cholesterol and enhancement of murine liver infection (Ghosh, Bose *et al.* 2013). In a second study, another group reported that *L. major* exosomes globally affected macrophage gene expression, which was in part GP63-dependent (Hassani, Shio *et al.* 2014). In summary, these results make a strong case for the importance of exosomes in *Leishmania sp.* pathogenesis.

In addition to their protein cargo, exosomes and microvesicles were recently shown to be carriers of nucleic acids in the form of RNA. This observation was first made in mast cell exosomes, which were found to contain mRNA as well as miRNA (Valadi, Ekstrom *et al.* 2007). Interestingly, these molecules were functional and could transduce signals in recipient cells. Since then, exosomal RNAs have been implicated in the pathogenesis of a variety of important, chronic infections. For example, Eppstein-Barr virus-infected B-cells were shown to release exosomes containing viral miRNAs which could regulate gene expression in recipient cells (Pegtel, Cosmopoulos *et al.* 2010). *Toxoplasma gondii*-infected fibroblasts released exosomes containing a set of host mRNAs and miRNAs that was distinct from that of uninfected, serum-starved cells (Pope and Lasser 2013). However, to date only two protozoan pathogens have been found to release RNA-containing extracellular vesicles directly. Thus, *Trichomonas vaginalis* exosomes were reported to contain RNA sequences, the biotype and function of which still remain to be determined (Twu, de *et al.* 2013). *Trypanosoma cruzi* was shown to release extracellular microvesicles containing a variety of non-coding RNAs including tRFs that have a suspected regulatory nature (Bayer-Santos, guilar-Bonavides *et al.* 2013, Garcia-Silva, Sanguinetti *et al.* 2014).

Based on the evidence that exosomes may serve as biologically important shuttle vectors for RNAs, in the present study, we sought to investigate the RNA content of *Leishmania sp.* exosomes. We indeed found that *Leishmania sp.* exosomes contained RNA cargo which they were capable of delivering to host cells *in vitro*. Using high throughput sequencing and bioinformatics analyses, we found that *Leishmania sp.* exosomes were enriched in small RNAs derived from largely non-coding RNAs. Notably, we discovered that these vesicles contained a relatively abundant and highly selective population of small RNAs derived from mature tRNAs. Furthermore, we found a number of novel transcripts, some of which were highly enriched in exosomes. Although exosomes released by both *L. donovani* and *L. braziliensis* had largely similar RNA content, *L. braziliensis* exosomes specifically contained transcripts derived from genes that also code for siRNAs.

Taken together, these findings show for the first time that *Leishmania sp.* exosomes are highly enriched in small non-coding RNAs, particularly tRNA-derived

small RNAs with potential regulatory functions. This suggests that these RNAs may have functions in intercellular communication. These findings hint at a previously unrecognized potential mechanism of *Leishmania sp.* pathogenesis, mediated through the exosomal delivery of small, principally non-coding RNAs to mammalian host cells.

## 4.3. Material and Methods

### 4.3.1. Cell culture

*L. donovani* Sudan strain S2 promastigotes are routinely cultured in M199 (Sigma-Aldrich) with 10% heat inactivated fetal bovine serum (FBS, Gibco), 20 mM HEPES (Stemcell), 6 µg/ml hemin (Sigma-Aldrich), 10 µg/ml folic acid (Sigma-Aldrich), 2 mM L-glutamine (Stemcell), 100 U/ml penicillin/streptomycin (Stemcell) and 100 uM adenosine (Sigma-Aldrich) at 26°C. Every 3 days the organisms were subcultured 1:10 in fresh medium and were kept in culture for a maximum of 20-25 passages. Fresh parasites were obtained by purification of amastigotes from spleens of infected Syrian Golden hamsters followed by *in vitro* transformation into promastigotes by culturing for 7 days at 26°C in promastigote media.

*L. braziliensis* (clinical isolate from the Peruvian Amazon region) promastigotes were routinely cultured in the same media as above except for supplementation with 20% FBS. *L. braziliensis* promastigotes were subcultured 1:5 every 3 days in fresh media and kept at 26°C.

### 4.3.2. Purification of exosomes

Exosomes were purified from *L. donovani* and *L. braziliensis* axenic amastigote culture supernatant as described previously (Silverman, Clos *et al.* 2010, Silverman, Clos *et al.* 2010). Briefly, 400-800 ml of day 5 promastigotes (at a concentration of $5x10E^7$ cells/ml) were washed 2X with Hank's buffered salt solution (HBSS, Sigma-Aldrich) followed by incubation in serum-free buffered exosome collection media at pH = 5.5, RPMI1640 supplemented with 1% D-glucose, 20 mM HEPES, 2 mM L-glutamine, 100 U/ml penicillin/streptomycin and 25 mM MES (all from Sigma-Aldrich), at 34°C for *L.*

*braziliensis* and 37°C for *L. donovani*. After 24 hours of incubation, exosomes were purified from the 400-800 ml culture supernatant under endotoxin-free conditions by a series of centrifugation and filtration steps, followed by flotation on a sucrose cushion, as described in (Silverman, Clos *et al.* 2010). After a final pelleting step at 100,000 g for 1 hour, purified exosomes were resuspended in 50-100 µl of PBS and processed immediately (in case of RNA extractions) or stored at 4°C for a maximum of 5 days (for macrophage uptake experiment and Nanosight analysis).

### 4.3.3. Nanosight particle tracking analysis

The size and concentration of the isolated exosomes were analysed using the NanoSight™ LM10-HS10 system (Malvern Instruments). For analysis, a monochromatic laser beam (405 nm) was applied to the diluted exosome solution (1:100 in 0.02 µm filtered PBS) that was injected into a LM12 viewing unit using a computer controlled syringe pump. NanoSight™ tracking analysis (NTA) software version 2.3 was used to produce the mean and median vesicle size together with an estimate of particle concentration. Samples were measured 3 times to confirm reproducibility.

### 4.3.4. Extraction and biochemical characterization of RNA

RNA was purified from *Leishmania sp.* exosomes by phenol/chloroform extraction using all RNA-grade reagents. For this purpose, 150 µL of LETS buffer (0.1M LiCl, 0.01 M $Na_2EDTA$, 0.01 M Tris-Cl pH=7.4, 0.2% SDS, all Sigma-Aldrich) was added to 50 µl of exosomes resuspended in PBS followed by addition of 200 µl Ultra-Pure buffer-saturated phenol pH = 7.4 (Life Technologies). The mixture was vortexed vigorously and centrifuged for 2 min at 13,000 g in a microcentrifuge at room temperature. The upper aqueous phase was collected and the phenol extraction was repeated once more followed by two extractions over 200 µl chloroform each (Fisher Scientific). RNA was precipitated by addition of 0.3 M NaCl, 2 µg/ml glycogen (Ambion) and 75% EtOH, and incubation at -20°C overnight. RNA was pelleted by centrifugation at 13,000 g for 30 min at 4°C. RNA pellets were washed with ice-cold 75% EtOH and resuspended in 10-20 µl $ddH_2O$. RNA concentration was determined by measuring the $OD_{260}$ with the nanodrop (Thermo).

To look at length profiles of exosome-derived RNA, 2 µg of purified *Leishmania sp.* total RNA and 1 µg of exosome RNA were first treated with 5 units DNase I (Thermo) to remove potential DNA contamination. After incubation for 30 min at 37°C, DNase was inactivated by addition of 2.5 mM EDTA and incubation at 65°C for 10 min followed by phenol-chloroform extraction and EtOH precipitation as above. RNA was resuspended in 4 µl $ddH_2O$ and RNA length profiles were obtained with the Agilent Bioanalyzer using the RNA 6000 Pico kit according to the manufacturer's instructions (Agilent). Alternatively, DNase-treated RNA was run on a 15% polyacrylamide gel, stained with SYBR green (Life Technologies) and imaged with UV-imaging.

To confirm identity of nucleic acid purified from exosomes as RNA, 1-2 µg of phenol/chloroform extracted RNA was treated with DNase (as above), followed by treatment with either 0.4 mg/ml RNase A (Thermo) for 15 min at 37°C or hydrolysis with 50 mM KOH (Sigma-Aldrich) for 15 min at 95°C. Samples were then 5' end labeled according to the manufacturer's instructions using polynucleotide kinase (PNK) (New England Biolabs, NEB) and [$\gamma$–$^{32}$P]-ATP (Life Technologies) and run on 15% polyacrylamide gels followed by imaging with a Typhoon phosphor-imager (GE Healthcare).

To assess whether the exosomal membrane was protecting the vesicular RNA content from degradation by exogenous RNases, intact exosomes resuspended in PBS (from 400 ml culture supernatant, split into 4 samples) were treated with 0.4 mg/ml RNAse A for 15 min at 37°C in the presence or absence of 0.1% Triton X-100 (Sigma-Aldrich). As a control for RNase activity, 1 µl of prepared RNA pico ladder (Agilent) was treated with RNase A under the same conditions. After incubation, samples were extracted with phenol/chloroform twice each and RNA was precipitated with EtOH as above. Samples were then treated with DNase, again phenol/chloroform extracted and EtOH precipitated, resuspended in 4 µl $ddH_2O$ and run on the Agilent Bioanalyzer to determine whether or not RNA had undergone degradation.

### 4.3.5. Vesicle delivery of RNA cargo to macrophages

Exosomes were purified from 400 ml culture supernatant of *L. donovani* axenic amastigotes as described above. Pelleted exosomes were resuspended in 100 µl PBS. Protein concentration in the exosome preparation was determined using the Micro BCA Protein Assay kit (Pierce). Exosomes were then stained with the membrane-permeant, RNA-specific dye SYTO RNASelect (Life Technologies) according to the manufacturer's recommendations. For this purpose, the SYTO dye was diluted in DMSO and added to the resuspended exosomes at a final concentration of 10 µM, followed by 20 min incubation at 37°C. Excess unbound dye was removed by washing twice with 1 ml PBS, pelleting the exosomes at 100,000 g for 1 hour at 4°C. Exosomes were then resuspended in the original volume of PBS (100 µl). Labelling efficiency was assessed by fluorescence microscopy using an Axioplan II epifluorescence microscope equipped with 63×/1.4 Plan-Apochromat objective (Carl Zeiss Inc). Images were recorded using an AxioCam MRm Camera coupled to the AxioVision software Version 4.8.2 (Carl Zeiss Inc.).

To investigate the exosome-mediated delivery of RNA to host macrophages, THP-1 cells were differentiated O/N with 10 ng/ml phorbol-12-myristate 13-acetate (PMA), followed by washing and resting cells for 24 hours. Differentiated cells were then treated with labeled exosomes for 2 hours at 37°C. As a negative control, cells were treated with labelled exosomes and incubated at 4°C for 2 hours, preventing phagocytosis. For quantification of exosome RNA uptake, exosome-treated THP-1 cells were washed 3 times with PBS to remove non-internalized exosomes. Cells were then fixed with 2% paraformaldehyde (Sigma) in PBS for 15 min at room temperature. After fixation, cells were again washed with PBS and then analyzed by flow cytometry (FACS Calibur, BD). To verify that exosomes were in fact internalized and not just bound to the cell membrane, the same experiment was performed with THP-1 cells grown on coverslips to be analysed by confocal microscopy. After incubation, cells on coverslips were washed and fixed as above, permabilized with 0.1% Triton X-100 in PBS for 5 min, and stained with Alexa Fluor 594 phalloidin (Life Technologies) for 1 hour at room temperature in the dark. After 3x washing with PBS, coverslips were mounted with Prolong Gold antifade mounting media containing DAPI (Life Technologies) to detect

macrophage nuclei. Confocal microscopy was done with a Leica DMIRE2 inverted microscope equipped with a SP2 AOBS laser scanning head. This is a filter-free spectral confocal and multiphoton microscope, and all imaging operations which include selections of laser, detection channels and other functions are fully automated and computer controlled. Pictures were taken with a 63X magnification oil immersion objective.

### 4.3.6. Library construction and sequencing

We used 1-2 µg of RNA extracted by phenol/chloroform extraction from *L. donovani* and *L. braziliensis* exosomes (from one individual exosome preparation each, from 800 ml supernatant) as starting material. RNA was first treated with DNase I (as described above) to remove potentially contaminating DNA. To remove 5' phosphates on the RNA, we first performed a calf intestinal alkaline phosphatase (CIP) treatment using 1 unit of CIP (Roche) per 10 µl reaction and incubation for one hour at 37°C. Once incubation was completed, samples were phenol-chloroform extracted twice and EtOH precipitated as described above. In order to monitor the efficiency of the CIP treatment, a parallel reaction was spiked with a 24 nt long radiolabelled RNA, and pre- and post-incubation with CIP were loaded in a 20% denaturing polyacrylamide gel to follow the disappearance of counts. Next, the CIP treated RNA sample (resuspended in 10 µl $ddH_2O$) was treated with tobacco acid phosphatase (TAP) to remove 5' caps. Half of the CIP treated RNA sample was combined with 2.5 U of TAP (Epicentre), 1X TAP buffer, brought to a final volume of 10 µl with $ddH_2O$ and incubated for one hour at 37°C. In order to control for the efficiency of 5' cap removal, a parallel reaction was spiked with [$\gamma$–$^{32}$P]-ATP and pre-and post-ligation samples were loaded in a 20% plyacrylamide gel to monitor the disappearance of counts. RNA was EtOH-precipitated and resuspended in 10 µl $ddH_2O$. The CIP and TAP treated RNA were then labeled with polynucleotide kinase (PNK) to have the same 5' phosphate in all RNA molecules about to be ligated. Ten U of PNK (NEB) were used along with 1X PNK buffer (NEB), and [$\gamma$–$^{32}$P]-ATP in a 10 µl reaction and incubation for one hour at 37°C. Next, ~ 10% of the sample was loaded onto a denaturing 15% polyacrylamide gel. The remainder of the PNK reaction was taken to a 30 µl volume with $ddH_2O$ and run through a dye terminator removal (DTR) cartridge (EdgeBio) following manufacture's indications in order to remove ions

and unincorporated [$\gamma$–$^{32}$P]-ATP. The sample was EtOH precipitated and resuspended in 5 µl of dd$H_2O$. The next step was to ligate a custom adenylated AppDNA adaptor (5' App-GAA GAG CCT ACG ACG A) to the 3' end of RNA molecules. This adaptor was slightly modified so that the 3' end was blocked in order to prevent self-ligation. Half of the pre-treated exosomal RNA sample was combined with T4 RNA ligase buffer (50 mM HEPES, pH 8.3, 10 mM $MgCl_2$, 3.3 mM DTT, 10 g/ml BSA and 8.3% Glycerol), 2.5 U of T4 RNA Ligase (Epicentre), and 20 µM AppDNA adaptor. Reactions were incubated at room temperature for one hour, and then the enzyme was denatured by heating at 65°C for 20 min. Samples were gel purified on 10% polyacrylamide gels to remove un-ligated adaptor. A second ligation reaction was set up to attach an RNA adaptor (5' rAUC GUA GGC ACC UGA AA) to the 5' end of the RNA-DNA hybrid. Conditions were the same as described above for the first ligation reaction with the only difference being that [$\gamma$–$^{32}$P]-ATP (final concentration of 0.4 mM) was added. The ligation reaction was gel purified from a denaturing 10% polyacrylamide gel and the recovered material was used as a template in a reverse transcription (RT) reaction. For this purpose, half of the recovered sample was combined with 100 µM RT primer (5' TCG TCG TAG GCT CTT C), dd$H_2O$ and incubated at 80°C for two min. After cooling samples down slowly, 1X First Strand Buffer (Life Technologies), 0.8 µM dNTP and 200 U of Superscript II Reverse Transcriptase (Life Technologies) were added. Controls with no enzyme and no template in the reaction were prepared in parallel. Reactions were incubated for 1 hour at 48°C. The RNA template was hydrolyzed by heating in the presence of 100 mM KOH followed by neutralization with 1 M Tris-HCl pH=5 (to a final pH=8), and the resulting cDNA was isolated on a 10% denaturing polyacrylamide gel. Twenty cycles of PCR amplification were performed in the presence of 5 mM $MgCl_2$, 100 µM dNTPs, 1 µM each forward primer (5' ATC GTA GGC ACC TGA AA) and reverse primer (same as RT primer), 1X Taq buffer and 2.5 Units Taq polymerase (UBI). PCR products were then gel purified and quantified by Qubit (Life Technologies) and used as input material for ligation of TruSeq adapters (Illumina) according to the manufacturer's recommendations. One hundred and fifty base pair, paired-end sequencing was performed using an Illumina MiSeq instrument (Illumina) at the Epigenomics core of Albert Einstein College of Medicine, NY.

### 4.3.7. Sequencing data analysis

After completion of Illumina paired-end sequencing and read quality control checking by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), both *L. donovani* and *L. braziliensis* exosomal RNA reads had their adapters trimmed by cutadapt version 1.0 (http://journal.embnet.org/index.php/embnetjournal/article/view/200/479). For each library, the output files from the trimming were separated into RNA adapter-trimmed reads and DNA adapter-trimmed reads, and the former was used to guide the assignment of correct orientation for all reads sequenced. We ran FLASh (settings:-M100 –x0.2) (Magoc and Salzberg 2011) and FASTX - Collapser (http://hannonlab.cshl.edu/ fastx_toolkit) to respectively combine the mates, for the cases where DNA inserts were shorter than twice the length of reads, and then collapsed identical sequences into single ones to facilitate handling the data in subsequent specific analyses. Bowtie2 version 2.1.0 (settings: very-sensitive-local –N1) (Langmead and Salzberg 2012) was used to align the collapsed reads (cReads) from both libraries against their respective reference genomes (LdBPK (*Leishmania donovani* strain BPK282A1) and LbrM (*Leishmania braziliensis MHOM/BR/75/M2904*) from TriTrypDB version 6.0), as well as against the species with the best assembled and annotated genome (LmjF, *L. major* MHOM/IL/81/Friedlin, TriTrypDB version 6.0). The very-sensitive-local setting of Bowtie 2 uses a seed length of 20 nt for the alignment, and the –N1 command allows for only one mismatch on that seed alignment. The alignments with LdBPK and LbrM were used to categorize the exosomal RNAs for the respective species, relying on htseq-count script (Simon, Paul Theodor *et al.* 2015) and the GFF files provided by TriTrypDB v6.0. The alignment with LmjF was done mainly to refine the analyses of reads mapping onto tRNAs. Of note, right after bowtie2 execution, samtools version 0.1.18 (Li, Handsaker *et al.* 2009) was applied to generate sorted bam files, which were then used as input to cufflinks (settings: -u --min-intron-length 3 --3-overhang-tolerance 25 --overlap-radius 10 --min-frags-per-transfrag 1) (Roberts, Pimentel *et al.* 2011) for the assembly of reads mapping on the same locus into individual “transcripts” or clusters. Artemis genome browser software (Carver, Harris *et al.* 2012) was used to manually inspect in greater detail and visualize the alignment of exosomal sequences with the reference genomes.

As mentioned above, we used reads mapping to *L. major* tRNAs for a better categorization of potential tRNA-derived small RNAs present within the exosomes. An *ad-hoc* PERL script was written to calculate the cReads position within each tRNA feature they mapped onto: 5' end (cReads mapping entirely on the 5' end half of the tRNA gene), mid-5' (cReads starting on the first 1/3 and ending before the last 1/3 of the tRNA gene length), 3' end (cReads mapping entirely on the 3' end half of the tRNA gene), mid-3' end (cReads starting after the first 1/3 and ending within the last 1/3 of tRNA gene length), mid (cRead overlaps both halves of the tRNA gene and not within the 1/3 extremity regions). The same method was used to calculate the cReads position within each rRNA feature for rRNA fragments found in exosomes.

The discovered 12 novel transcribed loci had their nucleotide sequences translated by the getorf program from the EMBOSS package (Rice, Longden *et al.* 2000) with the following paramaters: -minsize 33 -find 1 –noreverse, which sets a 10 amino acids minimum ORF length, translates solely from ATG to STOP codons and only on the three possible frames from the same strand where the exosome RNA reads mapped to, respectively. In order to check whether the putative ORFs outputted by the getorf program have similarity to any already known protein, we ran sensitive blastp against nr-NCBI (-word_size 2 -num_descriptions 5 -num_alignments 5 -evalue $1e^{-3}$) and no hits were found for any of them.

To determine whether there were transposable elements-derived RNA fragments within *Leishmania sp.* exosomes and also discard any possibility of cross-contamination between the libraries, we performed a BlastN search (Altschul, Madden *et al.* 1997) for *L. braziliensis*-specific SLACS/TATEs elements (extracted from TriTrypDB-6.0_LbraziliensisMHOMBR75M2904_ AnnotatedTranscripts.fasta downloadable file at tritrypdb.org). The following thresholds were applied during this screen: e-Value ≤ 1, identity ≥ 80% and query (cRead) coverage ≥70%.

To search for any sequences homologous to mammalian miRNAs within the *Leishmania sp.* exosomal RNA libraries, we ran blastn from the BLAST Plus package (Zhang, Schwartz *et al.* 2000) version 2.2.29+ querying the top thousand most abundant cReads on each library against the whole human and mouse miRNA dataset (hairpin

and mature) available at miRBase (mirbase.org). The blastn+ parameters were the following: -dust no -word_size 4 -evalue 1 -outfmt 6, and we also established a cutoff of 70% identity and 70% sequence coverage (ad-hoc PERL script). In a parallel approach to identify host genes that could potentially be targeted by putative regulatory RNAs in *Leishmania sp.* exosomes, we aligned the cReads from both libraries against human (hg19, NCBI) and vector (*Lutzomyia longipalpis* and *Phlebotomus papatasi,* https://www.vectorbase.org/) reference genomes. Bowtie2 version 2.1.0 (settings: --very-sensitive-local –N1) (Langmead and Salzberg 2012) and htseq-count script (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html) (using the option –s reverse, which reports reads mapping to annotated features on a reverse complement fashion) were used for this purpose.

### 4.3.8. Northern blotting

Aliquots of ~3 µg RNA per lane (*L. donovani* axenic amastigote total RNA from a single culture, or exosome RNA pooled from 4 separate exosome preparations) were loaded onto 8% denaturing polyacrylamide gels. Gels were stained with SYBR Green and visualized, then the samples were blotted onto Hybond $N^+$ nylon membrane (GE Healthcare). The membranes were UV cross-linked using a Stratalinker (1200 µJ for 30 seconds), blocked, probed and washed according to (Krieg 1996). Twenty one and 150 nt long *in vitro* transcribed RNAs were used as size markers. For hybridization, 5`end labelled DNA probes were used (LdBPK_291610_leftof probe 5' AAG GCG TCC CCA TGA TAA CG, LdBPK_301180_leftof probe 5' GAC CTC AAG TAT CTA CGG GAG A, tRNA-Asp probe ASP1 5' GGC GGG TAT ACT AAC CAC TAT AC, tRNA-Leu probe LEU1 5' AGA CCA CTC GAC CAT CTC A, tRNA-Leu probe LEU2 5' TGG AAC CTT AAT CCA ACG TCT T, 5.8S rRNA probe sequence was taken from (Dinhopl, Mostegl *et al.* 2011)). For 5' end labelling, the PNK labelling reaction was carried out as suggested by the manufacturer (NEB). The efficiency of [$\gamma$–$^{32}$P]-ATP incorporation was determined by running a small fraction of the PNK reaction on a native 20% polyacrylamide/urea gel and typically ranged between 80-95%, resulting in probes with high specific activity. The blotted membranes were placed in glass bottles containing a minimal amount of hybridization buffer (6X SSPE, 1% SDS, 2X Denhart's solution, 100 µg/ml of salmon sperm DNA; (Rose, Winston *et al.* 1990)) in a Hybaid$^{TM}$ mini oven MKII and pre-

hybridized with constant rotation for 4 hours at 37°C. After pre-hybridization, approximately 10 µCi of labelled probe was added and the membrane was hybridized for at least 18 hours at 37°C. The next day, the membrane was washed twice with a high stringency solution (2X SSPE and 0.1% SDS), and twice with a low stringency solution (0.2X SSPE and 0.1% SDS) for 15 min each at room temperature. The radioactive signal from the membranes was detected using a Storm 820 phosphorimager. Quantification of signals was performed in Imagequant.

### 4.3.9. Statistical analysis and graphs

R environment version 3.0 was used to generate read length distribution histograms for each library (calculating their mean and median values), as well as to perform the Pearson's Correlation analysis regarding the tRNA-derived small RNA reads abundance and the amino acid usage frequency of the respective predicted proteomes. Other graphs were generated with GraphPad Prism 4.0 and EXCEL.

## 4.4. Results

### 4.4.1. *L. donovani* and *L. braziliensis* exosomes contain short RNA sequences; and intact vesicles protect their RNA cargo from degradation

We have previously reported that *Leishmania sp.* use an exosome-based secretion mechanism in order to export proteins with potential virulence properties (Silverman, Chan *et al.* 2008, Silverman, Clos *et al.* 2010). Based on a number of studies in mammalian systems demonstrating the presence of RNA in exosomes, we were encouraged to expand on our findings and examine the RNA content of *Leishmania sp.* exosomes. We performed all experiments for this study with exosomes purified from supernatants of *L. donovani* or *L. braziliensis* cultured *in vitro* under infection-like stressors (acidic pH and elevated temperature for 24 h, see Methods), which induce the cells to transform into amastigotes. We had previously observed that these “early” axenic amastigotes release increased quantities of exosomes enriched in specific virulence factors (Silverman, Clos *et al.* 2010). Moreover, while undergoing

transformation into amastigotes, *Leishmania sp.* modulate critical macrophage processes to allow for establishment of chronic infection. Combined with the fact that amastigotes are literally the only life cycle stage found *in vivo* in the mammalian host once infection is established, we felt that exosomes purified from this life cycle stage were the most relevant to examine.

Exosomes were purified from supernatants of early axenic amastigotes and subjected to RNA extraction with phenol/chloroform. Results depicted in Figure 4.1A and 1B show that *L. donovani* axenic amastigote exosomes contained significant amounts of RNA that were detectable with the Agilent Bioanalyzer. Quantification with nanodrop revealed an average yield of 12.5 ng of RNA per µg of exosomal protein (data not shown). Notably, the length profile of exosomal RNA was distinct from that of *L. donovani* total RNA, with the bulk of exosomal sequences being short (25-250 nt). Furthermore, we did not detect full length ribosomal RNA (rRNA) peaks in exosome RNA profiles. In contrast, these full length rRNA peaks were prominent in the total RNA profiles. To confirm that the purified nucleic acid was in fact RNA, we incubated exosome RNA with DNase, RNase or KOH. As can be seen in Figure 4.1C, exosomal RNA was resistant to treatment with DNase, but was completely degraded upon exposure to either RNase or KOH.

To exclude the possibility that RNA was merely co-purified during exosome isolation but was not directly associated with or internal to the vesicles, we treated intact exosomes with RNase in the presence or absence of membrane-permeabilizing detergent. The results in Figure 4.1D show that when treated with RNase alone, exosome RNA remained intact. In contrast, when exosomes were treated with RNase and TritonX-100 simultaneously, the RNA signal was greatly diminished. These findings suggested that the RNA was confined within the exosomal membrane and thereby protected from degradation. The fact that we still saw a small residual signal after detergent and RNase treatment could indicate that a fraction of the RNA was bound to RNA-binding proteins and was thereby protected.

To investigate whether the release of RNA within exosomes is conserved between *Leishmania* species, we purified and analyzed RNA from exosomes released

by *L. braziliensis* early axenic amastigotes using the same procedures as described for *L. donovani*. This analysis showed that *L. braziliensis* exosomes also contain RNA, with similar characteristics to that of *L. donovani* exosome RNA (Supplementary* Figure B 1A and 1B in Appendix B, page 207). Taken together, these data represent the first description of RNA released by *Leishmania sp.* within exosomes.

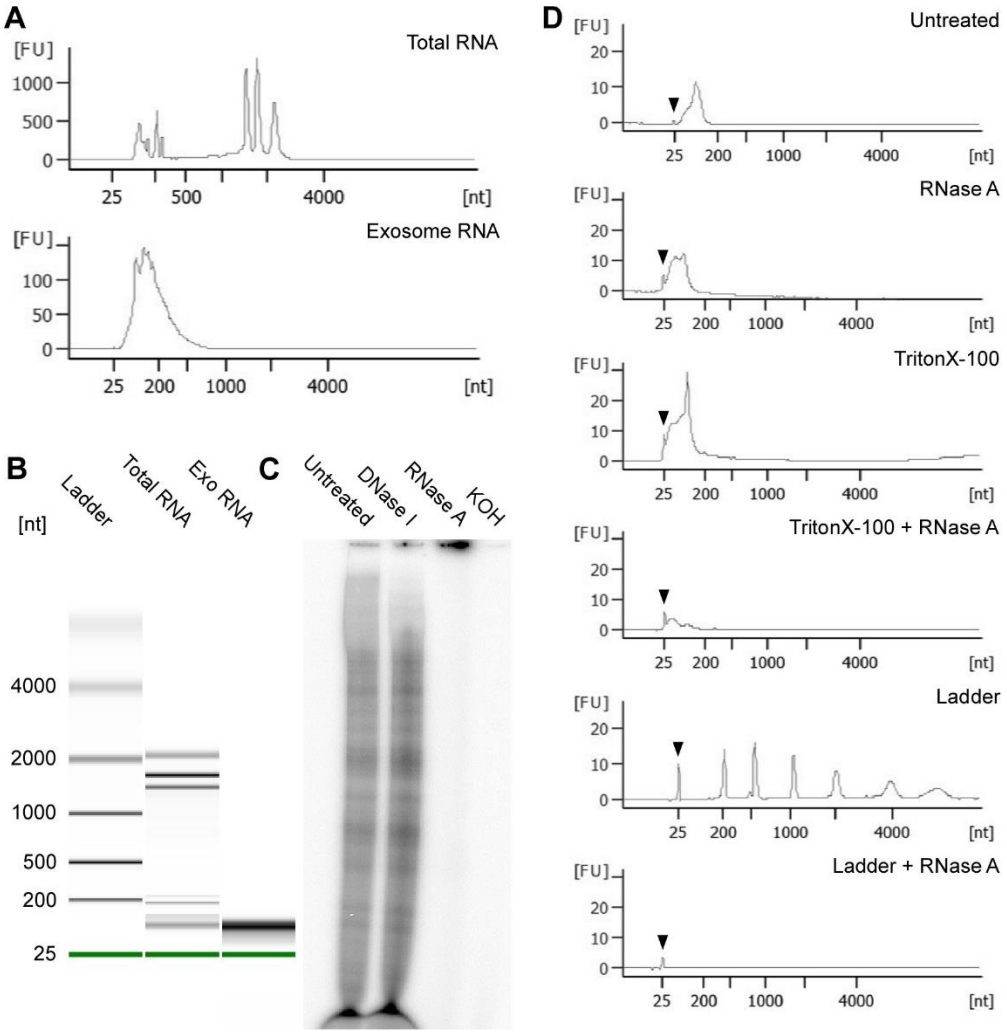



***Figure 4.1. L. donovani exosomes contain RNA cargo***

Exosomes were purified from *L. donovani* axenic amastigote culture supernatant as described in the Materials and Methods. RNA was extracted from exosomes or whole cells by phenol-chloroform extraction and then analyzed. A) Agilent Bioanalyzer RNA length profiles of exosome RNA alongside total RNA (~100 ng RNA were loaded for each). B) Gel-like image from Agilent Bioanalyzer measurement. C) Purified exosome RNA (~250 ng/sample) was either left untreated or treated with DNase I, RNase A or KOH followed by radiolabelling with [$\gamma$–$^{32}$P]-ATP and separation on a denaturing 15% polyacrylamide gel. D) RNA inside exosomes is resistant to degradation. Prior to RNA extraction, intact exosomes (purified from 400 ml culture supernatant) were either left untreated, or treated with RNase A or TritonX-100 or both. As a control for RNase A activity, 1 µl of the Agilent pico ladder was treated with the same concentration of RNase A. Samples were then subjected to RNA extraction and run on the Agilent Bioanalyzer. Arrowhead indicates internal 25 nt marker. nt, nucleotides. Exposure of the RNA ladder to RNase A was included as a positive control. All images are representative of at least 3 independent experiments.

### 4.4.2. *Leishmania sp.* exosomes deliver RNA cargo to human macrophages

In our previous studies, we observed that *Leishmania sp.* exosomes were released into infected macrophages, were taken up by uninfected bystander cells, and that exosomal proteins were delivered to host macrophage cytoplasm (Silverman, Clos *et al.* 2010). In order to investigate the potential delivery of exosomal RNA cargo to host macrophages, we labelled exosomes purified from the supernatant of *L. donovani* early amastigotes with an RNA specific fluorescent dye. Size and homogeneity of exosomes was assessed by Nanosight analysis (see Figure 4.2A) and the median size was determined to be 120 nm. Fluorescence of labeled exosomes was confirmed by microscopy (Supplementary Figure B2). PMA-differentiated THP-1 cells were incubated for 2 hours with fluorescently labelled exosomes and uptake was assessed by flow cytometry and confocal microscopy. As shown in Figure 4.2B, we observed a dose-dependent increase in fluorescence of cells, suggesting that macrophages readily take up exosomes and their RNA cargo. In contrast, control cells incubated at 4°C to inhibit phagocytosis showed only background fluorescence. To exclude the possibility that exosomes were just bound to the macrophage membrane but not internalized after incubation, we examined exosome-treated cells by confocal microscopy. Figure 4.2C shows that the fluorescence was localized to the cytoplasm of the macrophages and not to the membrane, indicating that the exosomes containing RNAs were indeed taken up by the cells. These results confirm that *Leishmania sp.* exosomes and their RNA cargo can be internalized by host cells and can access their cytoplasm.

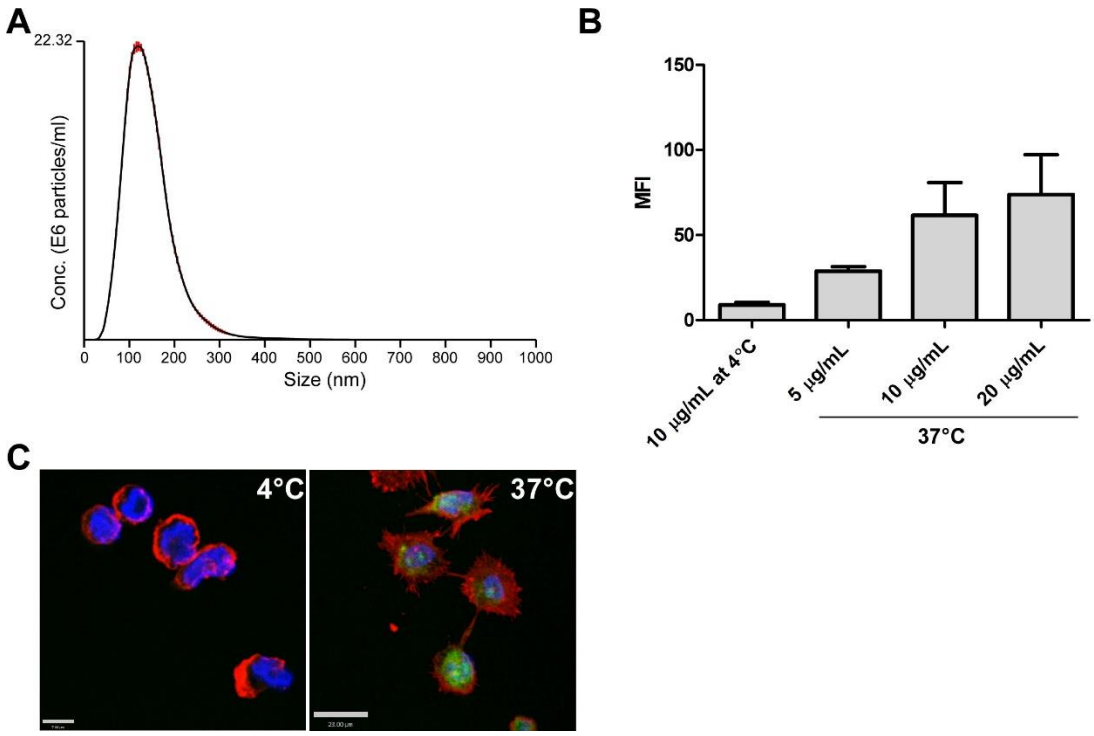



***Figure 4.2. Exosomal RNA cargo is delivered to macrophages***

Exosomes were purified from 400-800 ml supernatant of *L. donovani* axenic amastigotes, protein concentration was determined by Micro BCA, and exosomes were stained with a green fluorescent RNA-specific dye. PMA-differentiated THP-1 cells were incubated for 2 hours with labelled exosomes at either 37°C or 4°C to inhibit phagocytosis. A) Nanosight size profile of purified exosomes. B) Cells were treated with different concentrations of labelled exosomes as indicated and analysed by flow cytometry. Histograms were drawn, median fluorescence intensity (MFI) of cells was recorded, and the mean of 3 independent experiments was calculated (error bars represent standard error of the mean, SEM). C) Confocal microscopy of cells incubated with 10 µg/ml exosomes (green) at 4°C (left) or 37°C (right). Cells were stained with phalloidin-Alexa 594 to detect actin (red) and DAPI to detect nuclei (blue). Confocal microscopy was done with a Leica DMIRE2 inverted microscope equipped with a SP2 AOBS laser scanning head. Images were taken with a 63x magnification objective. Images are representative of 3 independent experiments.

### 4.4.3. Characterization of *Leishmania sp.* exosome RNA cargo: Exosomes are enriched in small non-coding RNAs derived from tRNAs and rRNAs

In order to assess the global transcriptome present in *Leishmania sp.* exosomes, we constructed complementary DNA libraries for high-throughput sequencing. We chose to compare RNA purified from exosomes released by early axenic amastigotes of *L. donovani* and *L. braziliensis* for three reasons: a) these two organisms cause distinct

disease manifestations and hence can be expected to differ in their mechanisms of pathogenesis; b) they are spread through different vectors: *L. donovani* is transmitted by sandflies of the genus Phlebotomus in the Old World, whereas *L. braziliensis* is transmitted by Lutzomyia in the New World; and c) *L. braziliensis* was found to have a functional RNA interference pathway, which seems to be absent in *L. donovani* (Lye, Owens *et al.* 2010). We hypothesized, therefore, that these two organisms could differ in their composition of exosomal RNA and chose to examine this directly. We used a strategy for library construction that was optimized for sequencing of small RNAs, as we had observed by gel electrophoresis that the exosomal RNA sequences were mainly short (Figure 4.1 and Supplementary Figure B1). We also incorporated a series of enzymatic treatments including dephosphorylation with calf intestinal alkaline phosphatase (CIP), 5' cap removal with tobacco acid phosphatase (TAP) and 5' re-phosphorylation with polynucleotide kinase (PNK) into the library construction procedure in order to pick up all sequences present in the exosomal transcriptome regardless of their 5' modification (see Materials and Methods). Sequencing of the libraries by paired end 150 bp MiSeq Illumina sequencing resulted in ~1.4 million paired reads for *L. donovani* and ~1.1 million paired reads for *L. braziliensis* (Table 4.1). After adapter trimming and adjustment of the orientation of all reads to correspond to that of the original RNA sequence, reads were collapsed into unique reads prior to further analysis. As shown in the histograms in Figure 4.3A, read length distributions of reads were clearly skewed towards shorter reads with the mean read length being 55 nt for *L. donovani* and 57 nt for *L. braziliensis* (medians 37 nt and 49 nt, respectively).

***Table 4.1. Sequencing statistics for exosomal RNA libraries from L. donovani and L. braziliensis***

Numbers of reads for *L. donovani* and *L. braziliensis* exosome RNA libraries as obtained by high-throughput sequencing. Reads were combined into unique reads by collapsing all identical reads to one read for downstream analysis. This also revealed the number of reads that were present in the dataset as a single copy.

| | **L. donovani library** | **L. braziliensis library** |
|---|---|---|
| Total paired reads | 1435277 | 1062571 |
| Collapsed/unique reads | 688524 | 538034 |
| Unique and single copy | 574049 | 421086 |

To get a general overview about what types of RNA transcripts were represented in our libraries, we aligned the reads of the *L. donovani* and the *L. braziliensis* libraries with reference genomes, respectively LdBPK (*Leishmania donovani* strain BPK282A1) and LbrM (*Leishmania braziliensis* MHOM/BR/75/M2904) using Bowtie 2 and the very-sensitive-local option which sets the seed length to 20 nucleotides, allowing for only one mismatch within the seed alignment (see Methods). We were able to align 58.61 % of the reads from the *L. donovani* library to the LdBPK reference genome and 22.87 % of the reads from the *L. braziliensis* library were aligned to the LbrM reference genome (see Supplementary Tables 1A and 1B for the full datasets). The comparatively low alignment rate especially in case of the *L. braziliensis* library is likely a result of incomplete assembly of the reference genome or the fact that we used a different *L. braziliensis* strain (a clinical isolate from the Peruvian Amazon region) than the strain used to generate the reference genome. This is supported by the fact that we were able to align significantly more reads (52.88%) from the *L. braziliensis* library with the *L. major* reference genome (*Leishmania major* MHOM/IL/81/Friedlin, LmjF, see Supplementary Table 4.1B). Other possible causes for low alignment rates could be misinterpretation of modified nucleosides by the sequencer or RNA editing of sequences prior to packaging into exosomes, making it difficult to compare our transcriptomic data with the available reference genomes derived from DNA sequencing. RNA editing is a well-described process in *Leishmania sp.* and other trypanosomatids (e.g. (Maslov 2010, Bayer-Santos, guilar-Bonavides *et al.* 2013)).

In order to ensure that our libraries were not contaminated with unrelated nucleic acids, we performed a BLAST search of all reads that failed to align with either the LdBPK, the LbrM or the LmjF reference genomes, against the NCBI nucleotide collection database (NCBI-NT). The results of this analysis showed that 28.1% of reads from the *L. donovani* exosome library and 36.3% of reads from the *L. braziliensis* exosome library aligned to sequences in the NCBI-NT database (see Supplementary Table 4.2A and 2B). Of these, 4.93% of *L. donovani* and 4.17% of *L. braziliensis* aligned with other *Leishmania sp.* genomes. The rest aligned with a promiscuous group of >6000 different plant, fungi, helminth and bacteria species, several of which were plant pathogens or soil inhabitants. Based on the observation that there was no enrichment of any particular species and that overall, the majority of reads from both libraries aligned with

*Leishmania sp.* genomes (in total 63.54% of reads of the *L. donovani* library and 57.05% of the *L. braziliensis* library, see summary of alignment statistics in Supplementary Table 4.3), we concluded that we did not have a contamination issue that would impugn our data. We think that many, if not all, of the reads mapping to bacteria or helminth genomes are likely false positive hits. Thus, even though our alignment rates were somewhat lower than we might have expected, we think that our datasets are valid and large enough to draw meaningful conclusions about the exosomal RNA content.

When categorizing reads into RNA biotypes based on reference genome annotations, we saw that for both libraries, the majority of reads were aligning with rRNA and tRNA genes, in the sense orientation (Figure 4.3B). In addition, a large number of the reads mapped to non-annotated (intergenic) regions of the reference genomes (42.47% for *L. donovani* and 34.46% for *L. braziliensis*) could potentially be novel transcripts. Interestingly, we only saw less than 4% of reads mapping to protein coding genes (CDS) or spliced leader (SL) RNA genes. These results described above indicated that the majority of sequences present in the *Leishmania sp.* exosome transcriptome are derived from non-coding RNAs and intergenic regions, while sequences derived from mRNAs are under represented.

### *Table 4.2. Top 20 most abundant clusters of transcripts present in Leishmania* sp. *exosomes*

Reads were clustered into genomic loci based on Bowtie 2 alignments with reference genomes (as described in Materials and Methods) to identify the RNA biotypes that were most abundant in exosomes. The details of the top 20 clusters with the highest numbers of reads falling into them are listed. Clusters of reads in the *L. donovani* library are listed in descending order of abundance, with the homologous cluster of reads in the *L. braziliensis* library given in the same row. Chr = chromosome number, annotation = annotation in reference genomes (LdBPK = *L. donovani*, LbrM = *L. braziliensis* or LmjF = *L. major*) followed by the gene name, No. of reads = # of reads from the library falling into this cluster, RNA biotype = type(s) of RNA that is annotated in the reference genomes in this region.

| ***L. donovani*** | | | | | | ***L. braziliensis*** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Chr** | **Coordinates of genomic locus** | | **Annotation** | **No. of reads** | **RNA biotype** | **Chr** | **Coordinates of genomic locus** | | **Annotation** | **No. of reads** | **RNA biotype** |
| 27 | 1014367 | 1019133 | LdBPK_27rRNA3<br>LdBPK_27rRNA4<br>LmjF.27.rRNA.13<br>LmjF.27.rRNA.22<br>LmjF.27.rRNA.29<br>LmjF.27.rRNA.31<br>LmjF.27.rRNA.33<br>LmjF.27.rRNA.42 | 344191 | 28S rRNA | 6 | 334041 | 334897 | LmjF.27.rRNA.31<br>LmjF.27.rRNA.34 | 58906 | 28S rRNA |
| 27 | 1019947 | 1021495 | LdBPK_27rRNA6 | 132730 | 18S rRNA | 00 | 463646 | 464099 | LbrM.27.rRNA1 | 11166 | 18S rRNA |
| 15 | 312758 | 313248 | LmjF.15.TRNAASP.01<br>LmjF.15.TRNAGLU.01<br>LmjF.09.5SrRNA.02<br>LmjF.05.5SrRNA.01<br>LmjF.15.5SrRNA.01 | 65737 | tRNA-Asp<br>tRNA-Glu<br>5S rRNA | 15 | 324587 | 325394 | LbrM.15.tRNA1<br>LbrM.15.tRNA2<br>LbrM.15.rRNA1 | 18229 | tRNA-Asp<br>tRNA-Glu<br>5S rRNA |
| 24 | 715730 | 715801 | LdBPK_24tRNA5 | 43207 | tRNA-Asp | 24 | 659346 | 659417 | LbrM.24.tRNA5 | 30583 | tRNA-Asp |
| 17 | 328838 | 328909 | LdBPK_17tRNA1 | 42601 | tRNA-Asp | 17 | 296223 | 296604 | LbrM.17.tRNA1<br>LbrM.17.tRNA2<br>LbrM.17.tRNA3 | 30523 | tRNA-Asp<br>tRNA-Ser<br>tRNA-Ala |
| 24 | 658796 | 658976 | LdBPK_24tRNA2 | 35611 | tRNA-Gln | 24 | 600448 | 600615 | LbrM.24.tRNA2 | 8749 | tRNA-Gln |
| 09 | 429809 | 430355 | LdBPK_09tRNA6 | 29246 | tRNA-Glu | 09 | 395278 | 395779 | LbrM.09.tRNA3 | 4467 | tRNA-Val |

| *L. donovani* | | | | | | *L. braziliensis* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Chr** | **Coordinates of genomic locus** | | **Annotation** | **No. of reads** | **RNA biotype** | **Chr** | **Coordinates of genomic locus** | | **Annotation** | **No. of reads** | **RNA biotype** |
| | | | LmjF.09.TRNAARG.01<br>LmjF.09.TRNAVAL.02<br>LmjF.05.5SrRNA.01<br>LmjF.11.5SrRNA.03<br>LmjF.21.5SrRNA.01 | | tRNA-Arg<br>tRNA-Val<br>5S rRNA | | | | LbrM.09.tRNA4<br>LbrM.09.rRNA1<br>LmjF.05.5SrRNA.01<br>LmjF.11.5SRRNA.03<br>LmjF.21.5SrRNA.02 | | tRNA-His<br>5S rRNA |
| | | | | | | 09 | 403494 | 403565 | LbrM.09.tRNA5 | 5226 | tRNA-Glu |
| 31 | 495812 | 496115 | LdBPK_31_tRNA3 | 18528 | tRNA-Glu | 31 | 582437 | 582738 | LbrM.31.tRNA2<br>LbrM.31.tRNA3 | 4719 | tRNA-Gly<br>tRNA-Glu |
| 27 | 1019543 | 1019804 | LdBPK_27rRNA5 | 18473 | 5.8S rRNA | | | | | | |
| 11 | 156707 | 157038 | LmjF.11.TRNAALA.01<br>LmjF.36.TRNALEU.01 | 15493 | tRNA-Ala<br>tRNA-Leu | 11 | 63421 | 63678 | LmjF.33.TRNAALA.01<br>LmjF.11.TRNALEU.02 | 3041 | tRNA-Ala<br>tRNA-Leu |
| 27 | 1014054 | 1014340 | LmjF.27.rRNA.47<br>LmjF.27.rRNA.48 | 14260 | 28S rRNA | | | | | | |
| 23 | 230438 | 230509 | LdBPK_23tRNA9 | 13814 | tRNA-Gly | 23 | 216842 | 216916 | LbrM.23.tRNA9 | 3679 | tRNA-Gly |
| 36 | 1630332 | 1630403 | LdBPK_36tRNA2 | 11529 | tRNA-Gln | | | | | | |
| 05 | 360707 | 361335 | LdBPK_05snRNA1 | 9971 | snRNA | 05 | 349991 | 350587 | LbrM.05.rRNA1-1<br>LbrM.05.ncRNA1-1 | 6306 | 5S rRNA<br>ncRNA |
| 33 | 104560 | 104930 | LdBPK_33tRNA1<br>LdBPK_33tRNA2<br>LdBPK_33tRNA3 | 9352 | tRNA-Ala<br>tRNA-Arg | 33 | 105787 | 105859 | LbrM.33.tRNA1 | 5730 | tRNA-Arg |
| 23 | 229645 | 229857 | LdBPK_23tRNA5<br>LdBPK_23tRNA6 | 8487 | tRNA-Leu<br>tRNA-Thr | 35 | 2472707 | 2472788 | LbrM.35.tRNA4 | 2592 | tRNA-Leu |
| 16 | 445957 | 446028 | LdBPK_161tRNA1 | 7916 | tRNA-Gln | 16 | 442089 | 442160 | LbrM.16.tRNA1 | 8695 | tRNA-Gln |
| 23 | 230585 | 230656 | LdBPK_23tRNA10 | 7804 | tRNA-Trp | 23 | 216992 | 217063 | LbrM.23.tRNA10 | 3657 | tRNA-Trp |
| | | | | | | 21 | 430678 | 430798 | LbrM.21.rRNA1 | 3273 | 5S rRNA |

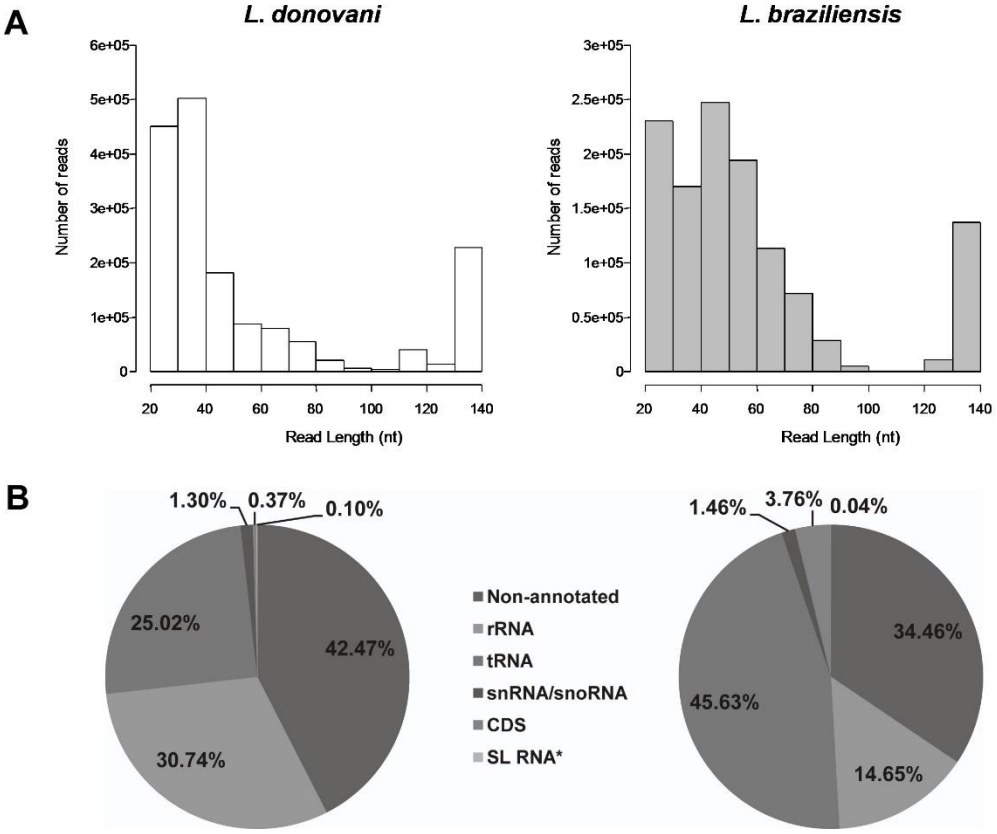



***Figure 4.3. Sequencing of Leishmania* sp. *exosomal RNA reveals conserved RNA cargo composed mainly of sequences derived from non-coding RNA***

Exosome RNA from *L. donovani and L. braziliensis* was purified and processed for high-throughput sequencing as described in Methods. A) Sequence length distribution of reads obtained from sequencing *L. donovani* and *L. braziliensis* exosome libraries. B) Categorization of reads according to their alignment with genomic features annotated in the *L. donovani* and *L. braziliensis* reference genomes. CDS, coding sequence; *SL RNA, spliced leader RNA. Numbers for reads mapping to SL RNA genes were obtained from alignment with the *L. major* reference genome, as these genes have currently only been annotated in this genome.

When working with the LdBPK and LbrM reference genomes, we had to take into account that both are limited in their annotations. Hence, it was not surprising that we found a large number of reads in our libraries mapping to intergenic regions. Whereas the annotations of CDS are thought to be comprehensive in these genomes, the assignments of SL RNAs as well as structural non-coding RNAs such as rRNAs, tRNAs,

snRNAs and snoRNAs are clearly lacking in completeness. Consequently, it has to be considered that the large number of reads mapping to intergenic regions may not necessarily all be novel transcripts, but could also have resulted from incomplete annotation of non-coding RNA types in these regions. Keeping this in mind and still trying to dissect what types of RNA sequences are highly represented in exosomes, we decided to inspect in greater detail the alignment of the most abundant exosomal sequences manually using the Artemis genome browser software (Carver, Harris *et al.* 2012). For this purpose, reads were clustered into unique regions of alignment and the regions were ranked by abundance (number of reads found per region). Considering that *L. major* is the species with the best assembled genome to date and presents the most complete annotation of non-coding RNAs, we also performed alignments of *L. major* annotated non-coding RNAs with the LdBPK and LbrM reference genomes, in order to identify non-annotated, non-coding RNA loci in our target genomes. The results of the screening using Artemis showed that the top 20 most abundant reads from both libraries mapped to three RNA classes in the sense orientation: rRNA, tRNA and snRNA (Table 2).

The high abundance of reads mapping to rRNA genes observed in both libraries is in compliance with other recent reports on RNA types found in exosomes. Upon closer inspection we saw that the majority of reads mapping to rRNA genes were shorter fragments (median length 39 nt for the *L. donovani* library and 52 nt for *L. braziliensis*, see Supplementary Figure 4.3). We then looked for enrichment of specific rRNA genes within our pool and saw that the majority of reads from both libraries mapped to 28S and 18S rRNA genes (>90%, Supplementary Table 4.4). Furthermore, we investigated the position of reads aligned within the various rRNA genes, and found that for both libraries, those reads covered the entire length of the rRNA genes (Supplementary Table 4.4). It was of particular interest to find a large number of reads mapping to tRNAs in both libraries, as tRNA-derived small RNAs have recently been discovered in *T. cruzi* (Garcia-Silva, Frugier *et al.* 2010, Franzen, Arner *et al.* 2011) , and these novel small RNAs are thought to participate in regulation of gene expression (Sobala and Hutvagner 2011, Garcia-Silva, Sanguinetti *et al.* 2014) (see below for a more detailed analysis of tRNA-derived small RNAs).

Notably, to our surprise, the overlap of the RNA profiles for *L. donovani* and *L. braziliensis* was striking. Thus, these parallel and independent RNA-seq replicates provide direct evidence for the reproducibility of our data.

### 4.4.4. Exosomes carry putative novel transcripts

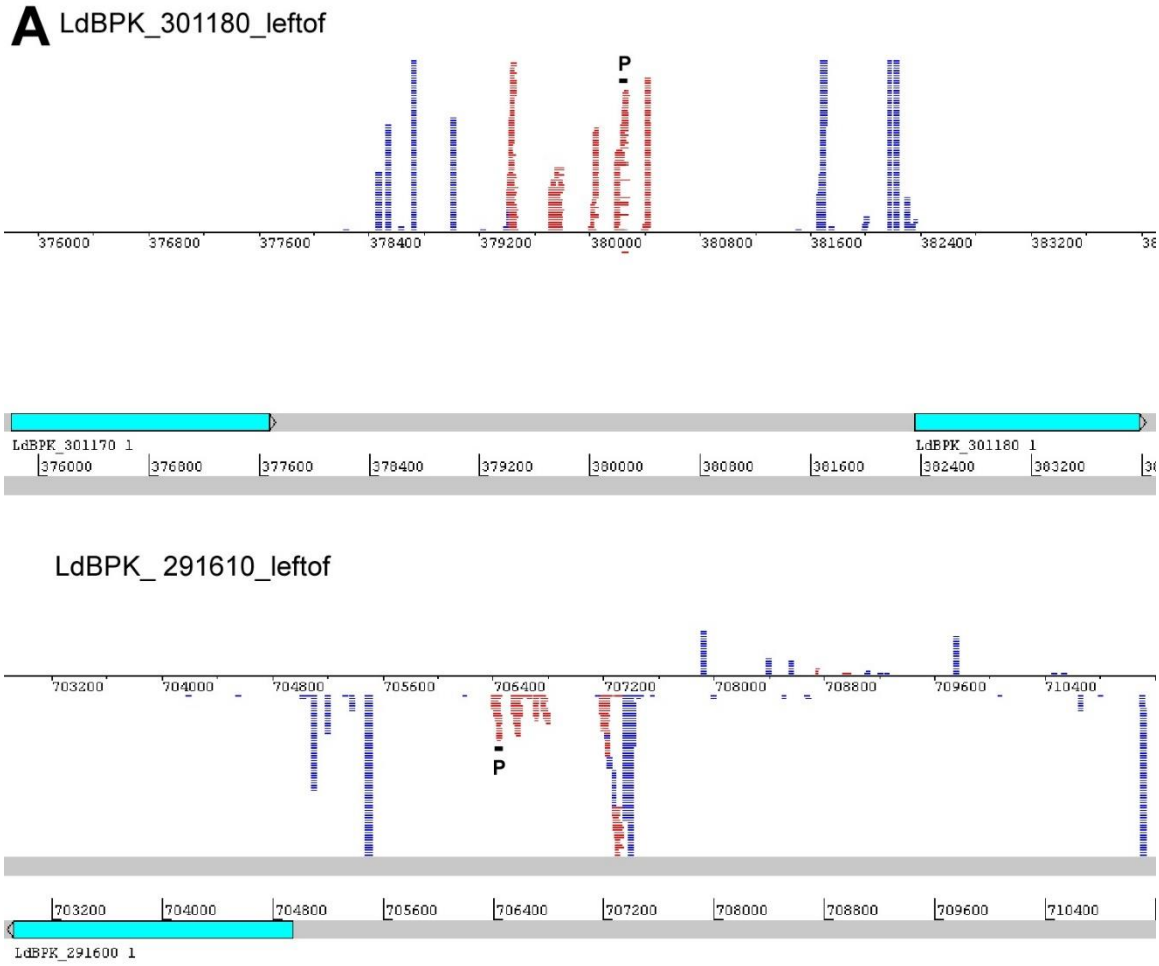
To make sure we did not miss any important information amongst the group of less abundant reads, we randomly selected a number of less abundant reads from both libraries and inspected their alignment with the reference genomes manually. Interestingly, we discovered several reads mapping to intergenic regions at different genomic loci (Table 4.3). These intergenic regions were neither annotated at those loci in any of the sequenced *Leishmania sp.* or trypanosome genomes, nor did they share homology to any known trypanosomatid gene (as assessed by performing BLAST searches on TriTrypDB and NCBI). These findings suggested that the sequences mapping to these regions corresponded to *bona fide* novel transcripts. Notably, we found homologous novel transcripts in both libraries, providing evidence that the transcripts are conserved between species as well as packaged into exosomes. When overlaying our sequencing data from the *L. donovani* exosome RNA library with a recently sequenced *L. donovani* spliced leader (SL) RNA library (P. Myler, unpublished data), we observed that the genomic loci giving rise to our identified novel transcripts had SL sites in the 5` region upstream of them (See Figure 4.4 for 2 examples). This indicates that they might be processed by *trans*-splicing and are hence likely to be functional mature transcripts rather than promiscuous transcriptional by-products.

We also searched for open reading frames (ORF) within the sequences of the novel transcripts to see whether they have the potential to code for a protein or peptide and found potential ORFs for the majority of them (Supplementary Table 4.5). However, when we translated the ORFs and looked for homologies to known proteins in the NCBI database using Blastp, we did not obtain any hits.

## Table 4.3. Intergenic regions coding for putative novel transcripts in exosomes

List of 12 intergenic regions identified with numbers of reads mapping to them listed by descending order of abundance in the *L. donovani* library, with the homologous genomic region in the *L. braziliensis* library given in the same row. Names are derived from the annotated genes adjacent to the intergenic region plus the designation "_leftof", indicating that the intergenic region is on the left side of the annotated gene on the same strand, regardless of transcriptional direction. There is 75% overlap of intergenic regions coding for novel transcripts in the *L. donovani* and *L. braziliensis* libraries. ORF, open reading frame. N.A., not applicable.

| **L. donovani** | | | | | **L. braziliensis** | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Name** | **Coordinates** | | **No. of reads** | **No. of ORF** | **Name** | **Coordinates** | | **No. of reads** | **No. of ORF** |
| LdBPK_301180_leftof | 379397 | 380435 | 363 | 7 | LbrM.30.1240_leftof | 394563 | 397304 | 317 | 13 |
| LdBPK_291610_leftof | 706385 | 707360 | 181 | 9 | LbrM.29.1600_leftof | 663702 | 664543 | 103 | 7 |
| LdBPK_360420_leftof | 109068 | 109733 | 139 | 3 | LbrM.35.0480_leftof | 132138 | 133299 | 185 | 5 |
| LdBPK_36300/0_leftof | 1183324 | 1183522 | 132 | 1 | LbrM.35.3080_leftof | 1176690 | 1176833 | 36 | 1 |
| LdBPK_362290_leftof | 872006 | 872406 | 89 | 5 | LbrM.35.2400_leftof | 892711 | 895894 | 29 | 28 |
| LdBPK_313190_leftof | 1452630 | 1452719 | 83 | 1 | LbrM.31.3490_leftof | 1508257 | 1510948 | 0 | N.A. |
| LdBPK_040550_leftof | 225336 | 230737 | 80 | 8 | LbrM.04.0610_leftof | 229925 | 233378 | 111 | 25 |
| LdBPK_131560_leftof | 555895 | 556192 | 75 | 2 | LbrM.13.1200_leftof | 433509 | 433481 | 0 | N.A. |
| LdBPK_364270_leftof | 1570740 | 1570991 | 57 | 1 | LbrM.35.4310_leftof | 1563654 | 1563940 | 29 | 2 |
| LdBPK_366120_leftof | 2270195 | 2272110 | 49 | 13 | LbrM.35.6160_leftof | 2242345 | 2244214 | 153 | 14 |
| LdBPK_330560_leftof | 173136 | 173362 | 40 | 0 | LbrM.33.0550_leftof | 184438 | 184751 | 19 | 1 |
| LdBPK_366590_leftof | 1903364 | 1906720 | 0 | N.A. | LbrM.35.6630_leftof | 2438694 | 2438825 | 155 | 0 |


A
LdBPK_301180_leftof
P
376000
376800
377600
378400
379200
380000
380800
381600
382400
383200
LdBPK_301170 1
LdBPK_301180 1
LdBPK_ 291610_leftof
703200
704000
704800
705600
706400
707200
708000
708800
709600
710400
P
LdBPK_291600 1

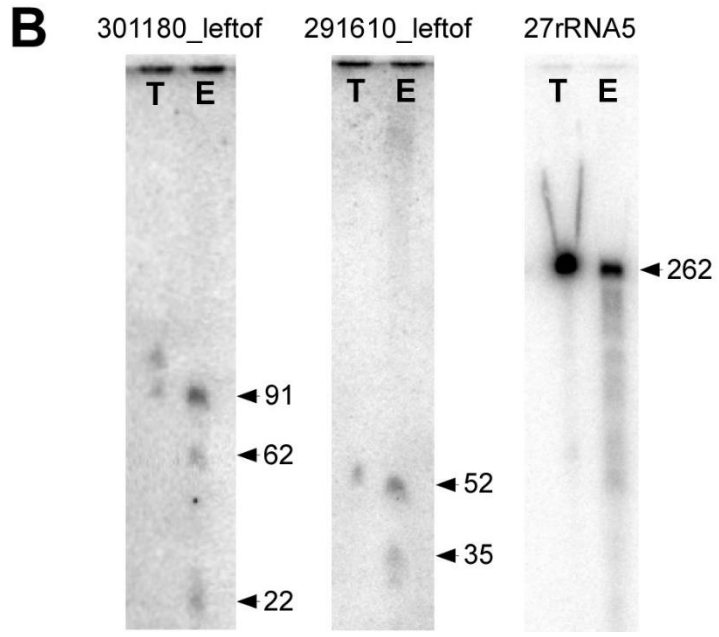

B
301180_leftof
291610_leftof
27rRNA5
T E
T E
T E
91
62
22
52
35
262

***Figure 4.4. Novel transcripts are found in Leishmania* sp. *exosomes***

A) Artemis genome browser alignments of the *L. donovani* exosome library and a *L. donovani* spliced leader sequence library (P. Myler, unpublished) with the *L. donovani* reference genome. Shown are two regions with reads mapping to them. Top: intergenic region on chromosome 30 (LdBPK_301180_leftof), bottom: intergenic region on chromosome 29 (LdBPK_291610_leftof). Light blue boxes on grey tracks are annotated genes. In dark blue are the reads from the spliced leader library, in red the reads from the exosome libraries. "P" designates the regions that were used for designing probes for Northern blotting. B) Northern blots with probes designed for novel transcripts found in exosomes, corresponding to the genomic regions shown in panel A (301180_leftof and 291610_leftof), plus an additional probe against a 5.8S rRNA (27rRNA5). *L. donovani* total (T) and exosome (E) RNA were probed on the same membrane. Equal amounts of RNA (3 µg) were loaded in each lane. Sizes of bands on membranes as indicated are in nucleotides (nt) and were calculated based on 262 nt, 150 nt and 21 nt size markers.

Based on the hypothesis that these novel transcripts could have a role in regulation of gene expression in either the mammalian or insect host or both, we performed Bowtie2 alignments to the human and the vector (Lutzomyia and Phlebotomous) genomes to search for potential targets. Out of the 1288 reads representing novel transcripts in the *L. donovani* library, we obtained 60 hits when searching against the human genome and 15 hits when searching against the Phlebotomus genome (Supplementary Table B6A). For the 1137 reads that comprise the novel transcripts in the *L. braziliensis* library, 25 hits were observed when searching against the human genome and 6 hits when searching against the Lutzomyia genome (Supplementary Table B6B). However, nearly all of these hits were determined to be in non-annotated regions of the genomes, implying that, at least in the human genome, there are no genes that could be regulated by the novel transcripts (analyses based on perfect complementarity). Of note, the annotation of the vector genomes is a very recent effort and far from complete. It is quite possible, therefore, that there are as yet non-annotated protein coding genes in the genomic regions where the novel exosomal transcripts aligned, which would have been missed, resulting in false negative findings. Moreover, in most animals, regulatory RNAs such as miRNAs have incomplete homology with their target sequences (Didiano and Hobert 2006) and, therefore, our predictions based on perfect complementarity may have missed some potential targets in the host genomes, again leading to false negative results. Unfortunately, as the tools to predict RNA-RNA interactions at the level of potential regulatory RNA-mRNA target pairs are fairly limited (generating a massive amount of ambiguous results when analysing large datasets), we were unable to carry out a comprehensive host mRNA

target prediction with the novel transcripts that was informative. We also performed an alignment search of the novel transcripts against the databases of human and mouse miRNAs (mirbase.org), but failed to detect homologous sequences. All of the novel transcripts we identified were present in the sense orientation of transcription in *Leishmania sp.*, which implies that they are unlikely to be present in exosomes as double strands, which is a characteristic of canonical siRNAs and miRNAs.

To validate the presence of the identified novel transcripts in exosomes and compare their expression in exosomes with total *Leishmania sp.* RNA, we designed probes for Northern blots. We selected the two most abundant novel transcripts identified in the *L. donovani* exosome library, one of which was positioned in between the genes 1170 and 1180 on chromosome 30 (LdBPK_301180_leftof) and the other in between the genes 1600 and 1610 on chromosome 29 (LdBPK_291610_leftof) for probe design. The regions that are complementary to the probes used are indicated in Figure 4.4A. We loaded equal amounts of *L. donovani* total and exosome RNA into a polyacrylamide gel, along with 21 nt and 150 nt size markers. Of interest, we detected bands of larger size that were present in both total and exosome RNA (which likely represent the primary transcript), but where the signal appeared to be stronger in the exosomal RNA lane (Figure 4.4B). In addition, we detected bands of smaller size in the exosome RNA, that were completely absent in the total RNA. For comparison, we also incubated a blot with a probe for 5.8S rRNA (LdBPK_27rRNA5, Ref), which appeared to be much more abundant in total than in exosome RNA. These results confirm the presence of the novel transcripts identified by sequencing in the exosomes and indicate that these transcripts with specific lengths, are uniquely present in exosomes and agrees with the idea of selective packaging.

### 4.4.5. *L. braziliensis* exosomes carry a low abundance of sequences derived from siRNA-coding regions

As discussed above, *L. braziliensis* can regulate gene expression through the RNAi pathway and produce small interfering RNAs (siRNAs) (Didiano and Hobert 2006). We were interested in exploring the possibility to find siRNAs as part of *L. braziliensis* exosomal RNA cargo. The main classes of *L. braziliensis* siRNAs are derived from the

spliced leader-associated conserved sequence (SLACS) retroposon, the telomere-associated transposable element (TATE) (Peacock, Seeger *et al.* 2007) , the *L. braziliensis*-specific telomere-associated sequence (TAS) (Fu and Barker 1998) and the chromosomal internal repeats, 74-nucleotide long (CIR74) (Atayde, Shi *et al.* 2013). We generated a BLAST database from a FASTA file with 41 SLACS/TATEs extracted from TriTrypDB-5.0_LbraziliensisMHOMBR75M2904_ AnnotatedTranscripts.fasta comprising the nucleotide sequences of the SLACS and TATEs genetic elements and performed a BLAST search with our libraries against this database. In the *L. braziliensis* library, we found 4471 reads mapping to these elements (Supplementary Table B7). Interestingly, about 50% of these reads were both sense and antisense, suggesting that the sequences were present in exosomes as double-stranded RNAs. The lengths of reads were somewhat heterogeneous, ranging from 20 nt to 70 nt, whereas *L. braziliensis* mature siRNAs (*Lbr*AGO1-bound) are believed to be 20-25 nt in length (Atayde, Shi *et al.* 2013). For comparison, we also performed the same BLAST analysis with the *L. donovani* library, where we only found 353 reads mapping to the siRNA-coding genetic elements (Supplementary Table B7). The fact that we found some reads in the *L. donovani* library mapping to these elements could be due to settings used for the BLAST search that were not stringent enough (cut off 80% identity and 70% query coverage), possibly resulting in false positive alignments. Despite this, it is clear that there were >10 times more reads in the *L. braziliensis* library mapping to siRNA-coding genetic elements, indicating that our results are specific and providing evidence that *L. braziliensis* may export siRNAs or their precursors within exosomes. However, we cannot rule out the possibility that the sequences we found in exosomes may originate from regions of the SLACS/TATEs genes other than the ones giving rise to siRNAs.

### 4.4.6. *Leishmania sp.* exosomes contain an abundance of specific tRNA-derived fragments

Remarkably, we found a large number of reads in both the *L. donovani* and *L. braziliensis* exosome RNA libraries that mapped to tRNA genes. A few recent studies characterizing the RNA content of mammalian exosomes had reported the presence of tRNAs or their fragments in these vesicles. For example, reads mapping to tRNAs were found in sequencing libraries made with RNA from exosomes released from neuronal

cells (13.5%) (Bellingham, Coleman *et al.* 2012) , immune cells (~7%) (Nolte-'t Hoen, Buermans *et al.* 2012) and plasma exosomes (1.24%) (Huang, Yuan *et al.* 2013). Strikingly, in our datasets, 351,919 reads (36.4%) and 135,149 reads (21.1%) from *L. donovani* and *L. braziliensis*, respectively, mapped to tRNA genes when aligned to the *Leishmania major* MHOM/IL/81/Friedlin (LmjF) reference genome (which has the best curation on tRNA annotation amongst *Leishmania* species). These frequencies exceeded by some measure those reported for mammalian exosomes in the studies cited above. Close inspection of the genome alignments revealed that a high percentage of these sequences were covering only parts of the respective tRNA genes (Figure 4.5), consistent with the occurrence of tRNA-derived small RNAs (tsRNAs), which has recently been recognized as a specific process. In light of these findings we decided to characterize the reads mapping to tRNAs in more detail. In case of both libraries, the vast majority (99.8%) of reads were in the sense direction of transcription. Looking at their length profiles, we found the mean read length to be slightly different between the two libraries, 38 nt for *L. donovani* and 46 nt for *L. braziliensis*, however, the median read length was similar (33 nt and 34 nt, respectively) (Figure 4.5A). For both *Leishmania sp.* libraries, tsRNAs derived from tRNA-Asp, tRNA-Gln, tRNA-Glu and tRNA-Leu were most abundantly present (Fig.5B and Table 4.4). To make a case that these tsRNAs were specific cleavage products selectively packaged into exosomes, we calculated the Pearson's correlation of the predicted cellular amino acid usage and the relative expression of our tsRNAs as determined by sequencing. The results showed that there was no correlation ($r = 0.163$ for *L. donovani* and $r = 0.114$ for *L. braziliensis*), indicating that the tsRNAs were unlikely to be random degradation products. Strikingly, we observed the same rank order frequency of tRNA isotypes as origins of tsRNAs in both libraries (Figure 4.5B and Table 4.4), indicating that the formation of specific tsRNAs and their appearance as exosomal cargo is an evolutionary conserved phenomenon in *Leishmania sp.*.

### *Table 4.4 Reads mapping to tRNAs*

Distribution of reads from *L. donovani* and *L. braziliensis* libraries over different tRNA isoacceptors. These are sorted by descending abundance in the *L. donovani* library, with the equivalent reads from the *L. braziliensis* library in the same row. nt = nucleotide. N.A. = not applicable

| | ***L. donovani* library** | | | ***L. braziliensis* library** | | |
|---|---|---|---|---|---|---|
| **tRNA** | **% of total tRNA reads** | **length [nt][A]** | **position[B]** | **% of total tRNA reads** | **length [nt][A]** | **position[B]** |
| Asp | 34.91 | 47 | 5’ | 43.51 | 58 | 5’ |
| Gln | 16.00 | 32 | mid-5’ | 13.73 | 36 | mid-5’ |
| Glu | 11.48 | 38 | 3’ | 9.65 | 46 | 5’ |
| Leu | 8.80 | 29 | 5’ | 7.66 | 31 | 5’ |
| Gly | 8.02 | 37 | 3’ | 3.11 | 43 | 5’ |
| Arg | 5.40 | 31 | mid-5’ | 9.62 | 28 | mid-5’ |
| Ala | 4.03 | 34 | 3’ | 2.21 | 34 | 3’ |
| Trp | 2.38 | 37 | mid-3’ | 2.42 | 40 | mid-3’ |
| Val | 2.24 | 36 | 5’ | 1.28 | 41 | 5’ |
| Thr | 1.33 | 31 | 3’ | 1.49 | 35 | 3’ |
| His | 1.08 | 39 | mid-3’ | 1.99 | 42 | mid-3’ |
| Tyr | 1.06 | 31 | 3’ | 0.38 | 28 | 3’ |
| Ser | 1.02 | 34 | 3’ | 0.82 | 39 | mid-3’ |
| Pro | 0.80 | 31 | 3’ | 1.02 | 34 | 3’ |
| Ile | 0.44 | 25 | 3’ | 0.31 | 31 | 3’ |
| Cys | 0.34 | 25 | mid | 0.48 | 26 | mid |
| Phe | 0.32 | 36 | 3’ | 0.07 | 36 | 3’ |
| Lys | 0.24 | 36 | 3’ | 0.09 | 37 | 3’ |
| Met | 0.10 | 29 | 3’ | 0.15 | 33 | 3’ |
| Asn | 0.01 | 33 | mid-5’ | 0.01 | 38 | mid-3’ |

[A] Average read length

[B] Most abundant read position (of all tRNA reads)

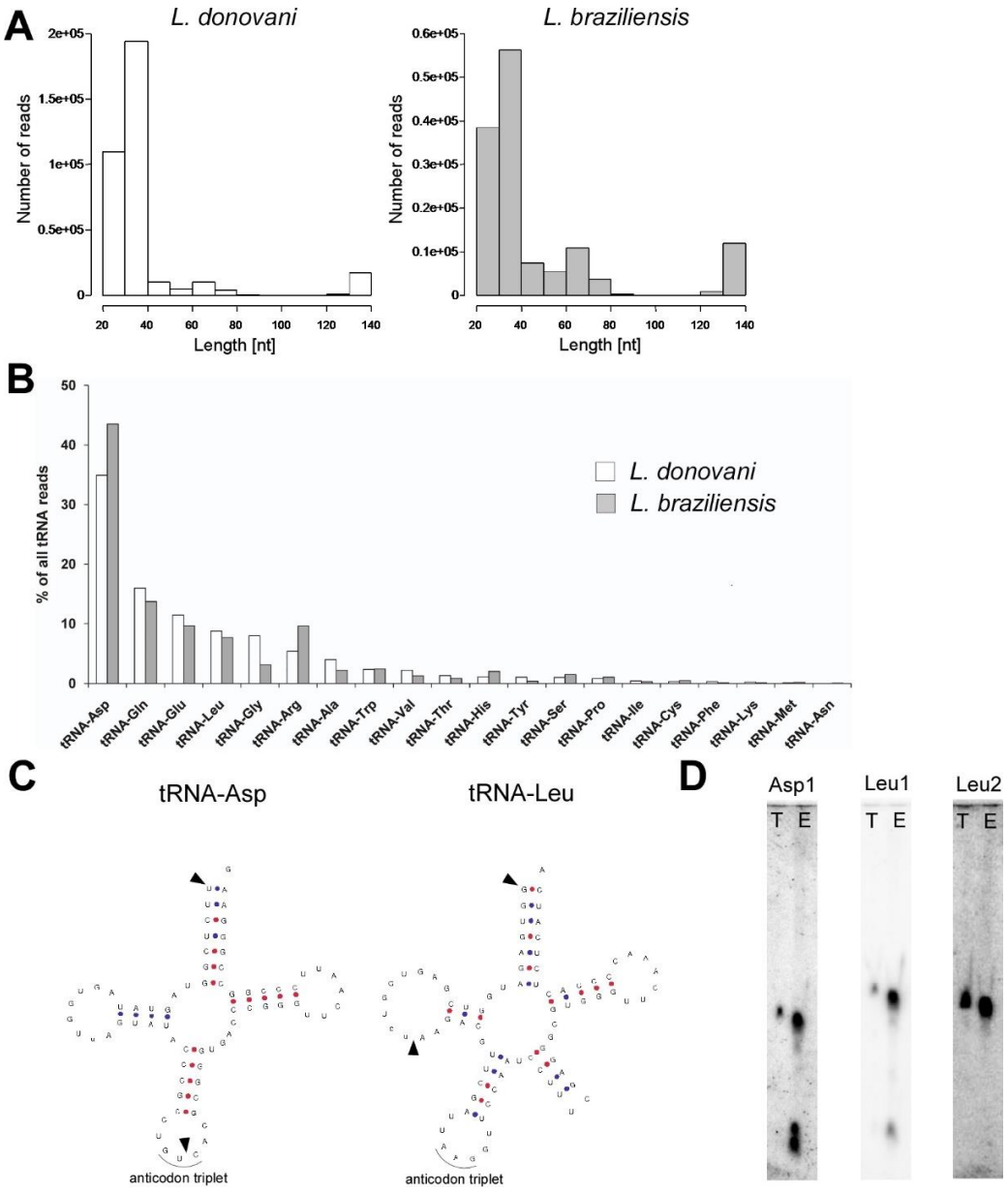



***Figure 4.5. tRNA-derived fragments are cargo of Leishmania* sp. *exosomes***

A. Length distributions of reads mapping to tRNAs in *L. donovani* and *L. braziliensis* exosome RNA sequencing libraries. B. Bar graph showing percentages of reads from *L. donovani* (white bars) and *L. braziliensis* (grey bars) mapping to the respective tRNA isoacceptors. C. tRNA secondary structures for *Leishmania sp.* tRNA-Asp and tRNA-Leu (downloaded from Surgucheva and collueagues (Surgucheva, Sharov *et al.* 2012)). Arrowheads indicate major cleavage products as observed in the sequenced libraries. D. Northern blots with probes designed against tRNA-Asp (Asp1) and tRNA-Leu (Leu1 and Leu2). *L. donovani* total (T) and exosome (E) RNA were probed on the same membrane. Equal amounts of RNA (3 µg) were loaded in each lane. Asp1 and Leu1 probes were designed to specifically detect full length tRNA as well as t-RNA-derived small RNAs seen in sequencing libraries, whereas Leu2 was designed against the mid region (anticodon loop) of tRNA-Leu and hence only detects the full length tRNA.

Next, we investigated whether the tsRNAs were derived from the 3' or 5' end of mature tRNAs, and found that the most abundant tsRNAAsp and tsRNAGln were derived from the 5' end or mid-5' end in both libraries (Table 4.4). Notably, this was not the case for all of the tsRNAs, as some appeared to be derived from the 3' end, and two were not derived from the same end in the two different species (tsRNAGlu and tsRNAGly) (Table 4.4). Furthermore, we saw tsRNAs of different lengths, some of them corresponding to the length of tRNA-halves (~30 nt, e.g. tsRNALeu), and others to the length of tRNA-derived RNA fragments (25 nt, e.g. tsRNAIle) (Table 4.4).

Based on the hypothesis that *Leishmania sp.* tRFs may act as miRNAs or siRNAs in the mammalian or invertebrate host, we performed additional Bowtie2 alignments with all reads mapping to *Leishmania sp.* tRFs against the human and vector genomes looking for complementarity to find potential targets. This search yielded a large number of hits (~20,000), the majority of which fell into the non-annotated regions of the host genomes (Supplementary Tables 6A and 6B). Analogous to our target prediction analyses with the novel transcripts described above, we were unable to perform a more comprehensive (based on more complex RNA-RNA interactions rather than simple complementarity alone) analysis of potential *Leishmania sp.* tsRNA-host target mRNA interactions due to the lack of appropriate tools for large datasets. We also performed a miRNA homology search against the human and mouse miRNA database, and found only one miRNA (miR-135b-5p) that shared 88% identity with one of the tsRNAs present in both libraries (tsRNAArg) (Supplementary Table B8). These results indicate that *Leishmania sp.* tsRNAs are not highly similar to canonical mammalian miRNAs.

In order to validate the presence of the identified tsRNAs in exosomes and to compare their abundance in exosomes with total RNA, we performed Northern blotting with probes specific for $tsRNA^{Asp}$ (Asp1) and $tsRNA^{Leu}$ (Leu1). The probes were designed to be complementary to the 5' end of each tRNA and hence recognize full length tRNA and 5' tRFs. We also included a probe that was designed against the middle region (anticodon loop) of $tRNA^{Leu}$ for comparison (Leu2). When hybridizing blots of RNA isolated from *L. donovani* total cells and versus exosomes with these probes, we detected a common band corresponding to the full length tRNA in total and exosome

RNA in case of both the probes Asp1 and Leu1 (72 nt and 82 nt respectively) (Figure 4.5D). In addition, we detected two smaller bands in the exosome RNA lane of the blot probed with Asp1 which were absent in the total RNA lane. Similarly, we detected a smaller band in the exosome RNA lane of the blot probed with Leu1 which was absent in the total RNA lane. These results demonstrate that 5' tRFs are produced from $tRNA^{Asp}$ and $tRNA^{Leu}$ and that these tRFs are highly enriched in exosomes. In the blot probed with Leu2 we only detected a band corresponding to the full length tRNA in both total and exosome RNA, confirming the absence of fragments that are derived from the anticodon loop of this tRNA. In summary, these findings are the first to show that *Leishmania sp.* producetRFs and that they are specifically enriched in exosomes.

## 4.5. Discussion

### *Leishmania* sp. *exosomes contain specific RNA cargo*

It has been firmly established that exosomes released by various mammalian cell types can serve as shuttle vehicles to deliver RNA molecules to recipient cells, thereby influencing gene expression. However, to date no protozoan pathogen has been shown to release bona fide exosomes containing RNAs with gene regulatory or other sequence-specific properties. *Leishmania sp.* have recently been shown to secrete exosomes that contain a plethora of protein virulence factors capable of affecting the phenotype of host mononuclear phagocytes (Silverman, Clos *et al.* 2010, Silverman, Clos *et al.* 2010). Considering the enormous potential impact of exosome-mediated delivery of regulatory RNAs to either recipient *Leishmania sp.* or mammalian host cells or both, we sought to investigate whether *Leishmania sp.* exosomes carry RNA cargo. Here, we provide unambiguous evidence that *Leishmania sp.* parasites of two distinct species, namely *L. donovani* and *L. braziliensis*, release exosomes containing RNA sequences. These RNA sequences were heterogeneous, but overall short in length (25-250 nt). Thus, despite the abundance of longer sequences in total cell RNA, we were unable to detect them in exosomes. The enrichment of short RNA sequences in *Leishmania sp.* exosomes is concordant with the majority of reports on exosome RNA in other organisms published thus far (e.g. (Nolte-'t Hoen, Buermans *et al.* 2012, Twu, de *et al.* 2013)). Although, there have been a few reports of the presence of longer RNAs

such as full length ribosomal RNAs and mRNAs (Valadi, Ekstrom *et al.* 2007) as well. The RNA cargo of exosomes is largely dependent on the cell of origin and appears to be affected by environmental conditions such as infection or nutritional stress (Pope and Lasser 2013), which likely explains the observed differences.

One important property of exosomes is their capacity to act as both short and long distance messengers. RNA-containing exosomes have been detected in a variety of human body fluids such as plasma, saliva and semen (Rabinowits, Gercel-Taylor *et al.* 2009, Michael, Bajracharya *et al.* 2010, Vojtech, Woo *et al.* 2014) , which supports a role in long distance communication. As RNases are ubiquitously present in all organisms, RNAs travelling long distances need to be protected from degradation. In our *in vitro* experiments, we were able to show that *Leishmania sp.* exosomal RNA cargo is protected from degradation by exogenous RNase. When we incubated PMA-differentiated THP-1 cells *in vitro* with exosomes purified from axenic amastigotes of *L. donovani*, we saw that the exosome RNA cargo was readily taken up by host cells. This finding suggests that it should be possible for *Leishmania sp.*-derived RNAs to gain access to host cells through exosomes *in vivo*.

Numerous studies on exosome RNA have reported the presence of small regulatory RNAs such as micro RNAs (miRNAs) in these vesicles. It was of interest, therefore, to examine the *Leishmania sp.* exosome RNA content in detail by high-throughput sequencing. It is important to mention here that *Leishmania sp.* are a special case with regard to small regulatory RNA pathways: *L. braziliensis* and other species from the new world *Leishmania sp.* (Viannia) subgenus have been shown to have a functional RNAi pathway and actively produce siRNAs (Lye, Owens *et al.* 2010, Atayde, Shi *et al.* 2013). Conversely, this pathway appears to have been evolutionarily lost in old world *Leishmania sp.* species, such as *L. major* and *L. donovani* (Lye, Owens *et al.* 2010). With this contrast in mind, we elected to sequence exosome RNA from *L. braziliensis* and *L. donovani* in parallel in order to compare the exosome RNA trascriptome of an RNAi-competent organism with an RNAi-deficient one.

When sequencing exosomal RNAs from these two *Leishmania sp.* species, we found that they both contained a variety of small non-coding RNA species, the majority

of which appeared to be cleavage products derived from longer known non-coding RNAs such as rRNA, tRNA, snoRNA and snRNA. We also saw a small number of reads mapping to protein coding genes. In addition, we discovered a number of novel transcripts that were conserved in both libraries, and *L. braziliensis* exosomes uniquely contained transcripts derived from siRNA-coding putative mobile elements and repeats, such as SLACS and TATEs (Atayde, Shi *et al.* 2013). Other studies looking at mammalian exosome RNA content by deep sequencing have reported a similar composition of the exosomal transcriptome, with a dominant fraction of sequences being derived from rRNA and other non-coding RNA (Nolte-'t Hoen, Buermans *et al.* 2012, Huang, Yuan *et al.* 2013, Schageman, Zeringer *et al.* 2013, Miranda, Bond *et al.* 2014). Conspicuously, sequences derived from protein coding genes seem to be underrepresented in exosomes. Thus, it appears that exosomes from diverse organisms selectively package non-coding RNAs, although the exact function of which still needs to be determined.

Importantly, our study is the first to purify bona fide exosomes from a protozoan parasite and provide a comprehensive analysis of high-throughput sequencing data of exosomal RNA cargo. By virtue of comparing two distinct (old and new world) species of *Leishmania sp.*, we have made the serendipitous discovery that the packaging of specific RNA sequences into exosomes appears to be a conserved phenomenon in *Leishmania sp.* At the present time it remains unclear whether our findings are illustrative of what happens in other eukaryotic pathogens; however, there is some evidence indicating that the release of RNA within vesicles might occur in other parasitic organisms as well. In particular, there have been two articles published by independent groups that demonstrate the release of tRNA-derived small RNAs and other types of RNA within extracellular vesicles shed by the protozoan *T. cruzi* (Bayer-Santos, Lima *et al.* 2014, Garcia-Silva, das Neves *et al.* 2014); however, these vesicles were not characterized or classified as bona fide exosomes. The distinction between exosomes and other extracellular vesicles is important, as their origin, mechanism of biogenesis and thus loading of cargo differs substantially (Colombo, Raposo *et al.* 2014). Three other articles have been published looking at the RNA cargo of exosomes released by parasitic pathogens; one of them a protozoan (*Trichomonas vaginalis*) and the other two helminths (*Heligmosomoides polygyrus* and *Dicrocoelium dendriticum*). However, all of

these studies have significant limitations in their experimental design and RNA analysis when compared with the present study. The study on *T. vaginalis* only shows a size profile of RNA purified from exosomes measured by Bioanalyzer, but no sequencing data on exosomal RNA (Twu, de *et al.* 2013). The article on *D. dendriticum* describes the analysis of exosomes by high-throughput sequencing; however, data analysis was focused on looking at micro RNA and does not report on other types of RNA in the exosomes (Bernal, Trelis *et al.* 2014). Lastly, a very recent report on *H. polygyrus* reports sequencing data of libraries that have been generated with RNA obtained from parasite secretions and a vesicular fraction collected by ultracentrifugation, but not RNA from bona fide exosomes that were specifically purified (Buck, Coakley *et al.* 2014). The limitations of these studies do not allow for a direct comparison with our data and do not definitely answer the question whether the release of specific types of RNA within exosomes is a widespread phenomenon among parasites. While the data available suggest that this may certainly well be the case, further research will be needed to confidently answer this question.

Results from a number of studies (Karlsson, Lundin *et al.* 2001, Skokos, Le Panse *et al.* 2001, Skokos, Botros *et al.* 2003, Haussecker, Huang *et al.* 2010) led to the suggestion that fragments derived from non-coding RNA species such as rRNA, snoRNA, vault RNA (vRNA) and tRNA can act as regulatory RNAs similar to miRNAs in RNAi. This hypothesis was based upon the finding that these fragments were shown to bind to Argonaute (AGO) proteins and formed RNA-induced silencing complexes (RISCs) which regulate expression of target mRNAs. *L. donovani* does not have the canonical proteins that are required for functional RNAi including AGO. However, an AGO/PIWI-like protein homolog, has been found in RNAi-deficient *Leishmania sp.* and other trypanosomes (Falaleeva and Stamm 2013). The function of this AGO homolog is currently unknown; however, one study suggested that it is not involved in the biogenesis or stability of siRNAs (Ullu, Tschudi *et al.* 2004). The presence of an alternative pathway of regulation of gene expression in RNAi-deficient trypanosomatids is likely, since these organisms perform their transcription polycistronically and hence regulation of expression of individual genes can only take place at the post-transcriptional level. A number of studies have indicated that post-transcriptonal regulation of gene expression in trypanosomes may involve cis-acting regulatory motifs

within the 3' UTRs of mRNAs and trans-acting RNA-binding proteins (D'Orso and Frasch 2001, De Gaudenzi, Carmona *et al.* 2013). Other evidence for the presence of an alternative RNAi pathway in RNAi-deficient trypanosomatids comes from a recent series of studies in *T. cruzi*. The authors identified a unique AGO/PIWI protein termed TcPIWI-tryp that is expressed in all life cycle stages of the parasite and localizes to the cytoplasm (Garcia Silva, Tosar *et al.* 2010). Interestingly, they found that TcPIWI-tryp bound to a repertoire of RNAs distinct from siRNAs, namely small RNAs derived from rRNAs and tRNAs (Garcia-Silva, Sanguinetti *et al.* 2014). However, while these findings are intriguing, it remains to be established whether these complexes function in regulation of gene expression.

A large portion of reads in both our libraries mapped to rRNA genes in the reference genomes and they appeared to be shorter fragments. The presence of rRNA-derived sequences in *Leishmania sp.* exosomes is in agreement with other recent reports on exosome RNA cargo (Jenjaroenpun, Kremenska *et al.* 2013, Miranda, Bond *et al.* 2014). Sequences mapping to rRNAs have also been found in other types of extracellular vesicles, for example shed vesicles released by *T. cruzi* (Bayer-Santos, Lima *et al.* 2014, Garcia-Silva, das Neves *et al.* 2014). At this point it is unknown whether rRNA fragments have any specific function. Some limited evidence has been presented to show that rRNA fragments are produced by specific cleavage rather than random degradation in humans and mice (Li, Ender *et al.* 2012). These specific products were characterized by termini-specific processing and asymmetric stabilization. However, in our data, no bias for either 5' or 3' processing was observed (Supplementary Table B4), but mapping of reads was rather scattered along the entire length of the rRNA gene. Moreover, we did not see enrichment of specific rRNA fragments derived from a subset of genes. Further study will be needed to elucidate whether rRNA fragments in *Leishmania sp.* exosomes are specifically enriched or selectively packaged.

Our finding that *Leishmania sp.* exosomes are overall enriched in non-coding RNA fragments which are taken up by host mononuclear cells, raises the interesting possibility that these RNA fragments may interfere with gene expression in the host possibly by binding to host AGO. This type of epigenetic regulation of gene expression

across kingdoms has been proposed, but so far little consistent and conclusive evidence has been presented. One elegant study recently showed that small RNAs from the plant fungal pathogen *Botrytis cinera* could silence Arabidopsis and tomato genes involved in plant immunity by binding host AGO (Weiberg, Wang *et al.* 2013). This was the first time that naturally occurring cross-kingdom RNAi was shown to be a potential, novel virulence mechanism. Regarding human pathogens, some evidence has been presented that *T. cruzi* produces tRNA-derived small RNAs (tsRNAs) that may be delivered to susceptible mammalian cells (Garcia-Silva, das Neves *et al.* 2014). Moreover, it was shown that transfection of host HeLa cells with synthetic *T. cruzi* tsRNAs can modify the expression of specific genes as assessed by microarray (Garcia-Silva, Cabrera-Cabrera *et al.* 2014). It remains to be established, how these tsRNA-induced changes of host gene expression were brought about and if this process involves hijacking of host RNAi pathways. In what follows below, we discuss potential roles of the non-coding, small RNA species found in *Leishmania sp.* exosomes which we believe are most likely to have regulatory functions, namely novel transcripts, siRNAs and tRNA-derived small RNAs.

### *Novel transcripts*

When examining reads that were less abundant in the libraries to make sure we did not miss any important information amongst this group of reads, we discovered 12 distinct genomic loci (Table 4.3 and Figure 4.4) that apparently gave rise to transcripts which have not been previously described. None of these transcripts had homology to any annotated gene in TriTrypDB or GenBank. A BLAST search confirmed that these non-annotated genomic sequences were conserved among most *Leishmania sp.* species. In considerable interest, we found that all of these novel transcripts had a spliced leader site upstream of their 5' end (see Figure 4.4 for two examples) implying that they are processed alongside other transcripts during trans-splicing. The fact that the majority of novel transcripts contained one or more ORFs suggests that they may be translated into peptides or proteins. However, we were not able to find any homologous protein in any other species.

Two of the twelve novel transcripts that were most abundant from the group were further examined by quantifying their expression in exosomes in comparison to total cells

by Northern analysis (Figure 4.4B). We found that both of these transcripts produced shorter processing products that were uniquely present in exosomes. This indicates that cleavage products of these transcripts may be specifically targeted for packaging into exosomes for release from the cell. The lack of a signal for these shorter products in Northerns of *Leishmania sp.* total RNA may also explain why these transcripts have not been reported in any of the previous studies on sequencing the *Leishmania sp.* transcriptome. Another possibility for why they have not previously been found is that they could be differentially expressed in the different life cycle stages (as we focussed only on axenic amastigotes).

One question that remains to be answered is what type of RNA these novel transcripts represent (protein coding, structural, regulatory) or whether they represent novel type(s) of RNA. Further studies will be needed to properly address the functions of these novel transcripts.

### *siRNAs*

We detected a number of sequences in the *L. braziliensis* exosome RNA library that mapped to siRNA coding loci such as SLACS and TATEs in the *L. braziliensis* genome. Even though the functional significance of endogenous siRNAs in *L. braziliensis* is still unclear, they are thought to be a genome defence mechanism to control the spread of mobile elements, repeats and viruses (Atayde, Shi *et al.* 2013). Although these sequences were detectable in our library, they were not very abundant when compared to fragments of rRNA or tRNA. The fact that half of these sequences were each sense and antisense supported their tentative identity as siRNAs, given that one cardinal feature of siRNAs is that they are double stranded. Moreover, these were the only type of sequences in our libraries where this phenomenon was observed, as the vast majority of the other sequences (rRNA and tRNA fragments) were present only in the sense direction. The lengths of these putative siRNA sequences did not correspond exactly to what has been reported for *L. braziliensis* mature siRNAs (LbrAGO1-bound) (Atayde, Shi *et al.* 2013). This might be due to different library construction strategies (size selection), it may be that what we detected were siRNA precursors, or that the sequences we detected were derived from distinct regions within the SLACS and TATEs loci. The finding that *L. braziliensis* packages these putative siRNA sequences as cargo

of exosomes is of significant interest. It implies that these RNAs may function not only in the cell where they originated, but that they may also act in intercellular communication when taken up by other *Leishmania sp.* or by host cells or both. To our knowledge, no other parasite has been shown to release pathogen-derived siRNAs within vesicles directly. Further studies will be needed to confirm the identity of the sequences as siRNAs and delineate their function in parasite biology and in host-pathogen interaction.

### *tRNA-derived small RNAs*

A striking finding of the present study was the abundance of tRNA fragments principally originating from a small subset of tRNA isoacceptors (Figure 4.4B) that were highly conserved in the *L. donovani* and *L. braziliensis* exosome transcriptomes. Only recently tRNA fragmentation has been recognized as a specific process. tRNA fragments have been found in all domains of life and can be divided into several categories, depending on the cleavage site: cleavage within the anticodon loop gives rise to 5' and 3' tRNA halves (30-35 nt), and cleavage within the D-arm (5') or T-arm (3') gives rise to smaller tRNA-derived RNA-fragments (tRFs) (13-20 nt). Each of these fragments appears to be generated through distinct pathways. tRNA halves are known to be produced in response to stress (Raina and Ibba 2014), whereas the smaller tRFs, on the other hand, can be generated at any point during tRNA processing, by Dicer-dependent or –independent mechanisms (Kumar, Anaya *et al.* 2014, Raina and Ibba 2014). Together, tRNA halves and tRFs are referred to as tRNA-derived small RNAs (tsRNAs). tsRNAs have recently been described in higher as well as lower eukaryotes. Their physiological function is not well understood, but they have commanded increasing interest due to their suspected regulatory nature. Notably, it appears that tRNA halves and tRFs have distinct biological functions. In human cells, tRNA halves were found to inhibit protein translation by targeting the translation initiation machinery and displacing elongation initiation factors (Ivanov, Emara *et al.* 2011). tRFs, on the other hand, were shown to be involved in regulation of translation by directly binding to the small ribosomal subunit and interfering with peptidyl-transferase activity in archaea (Gebetsberger, Zywicki *et al.* 2012). Furthermore, a similar mechanism was observed in human cells, where a tRF was shown to inhibit translation by affecting peptide bond formation (Sobala and Hutvagner 2013). In addition to the effects on translation, tRFs have also been shown to function during gene silencing. Haussecker *et al* showed that

tRFs can associate with Argonaute proteins, however, they associated more efficiently with the non-silencing AGO3 and AGO4 (Haussecker, Huang *et al.* 2010). Furthermore, they saw that tRFs can affect the silencing activities of miRNAs and siRNAs, indicating a potential broad based role in regulating RNA silencing. In another study it was found that tRFs can function like miRNAs in RNAi and inhibit the expression of RPA1 (a protein involved in DNA repair) by binding to the 3' UTR of its mRNA (Maute, Schneider *et al.* 2013).

A small number of studies have looked at the presence of tsRNAs in protozoan parasites. In *Plasmodium berghei* and *Toxoplasma gondii*, tRNA-halves were detected, and a connection between tRNA-half production and growth rate was observed (Galizi, Spano *et al.* 2013). However, the precise function of these parasite-derived tRNA-halves remains unknown. Interestingly, tsRNAs were recently discovered in *Leishmania sp.*'s close relative *T. cruzi*. Despite the fact that the lengths and origins of tsRNAs differed slightly from study to study, their production was convincingly demonstrated in both trypomastigotes and epimastigotes in a number of reports (Garcia-Silva, Frugier *et al.* 2010, Franzen, Arner *et al.* 2011, Reifur, Garcia-Silva *et al.* 2012). Importantly, it was found that *T. cruzi* tsRNAs were bound to TcPIWI-tryp, a distinct Argonaute protein that has been described in this RNAi-deficient organism (Garcia Silva, Tosar *et al.* 2010, Garcia-Silva, Sanguinetti *et al.* 2014). The majority of these TcPIWI-tryp bound tsRNAs were derived from the 5` halves of $tRNA^{Glu}$. Importantly, TcPIWI-tryp-tsRNA complexes were also found in vesicles from *T. cruzi*. The authors proposed that these vesicles may have a role in life cycle transition from epimastigote to trypomastigote as well as contribute to infection susceptibility of mammalian cells (Garcia-Silva, das Neves *et al.* 2014). These findings provide evidence that tsRNAs in *T. cruzi* may participate in non-canonical regulatory pathways and raise the question as to whether a similar phenomenon may be operative in *Leishmania sp.*.

In the present study, we provide evidence that *Leishmania sp.* also produces specific tsRNAs, and that these potential regulatory RNAs are released by the intracellular amastigote stage within bona fide exosomes, competent for delivery to mammalian cells. The major fraction of tsRNAs found in both *L. donovani* and *L. braziliensis* exosomes were 5' tRNA halves, however, we also found shorter tsRNAs

derived from the D-arm or T-arm of the tRNA, corresponding to 5' tRFs and 3' tRFs. The production and presence of tRNA halves in exosomes from *Leishmania sp.* amastigotes might be a result of the elevated temperature and acidic pH the parasites were exposed to. We found that the vast majority of tsRNAs in *Leishmania sp.* exosomes were derived from tRNA-Asp, tRNA-Gln, tRNA-Leu, tRNA-Glu and tRNA-Gly (Table 4.4 and Figure 4.5) and this was highly conserved between *L. donovani* and *L. braziliensis*. Although we did not carry out a comprehensive and quantitative analysis of the frequencies of all tsRNAs in *Leishmania sp.* whole cells, strikingly, we found by Northern blotting that tsRNAs from tRNA-Asp and tRNA-Leu were highly enriched in exosomes, with no detectable amounts in *Leishmania sp.* total RNA. This indicates that these tsRNAs are preferentially and quantitatively packaged into exosomes to be released from the cell rather than being retained in the whole cell (minus exosomal) RNA pool.

The mechanism of biogenesis and function of tsRNAs in *Leishmania sp.* remains unknown. Therefore, we cannot be certain that the same classification of tRNA halves and tRFs as recently proposed by several groups (Gebetsberger and Polacek 2013, Kumar, Anaya *et al.* 2014, Raina and Ibba 2014) applies to our data. However, as many of the characteristics (length, cleavage site, isoacceptor origin) correspond to what has been reported in other organisms, we conclude that the phenomenon of specific tsRNA generation is evolutionarily conserved in *Leishmania sp.* as well. Based upon our initial functional predictions it appears clear that *Leishmania sp.* tsRNAs are not highly similar to canonical mammalian miRNAs or siRNAs. Further detailed investigations will be needed to delineate the functions of tsRNAs in *Leishmania sp.* biology, what roles they play in parasite-parasite, parasite-vector or parasite-host interactions, whether this involves their association with the host RNAi machinery and how they are targeted for exosomal packaging.

## 4.6. Conclusions

In summary, this report provides evidence that *Leishmania sp.* exosomes are enriched in short sequences derived from non-coding RNAs such as rRNAs and tRNAs. Moreover, exosomes contain a number of novel transcripts, albeit in relatively low abundance. The RNAi-proficient *L. braziliensis* appears to package putative siRNAs or

their precursors into exosomes, whereas RNAi deficient *L. donovani* does not. Based on Northern analyses, our data indicate that specific RNA sequences are selectively, and in some cases quantitatively packaged into exosomes. This conclusion is supported further by the highly biased distribution of sequences detected in exosomes over only a subset of genes in the *Leishmania sp.* genome combined with a striking paucity of transcripts derived from protein coding genes which are otherwise abundant in total cellular RNA.

Our findings provide at least three lines of evidence arguing for the presence of an evolutionarily conserved mechanism for specific packaging of small non-coding RNAs into exosomes in *Leishmania sp..* First, the high degree overlap between the top 20 most abundant sequences found in *L. donovani* and *L. braziliensis* exosomes, second the vast majority of identified novel transcripts were present in exosomes from both species, and finally the most abundant tsRNAs found in exosomes were derived from a highly biased subset of the same tRNA isoacceptors in both species. Taken together, the data strongly argues that *Leishmania sp.* exosomal RNA sequences are specifically produced and packaged into exosomes for release, likely with the purpose of modifying host cell phenotype to support chronic infection. The investigation of the precise functions of these small, non-coding, *Leishmania sp.* RNAs should contribute significantly to our understanding of parasite biology and mechanisms of pathogenesis.

# Chapter 5.

# Conclusion

My interest in how RNAs can regulate cellular pathways (with a particular focus on pathogens) has been the driving force of all the work I've done in Dr. Unrau's Lab. The discovery of short RNA molecules that regulate the expression of genes has contributed to my work and is probably one of the most outstanding discoveries in recent years (Fire 1998). RNAi is not only an ancient genome defense system used by eukaryotic organisms, but also a very powerful experimental tool. RNA can control gene expression by various mechanisms, including: I) Double-stranded RNA (siRNA) tags mRNA for cleavage. II) miRNA inhibiting translation, and III) the induction of chromatin remodeling (Novina and Sharp 2004). One of the first projects I worked on was a "Small RNA survey across the Kinetoplastida". The goal of this project was to create small RNA libraries from different species of kinetoplastids and analyze them with high-throughput sequencing. I expected that comparing the data from the different libraries would lead me to find small RNAs or common RNA motifs that could be related to pathogenesis. Although the project was discontinued due to the logistical complexities of collecting pathogens from different developing countries, the research revealed that RNA is a multi-faceted molecule with a wide variety of roles and that many of these roles are not fully understood, or not yet even discovered.

Studying the release of 6S RNA and mutant versions of it verified that the secondary structure of 6S RNA plays a key role in determining its function, as had been suggested by my group's previous work as well as the previous work of other researchers (Barrick, Sudarsan *et al.* 2005, Wassarman and Saecker 2006, Shephard, Dobson *et al.* 2010, Beckmann, Hoch *et al.* 2012, Cavanagh, Sperger *et al.* 2012, Panchapakesan and Unrau 2012, Steuten and Wagner 2012). Nonetheless, the secondary structure is not the only not essential feature that determines the function of

6S RNA. Sequence was also found to play an important role, especially when changes were introduced around the TSS or the LOB. It is probable that a combination of sequence and structure determines the nature of the pRNAs that will be produced and largely governs their release rates. Studying release mutants led me to observe a wide range of pRNAs lengths and release rates. The R9-33 (Class I) was the most interesting mutant 6S RNA construct and was found to be tightly bound to the $E\sigma^{70}$ and to only produce a short pRNA product. Given that R9-33 has been shown to bind with the same efficiency as T1, we think it is a good candidate for crystallization experiments that attempt to provide a better understanding of how 6S RNA interacts with the $E\sigma^{70}$. In addition, since we know that it is possible to delete the gene that codes for 6S RNA (Trotochaud and Wassarman 2004), replacing it with R9-33 could shed more light on the consequences of its expression in cellular growth, cell survival and recovery from the stationary phase. During our *in vivo* experiments with R9-33, I observed that its expression induces a stationary phase type of behavior. This could be a useful tool for the study of pathogens since it is clearly evident that several of the pathogenesis factors are under the control of $\sigma^S$ factor, which is also active during the stationary phase. If we attain a better understanding of the consequences of expressing R9-33 in cells, we can then use the suspended stationary phase R9-33 triggers as a way of mimicking stress in bacteria. Studying how bacteria adapt quickly to stress and changing environments might help us to understand how they behave while invading a host, which is a stressful situation.

In attempting to understand how *E. coli* transformed with pEcoli-R9-33 can resume exponential growth while plasmid-derived R9-33 expression decays, I first looked for additional proteins that interact with 6S RNA. The idea is that additional proteins that interact with the 6S RNA can provide clues as to what additional roles 6S RNA might have in the cell and how 6S RNA is degraded once it is no longer needed. Finding proteins that interact with 6S RNA by using a cell extract as the starting material for our experiments proved to be more difficult than initially expected. Although we could not conclusively identify proteins that interact specifically, the knowledge and experience we gained can be the starting point for further experiments and eventually aid the achievement of the previously mentioned goals. One of the questions that still remain unanswered is how the 6S RNA is degraded once it is no longer needed. Although this

question is tangential, our research suggests that 6S RNA might be degraded by an RNase that works independently from the degradosome. Previous work by other labs supports this idea due to the fact that Hfq does not interact with 6S RNA. Hfq is the chaperone molecule that will bring RNAs close to the RNase E in order to be degraded by the degradosome. Another question that still remains unanswered is why there are high levels of 6S RNA molecules throughout cellular growth, when they only assume an active role during the stationary phase binding to $E\sigma^{70}$. It is possible that 6S RNA could behave like $\sigma^{S}$, which maintains high levels of mRNA in the cell even during the exponential phase (Hengge-Aronis 2002, Hengge-Aronis 2002). It is important to continue researching the 6S RNA behavior over the long-term not only because it allows us to have a better understanding of the transcription process, but also because regulatory RNAs like 6S RNA are known to be involved in bacterial antibiotic resistance. Although 6S sRNA has not been shown to directly affect antibiotic resistance by regulating transcription, it is well-known that cells lacking 6S RNA are defective in terms of persistence (Trotochaud and Wassarman 2004). Persistence is a process that makes bacteria highly tolerant to different antibiotics and current data suggest there is a functional link between 6S RNA expression and antibiotic sensitivity that still remains to be explored (Lalaouna, Eyraud *et al.* 2014). Advancing a detailed understanding of how R9-33 irreversibly binds $E\sigma^{70}$ (*in vitro*) and stalls bacterial growth (at least transiently) could potentially make a valuable contribution to uncovering a novel antibiotic target.

The finding that ncRNA are packaged in exosomes from *Leishmania sp.* is the first of its kind for kinetoplastids. This finding emerged from a technical point of view and within the context a very challenging project and only begins to advance a better understanding of ncRNAs' relevance and role during pathogenesis. One of the challenges encountered in the course of the project was that the kit used for enriching exosomes does not explain what the principle of purification is. Another challenge was that the additives used to supplement the growth media might produce nucleic acid contamination. These challenges were difficult to overcome because there are no other alternatives for the *in vitro* growth of trypanosomatids. Further, we had to work with the limitation of there being a lack of in-depth knowledge about how RNA processing and maturation occur in trypanosomatids. Although we tried to remove all the potential caps (with CIAP and TAP treatments) present in RNAs, it is important to acknowledge that we

may have unknowingly introduced a bias the construction of the library by failing to remove additional groups that protect the RNA ends. An essential next step is to generate in parallel exosomes and whole cell-derived libraries in order to completely confirm the specific enrichment that was detected by comparing two exosome libraries from different *Leishmania* species. Once the ncRNAs present in the exosomes are better characterized, their role in the invasion of the host cells as well as the changes they induce in host cells needs to be studied. It is my hope that a better comprehension of the role of exosomal RNA will enhance our understanding of how these pathogens manage to establish chronic infections.

# References


Adams, M. J.*, et al.* (2013). "Recently agreed changes to the International Code of Virus Classification and Nomenclature." Archives of Virology **158**(12): 2633-2639.

Alberts, B. (2002). Molecular biology of the cell. New York, Garland Science.

Altschul, S. F.*, et al.* (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. **25**(17): 3389-3402.

Alwine, J. C.*, et al.* (1977). "Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes." Proceedings of the National Academy of Sciences of the United States of America **74**(12): 5350-5354.

Amaral, P. P.*, et al.* (2008). "The eukaryotic genome as an RNA machine." Science **319**(5871): 1787-1789.

Ames, T. D. and R. R. Breaker (2010). Bacterial Riboswitch Discovery and Analysis. The Chemical Biology of Nucleic Acids, John Wiley & Sons, Ltd**:** 433-454.

Aravin, A. A. (2001). "Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline." Curr. Biol. **11**: 1017-1027.

Aravin, A. A. (2008). "A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice." Mol. Cell **31**: 785-799.

Archambault, J. and J. D. Friesen (1993). "Genetics of eukaryotic RNA polymerases I, II, and III." Microbiol Rev **57**(3): 703-724.

Argaman, L.*, et al.* (2001). "Novel small RNA-encoding genes in the intergenic regions of Escherichia coli." Curr Biol **11**(12): 941-950.

Arluison, V.*, et al.* (2007). "Sm-like protein Hfq: Location of the ATP-binding site and the effect of ATP on Hfqâ€“RNA complexes." Protein Science : A Publication of the Protein Society **16**(9): 1830-1841.

Arnez, J. G. and D. Moras (1997). "Structural and functional considerations of the aminoacylation reaction." Trends Biochem Sci **22**(6): 211-216.

Atayde, V. D.*, et al.* (2013). "The structure and repertoire of small interfering RNAs in Leishmania (Viannia) braziliensis reveal diversification in the trypanosomatid RNAi pathway." Mol.Microbiol. **87**(3): 580-593.

Babu, M.*, et al.* (2011). "A dual function of the CRISPR–Cas system in bacterial antivirus immunity and DNA repair." Molecular Microbiology **79**(2): 484-502.

Baek, D. (2008). "The impact of microRNAs on protein output." Nature **455**: 64-71.

Baj-Krzyworzeka, M.*, et al.* (2006). "Tumour-derived microvesicles carry several surface determinants and mRNA of tumour cells and transfer some of these determinants to monocytes." Cancer Immunol Immunother **55**(7): 808-818.

Balakrishnan, L. and R. A. Bambara (2013). "Okazaki Fragment Metabolism." **5**(2).

Ban, N.*, et al.* (2000). "The complete atomic structure of the large ribosomal subunit at 2.4 A resolution." Science **289**(5481): 905-920.

Bardill, J. P.*, et al.* (2011). "The Vibrio cholerae quorum sensing response is mediated by Hfq-dependent sRNA/mRNA base pairing interactions." Molecular Microbiology **80**(5): 1381-1394.

Bardwell, J. C.*, et al.* (1989). "Autoregulation of RNase III operon by mRNA processing." The EMBO Journal **8**(11): 3401-3407.

Barrangou, R.*, et al.* (2007). "CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes." Science **315**(5819): 1709-1712.

Barrett, M. P.*, et al.* (2003). "The trypanosomiases." Lancet **362**(9394): 1469-1480.

Barrick, J. E. and R. R. Breaker (2007). "The distributions, mechanisms, and structures of metabolite-binding riboswitches." Genome Biol **8**(11): R239.

Barrick, J. E.*, et al.* (2005). "6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter." RNA **11**(5): 774-784.

Bayer-Santos, E.*, et al.* (2013). "Proteomic analysis of Trypanosoma cruzi secretome: characterization of two populations of extracellular vesicles and soluble proteins." J.Proteome.Res. **12**(2): 883-897.

Bayer-Santos, E.*, et al.* (2014). "Characterization of the small RNA content of Trypanosoma cruzi extracellular vesicles." Mol.Biochem.Parasitol. **193**(2): 71-74.

Beckmann, B. M.*, et al.* (2012). "A pRNA-induced structural rearrangement triggers 6S-1 RNA release from RNA polymerase in Bacillus subtilis." EMBO J **31**(7): 1727-1738.

Belasco, J. and G. Brawerman (1993). Experimental approaches to the study of mRNA decay. In: Belasco JG, Brawerman G (eds) Control of messenger RNA stability. New York, Academic Press

Bellingham, S. A*., et al.* (2012). "Small RNA deep sequencing reveals a distinct miRNA signature released in exosomes from prion-infected neuronal cells." Nucleic Acids Res. **40**(21): 10937-10949.

Bernal, D*., et al.* (2014). "Surface analysis of Dicrocoelium dendriticum. The molecular characterization of exosomes reveals the presence of miRNAs." J.Proteomics. **105**: 232-241.

Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." Nature **321**(6067): 209-213.

Birnboim, H. C. a. D., J (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucleic Acids Res. **7**(6): 1513-1523.

Birney, E. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**: 799-816.

Boeke, J. D. (2003). "The Unusual Phylogenetic Distribution of Retrotransposons: A Hypothesis." Genome Research **13**(9): 1975-1983.

Borukhov, S*., et al.* (1993). "Two modes of transcription initiation in vitro at the rrnB P1 promoter of Escherichia coli." J Biol Chem **268**(31): 23477-23482.

Breaker, R. R. (2012). "Riboswitches and the RNA world." Cold Spring Harb Perspect Biol **4**(2).

Breaker, Ronald R. and Gerald F. Joyce (2014). "The Expanding View of RNA and DNA Function." Chemistry & Biology **21**(9): 1059-1065.

Brennecke, J. (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila." Cell **128**: 1089-1103.

Brennecke, J*., et al.* (2005). "Principles of microRNA-target recognition." PLoS Biol. **3**: e85.

Brimacombe, R*., et al.* (1978). "Ribosome Structure." Annual Review of Biochemistry **47**(1): 217-249.

Brouns, S. J. J*., et al.* (2008). "Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes." Science **321**(5891): 960-964.

Buck, A. H., *et al.* (2014). "Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity." Nat.Commun. **5**: 5488.

Bullido, R. (2000). "RNA virus replication." Journal of Clinical Virology **18**(1â€“3): 105-112.

Bunz, U. H. F. (2005). "How Are Alkynes Scrambled?" Science **308**(5719): 216-217.

Burger, A. l., *et al.* (2011). "Current perspectives of the Escherichia coli RNA degradosome." Biotechnology Letters **33**(12): 2337-2350.

Cabrera-Ostertag, I. J., *et al.* (2013). "Initiating nucleotide identity determines efficiency of RNA synthesis from 6S RNA templates in Bacillus subtilis but not Escherichia coli." Nucleic Acids Res **41**(15): 7501-7511.

Campbell, E. A., *et al.* (2001). "Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase." Cell **104**(6): 901-912.

Carpousis, A. J. (2007). "The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E." Annu Rev Microbiol **61**: 71-87.

Carver, T., *et al.* (2012). "Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data." Bioinformatics. **28**(4): 464-469.

Cavanagh, A. T., *et al.* (2008). "Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of sigma70." Mol Microbiol **67**(6): 1242-1256.

Cavanagh, A. T., *et al.* (2012). "Regulation of 6S RNA by pRNA synthesis is required for efficient recovery from stationary phase in E. coli and B. subtilis." Nucleic Acids Res **40**(5): 2234-2246.

Cavanagh, A. T. and K. M. Wassarman (2013). "6S-1 RNA function leads to a delay in sporulation in Bacillus subtilis." J Bacteriol **195**(9): 2079-2086.

Cech, T. R. and B. L. Bass (1986). "Biological catalysis by RNA." Annu Rev Biochem **55**: 599-629.

Cech, T. R., *et al.* (1981). "In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence." Cell **27**(3 Pt 2): 487-496.

Cerutti, H. and J. A. Casas-Mollano (2006). "On the origin and functions of RNA-mediated silencing: from protists to man." Curr Genet **50**(2): 81-99.

Chae, H.*, et al.* (2011). "Rho-dependent Termination of ssrS (6S RNA) Transcription in Escherichia coli: IMPLICATION FOR 3' PROCESSING OF 6S RNA AND EXPRESSION OF DOWNSTREAM ygfA (PUTATIVE 5-FORMYL-TETRAHYDROFOLATE CYCLO-LIGASE)." Journal of Biological Chemistry **286**(1): 114-122.

Chan, S. W.-L.*, et al.* (2004). "RNA Silencing Genes Control de Novo DNA Methylation." Science **303**(5662): 1336.

Chandran, V. and B. F. Luisi (2006). "Recognition of Enolase in the Escherichia coli RNA Degradosome." Journal of Molecular Biology **358**(1): 8-15.

Cheetham, G. M. and T. A. Steitz (1999). "Structure of a transcribing T7 RNA polymerase initiation complex." Science **286**(5448): 2305-2309.

Chen, J.*, et al.* (2010). "Promoter melting triggered by bacterial RNA polymerase occurs in three steps." Proc Natl Acad Sci U S A **107**(28): 12523-12528.

Clancy, S. (2014). "Chemical Structure of RNA." Nature Education **7**(1): 60.

Claude, A. (1943). "The Constitution of Protoplasm." Science **97**(2525): 451-456.

Colombo, M.*, et al.* (2014). "Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles." Annu.Rev.Cell Dev.Biol. **30**: 255-289.

Cooper, G. M. and R. E. Hausman (2007). The cell : a molecular approach. Washington, D.C.

Sunderland, Mass., ASM Press ;

Sinauer Associates.

Corn, J. E. and J. M. Berger (2006). "Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions." Nucleic Acids Research **34**(15): 4082-4088.

Cramer, P.*, et al.* (2001). "Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution." Science **292**(5523): 1863-1876.

Crick, F. (1970). "Central dogma of molecular biology." Nature **227**(5258): 561-563.

Czech, B. (2008). "An endogenous small interfering RNA pathway in Drosophila." Nature **453**: 798-802.

D'Orso, I. and A. C. Frasch (2001). "TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes." J.Biol.Chem. **276**(37): 34801-34809.

Darfeuille, F.*, et al.* (2007). "An Antisense RNA Inhibits Translation by Competing with Standby Ribosomes." Molecular Cell **26**(3): 381-392.

De Gaudenzi, J. G.*, et al.* (2013). "Genome-wide analysis of 3'-untranslated regions supports the existence of post-transcriptional regulons controlling gene expression in trypanosomes." PeerJ. **1**: e118.

Deltcheva, E.*, et al.* (2011). "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Nature **471**(7340): 602-607.

Deutscher, M. P. (2006). "Degradation of RNA in bacteria: comparison of mRNA and stable RNA." Nucleic Acids Res **34**(2): 659-666.

Didiano, D. and O. Hobert (2006). "Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions." Nat.Struct.Mol.Biol. **13**(9): 849-851.

Diener, T. O. (1971). "Potato spindle tuber virus: a plant virus with properties of a free nucleic acid. 3. Subcellular location of PSTV-RNA and the question of whether virions exist in extracts or in situ." Virology **43**(1): 75-89.

Diener, T. O. (2001). "The viroid: biological oddity or evolutionary fossil?" Adv Virus Res **57**: 137-184.

Dinhopl, N.*, et al.* (2011). "In situ hybridisation for the detection of Leishmania species in paraffin wax-embedded canine tissues using a digoxigenin-labelled oligonucleotide probe." Vet.Rec. **169**(20): 525.

Djikeng, A.*, et al.* (2001). "RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs." Rna **7**(11): 1522-1530.

Dolgosheina, E. V.*, et al.* (2014). "RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking." ACS Chemical Biology **9**(10): 2412-2420.

Dolgosheina, E. V.*, et al.* (2008). "Conifers have a unique small RNA silencing signature." RNA **14**(8): 1508-1515.

Dong, T. and H. E. Schellhorn (2009). "Control of RpoS in global gene expression of Escherichia coli in minimal media." Mol Genet Genomics **281**(1): 19-33.

Dove, S. L.*, et al.* (2003). "Region 4 of sigma as a target for transcription regulation." Mol Microbiol **48**(4): 863-874.

Durrenberger, M.*, et al.* (1988). "Intracellular location of the histonelike protein HU in Escherichia coli." Journal of Bacteriology **170**(10): 4757-4768.

Duval, M.*, et al.* (2013). "Escherichia coli ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation." PLoS Biol **11**(12): e1001731.

Ebhardt, H. A.*, et al.* (2012). "Isolation and biochemical analysis of plant small RNAs." Methods Mol Biol **894**: 223-239.

Ebhardt, H. A.*, et al.* (2005). "Extensive 3' modification of plant small RNAs is modulated by helper component-proteinase expression." Proc Natl Acad Sci U S A **102**(38): 13398-13403.

Ebhardt, H. A. and P. J. Unrau (2009). "Characterizing multiple exogenous and endogenous small RNA populations in parallel with subfemtomolar sensitivity using a streptavidin gel-shift assay." Rna **15**(4): 724-731.

Ederth, J.*, et al.* (2002). "The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing." J Biol Chem **277**(40): 37456-37463.

Elbashir, S. M.*, et al.* (2001). "RNA interference is mediated by 21- and 22-nucleotide RNAs." Genes & Development **15**(2): 188-200.

Emoődy, L.*, et al.* (2002). A Role for the Ï*f*s Subunit of RNA Polymerase in the Regulation of Bacterial Virulence. Genes and Proteins Underlying Microbial Urinary Tract Virulence, Springer US. **485:** 85-93.

Falaleeva, M. and S. Stamm (2013). "Processing of snoRNAs as a new source of regulatory non-coding RNAs: snoRNA fragments form a new class of functional RNAs." Bioessays **35**(1): 46-54.

Farewell, A.*, et al.* (1998). uspB, a New Ï‚S-Regulated Gene in Escherichia coli Which Is Required for Stationary-Phase Resistance to Ethanol. **180:** 6140-6147.

Felden, B.*, et al.* (1997). Probing the structure of the Escherichia coli 10Sa RNA (tmRNA). **3:** 89-103.

Fire, A. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**: 806-811.

Fischer, G. and F. X. Schmid (1990). "The mechanism of protein folding. Implications of in vitro refolding models for de novo protein folding and translocation in the cell." Biochemistry **29**(9): 2205-2212.

Flores, R.*, et al.* (2005). "Viroids and Viroid-Host Interactions." Annual Review of Phytopathology **43**(1): 117-139.

Frank, D. N. and N. R. Pace (1998). "Ribonuclease P: unity and diversity in a tRNA processing ribozyme." Annu Rev Biochem **67**: 153-180.

Franzen, O.*, et al.* (2011). "The short non-coding transcriptome of the protozoan parasite Trypanosoma cruzi." PLoS.Negl.Trop.Dis. **5**(8): e1283.

Fu, G. and D. C. Barker (1998). "Characterisation of Leishmania telomeres reveals unusual telomeric repeats and conserved telomere-associated sequence." Nucleic Acids Res. **26**(9): 2161-2167.

Galizi, R.*, et al.* (2013). "Evidence of tRNA cleavage in apicomplexan parasites: Half-tRNAs as new potential regulatory molecules of Toxoplasma gondii and Plasmodium berghei." Mol.Biochem.Parasitol. **188**(2): 99-108.

Garcia-Silva, M. R.*, et al.* (2014). "Gene expression changes induced by Trypanosoma cruzi shed microvesicles in mammalian host cells: relevance of tRNA-derived halves." Biomed.Res.Int. **2014**: 305239.

Garcia-Silva, M. R.*, et al.* (2014). "Extracellular vesicles shed by Trypanosoma cruzi are linked to small RNA pathways, life cycle regulation, and susceptibility to infection of mammalian cells." Parasitol.Res. **113**(1): 285-304.

Garcia-Silva, M. R.*, et al.* (2010). "A population of tRNA-derived small RNAs is actively produced in Trypanosoma cruzi and recruited to specific cytoplasmic granules." Mol.Biochem.Parasitol. **171**(2): 64-73.

Garcia-Silva, M. R.*, et al.* (2014). "A particular set of small non-coding RNAs is bound to the distinctive Argonaute protein of Trypanosoma cruzi: insights from RNA-interference deficient organisms." Gene **538**(2): 379-384.

Garcia Silva, M. R.*, et al.* (2010). "Cloning, characterization and subcellular localization of a Trypanosoma cruzi argonaute protein defining a new subfamily distinctive of trypanosomatids." Gene **466**(1-2): 26-35.

Garneau, J. E.*, et al.* (2010). "The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA." Nature **468**(7320): 67-71.

Gebetsberger, J. and N. Polacek (2013). "Slicing tRNAs to boost functional ncRNA diversity." RNA.Biol. **10**(12): 1798-1806.

Gebetsberger, J.*, et al.* (2012). "tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii." Archaea. **2012**: 260909.

Ghildiyal, M. and P. D. Zamore (2009). "Small silencing RNAs: an expanding universe." Nat Rev Genet **10**(2): 94-108.

Ghosh, A. and M. Bansal (2003). A glossary of DNA structures from A to Z. Acta Crystallographica Section D. **59:** 620-626.

Ghosh, J.*, et al.* (2013). "Leishmania donovani targets Dicer1 to downregulate miR-122, lower serum cholesterol, and facilitate murine liver infection." Cell Host.Microbe **13**(3): 277-288.

Giri, P. K.*, et al.* (2010). "Proteomic analysis identifies highly antigenic proteins in exosomes from M. tuberculosis-infected and culture filtrate protein-treated macrophages." Proteomics **10**: 1-13.

Giuliodori, A. M.*, et al.* (2007). "Review on Bacterial Stress Topics." Annals of the New York Academy of Sciences **1113**(1): 95-104.

Goldman, S. R.*, et al.* (2009). "Direct Detection of Abortive RNA Transcripts in Vivo." Science (New York, N.Y.) **324**(5929): 927-928.

Gottesman, S. (2004). "The small RNA regulators of Escherichia coli: roles and mechanisms*." Annu Rev Microbiol **58**: 303-328.

Green and, R. and H. F. Noller (1997). "RIBOSOMES AND TRANSLATION." Annual Review of Biochemistry **66**(1): 679-716.

Gruber, T. M. and C. A. Gross (2003). "MULTIPLE SIGMA SUBUNITS AND THE PARTITIONING OF BACTERIAL TRANSCRIPTION SPACE." Annual Review of Microbiology **57**(1): 441-466.

Grundy, F. J.*, et al.* (2002). "tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome." Proc Natl Acad Sci U S A **99**(17): 11121-11126.

Gualerzi, C. O.*, et al.* (2003). "Transcriptional and Post-transcriptional Control of Cold-shock Genes." Journal of Molecular Biology **331**(3): 527-539.

Guerrier-Takada, C.*, et al.* (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." Cell **35**(3 Pt 2): 849-857.

Gur, E. and R. T. Sauer (2008). Evolution of the ssrA degradation tag in Mycoplasma: Specificity switch to a different protease. **105:** 16113-16118.

Gusarov, I. and E. Nudler (1999). "The Mechanism of Intrinsic Transcription Termination." Molecular cell **3**(4): 495-504.

Hajnsdorf, E. and I. V. Boni (2012). "Multiple activities of RNA-binding proteins S1 and Hfq." Biochimie **94**(7): 1544-1553.

Hamilton, A. J. and D. C. Baulcombe (1999). "A species of small antisense RNA in posttranscriptional gene silencing in plants." Science **286**(5441): 950-952.

Hammer, B. K. and B. L. Bassler (2007). Regulatory small RNAs circumvent the conventional quorum sensing pathway in pandemic Vibrio cholerae. **104:** 11145-11149.

Hammerle, H.*, et al.* (2012). "Structural and biochemical studies on ATP binding and hydrolysis by the Escherichia coli RNA chaperone Hfq." PLoS One **7**(11): e50892.

Han, D.*, et al.* (2009). "SSO1450--a CAS1 protein from Sulfolobus solfataricus P2 with high affinity for RNA and DNA." FEBS Lett **583**(12): 1928-1932.

Hanada, T.*, et al.* (2013). "CLP1 links tRNA metabolism to progressive motor-neuron loss." Nature **495**(7442): 474-480.

Harley, C. B. and R. P. Reynolds (1987). "Analysis of E.Coli Pormoter sequences." Nucl. Acids Res. **15**(5): 2343-2361.

Hassani, K.*, et al.* (2014). "Absence of metalloprotease GP63 alters the protein content of Leishmania exosomes." PLoS.One. **9**(4): 95007.

Haussecker, D.*, et al.* (2010). "Human tRNA-derived small RNAs in the global regulation of RNA silencing." RNA. **16**(4): 673-695.

Hengge-Aronis, R. (1993). "Survival of hunger and stress: The role of rpoS in early stationary phase gene regulation in E. coli." Cell **72**(2): 165-168.

Hengge-Aronis, R. (1996). "Back to log phase: sigma S as a global regulator in the osmotic control of gene expression in Escherichia coli." Mol Microbiol **21**(5): 887-893.

Hengge-Aronis, R. (2002). "Recent insights into the general stress response regulatory network in Escherichia coli." J Mol Microbiol Biotechnol **4**(3): 341-346.

Hengge-Aronis, R. (2002). A Role for the $\sigma^S$ Subunit of RNA Polymerase in the Regulation of Bacterial Virulence

Genes and Proteins Underlying Microbial Urinary Tract Virulence. Genes and Proteins Underlying Microbial Urinary Tract Virulence. L. Emoődy, T. Pál, J. Hacker and G. Blum-Oehler, Springer US. **485:** 85-93.

Hengge-Aronis, R. (2002). "Signal Transduction and Regulatory Mechanisms Involved in Control of the σS (RpoS) Subunit of RNA Polymerase." Microbiology and Molecular Biology Reviews **66**(3): 373-395.

Herr, A. J.*, et al.* (2005). "RNA Polymerase IV Directs Silencing of Endogenous DNA." Science **308**(5718): 118-120.

Herwaldt, B. L. (1999). "Leishmaniasis." The Lancet **354**(9185): 1191-1199.

Holley, R. W. (1966). "The nucleotide sequence of a nucleic acid." Sci Am **214**(2): 30-39.

Hooper, D. C. (2001). "Emerging mechanisms of fluoroquinolone resistance." Emerg Infect Dis **7**(2): 337-341.

Horwich, M. D.*, et al.* (2007). "The Drosophila RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC." Current Biology **17**(14): 1265-1272.

Hsu, L. M.*, et al.* (1985). "Escherichia coli 6S RNA gene is part of a dual-function transcription unit." J Bacteriol **161**(3): 1162-1170.

Huang, X.*, et al.* (2013). "Characterization of human plasma-derived exosomal RNAs by deep sequencing." BMC.Genomics **14**: 319.

Huffman, J. L. and R. G. Brennan (2002). "Prokaryotic transcription regulators: more than just the helix-turn-helix motif." Current Opinion in Structural Biology **12**(1): 98-106.

Huntzinger, E.*, et al.* (2005). "Staphylococcus aureus RNAIII and the endoribonuclease III coordinately regulate spa gene expression." The EMBO Journal **24**(4): 824-835.

Hutvágner, G.*, et al.* (2001). "A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA." Science **293**(5531): 834-838.

Ikeda, Y.*, et al.* (2011). "Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli." Mol Microbiol **79**(2): 419-432.

Ishihama, A. (2000). "FUNCTIONAL MODULATION OF ESCHERICHIA COLI RNA POLYMERASE." Annual Review of Microbiology **54**(1): 499-518.

Ishino, Y.*, et al.* (1987). "NUCLEOTIDE-SEQUENCE OF THE IAP GENE, RESPONSIBLE FOR ALKALINE-PHOSPHATASE ISOZYME CONVERSION IN ESCHERICHIA-COLI, AND IDENTIFICATION OF THE GENE-PRODUCT." Journal of Bacteriology **169**(12): 5429-5433.

Ivanov, P.*, et al.* (2011). "Angiogenin-induced tRNA fragments inhibit translation initiation." Mol.Cell **43**(4): 613-623.

Jaffe, A.*, et al.* (1997). The Escherichia coli histone-like protein HU affects DNA initiation, chromosome partitioning via MukB, and cell division via MinCDE. **179:** 3494-3499.

Jenjaroenpun, P.*, et al.* (2013). "Characterization of RNA in exosomes secreted by human breast cancer cell lines using next-generation sequencing." PeerJ. **1**: e201.

Jiang, F. (2005). "Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila." Genes Dev. **19**: 1674-1679.

Jung, Y.-D.*, et al.* (2013). "Retroelements: molecular features and implications for disease." Genes & Genetic Systems **88**(1): 31-43.

Junghae, M. and J. G. Raynes (2002). "Activation of p38 mitogen-activated protein kinase attenuates Leishmania donovani infection in macrophages." Infect.Immun. **70**(9): 5026-5035.

Kaer, K. and M. Speek (2013). "Retroelements in human disease." Gene **518**(2): 231-241.

Karlsson, M.*, et al.* (2001). ""Tolerosomes" are produced by intestinal epithelial cells." Eur J Immunol **31**(10): 2892-2900.

Kazantsev, A. V. and N. R. Pace (2006). "Bacterial RNase P: a new view of an ancient enzyme." Nat Rev Microbiol **4**(10): 729-740.

Keeling, P. J.*, et al.* (2005). "The tree of eukaryotes." Trends Ecol Evol **20**(12): 670-676.

Keiler, K. C. (2008). Biology of trans-Translation. **62:** 133-151.

Keiler, K. C.*, et al.* (1996). "Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA." Science **271**(5251): 990-993.

Kennedy, W. P.*, et al.* (2007). "Mechanism for De Novo RNA Synthesis and Initiating Nucleotide Specificity by T7 RNA Polymerase." Journal of Molecular Biology **370**(2): 256-268.

Ketting, R. F.*, et al.* (2001). "Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans." Genes & Development **15**(20): 2654-2659.

Kim, K. S. and Y. Lee (2004). "Regulation of 6S RNA biogenesis by switching utilization of both sigma factors and endoribonucleases." Nucleic Acids Res **32**(20): 6057-6068.

Köhler, K.*, et al.* (2015). "Structural and mechanistic characterization of 6S RNA from the hyperthermophilic bacterium Aquifex aeolicus." Biochimie(0).

Kolosova, N.*, et al.* (2004). "Isolation of high-quality RNA from gymnosperm and angiosperm trees." Biotechniques **36**(5): 821-824.

Konarska, M. M., and Sharp P. A. (1989). " Replication of RNA by the DNA-dependent RNA polymerase of phage T7". Cell **57** (3):423-431

Kowal, J.*, et al.* (2014). "Biogenesis and secretion of exosomes." Curr.Opin.Cell Biol. **29C**: 116-125.

Kowalczyk, M. S.*, et al.* (2012). "Molecular biology: RNA discrimination." Nature **482**(7385): 310-311.

Krieg, P. A. (1996). A Laboratory guide to RNA: isolation, analysis, and synthesis. New York, Wiley-Liss.

Kuhnel, K. and B. F. Luisi (2001). "Crystal structure of the Escherichia coli RNA degradosome component enolase1." Journal of Molecular Biology **313**(3): 583-592.

Kumar, P.*, et al.* (2014). "Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets." BMC Biol. **12**(1): 78.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-685.

Lagos-Quintana, M.*, et al.* (2001). "Identification of novel genes coding for small expressed RNAs." Science **294**(5543): 853-858.

Lai, F. and R. Shiekhattar (2014). "Where long noncoding RNAs meet DNA methylation." Cell Res **24**(3): 263-264.

Lake, J. A. (1981). "The ribosome." Scientific American **245**(2): 84-97.

Lake, J. A. (1985). "Evolving Ribosome Structure: Domains in Archaebacteria, Eubacteria, Eocytes and Eukaryotes." Annual Review of Biochemistry **54**(1): 507-530.

Lalaouna, D.*, et al.* (2014). "Regulatory RNAs Involved in Bacterial Antibiotic Resistance." PLoS Pathog **10**(8): e1004299.

Lalaouna, D.*, et al.* (2013). "Regulatory RNAs and target mRNA decay in prokaryotes." Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1829**(6â€“7): 742-747.

Lambertz, U.*, et al.* (2015). "Small RNAs derived from tRNAs and rRNAs are highly enriched in exosomes from both old and new world Leishmania providing evidence for conserved exosomal RNA Packaging." BMC Genomics **16**(1): 151.

Lange, R. and R. Hengge-Aronis (1994). "The cellular concentration of the sigma S subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability." Genes & Development **8**(13): 1600-1612.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat.Methods **9**(4): 357-359.

Lau, N. C.*, et al.* (2001). "An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans." Science **294**(5543): 858-862.

Lavelle, F.*, et al.* (1977). "A pulse-radiolysis study of the catalytic mechanism of the iron-containing superoxide dismutase from Photobacterium leiognathi." Biochem J **161**(1): 3-11.

Law, J. A. and S. E. Jacobsen (2010). "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." Nat Rev Genet **11**(3): 204-220.

Le Derout, J.*, et al.* (2003). "Hfq affects the length and the frequency of short oligo(A) tails at the 3' end of Escherichia coli rpsO mRNAs." Nucl. Acids Res. **31**(14): 4017-4023.

Lee, J.*, et al.* (2013). "Regulation of transcription from two ssrS promoters in 6S RNA biogenesis." Molecules and Cells **36**(3): 227-234.

Lee, R. C. and V. Ambros (2001). "An extensive class of small RNAs in Caenorhabditis elegans." Science **294**(5543): 862-864.

Lee, S. Y.*, et al.* (1978). "Small stable RNAs from Escherichia coli: evidence for the existence of new molecules and for a new ribonucleoprotein particle containing 6S RNA." J Bacteriol **133**(2): 1015-1023.

Lee, Y. (2003). "The nuclear RNase III Drosha initiates microRNA processing." Nature **425**: 415-419.

Lee, Y.*, et al.* (2002). "MicroRNA maturation: stepwise processing and subcellular localization." EMBO J. **21**: 4663-4670.

Lee, Y. S.*, et al.* (2004). "Distinct Roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways." Cell **117**(1): 69-81.

Lee, Y. S.*, et al.* (2009). "A novel class of small RNAs: tRNA-derived RNA fragments (tRFs)." Genes & Development **23**(22): 2639-2649.

Lehninger, A. L.*, et al.* (2008). Lehninger principles of biochemistry. New York, W.H. Freeman.

Lewis, B. P.*, et al.* (2003). "Prediction of mammalian microRNA targets." Cell **115**: 787-798.

Li, H.*, et al.* (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics. **25**(16): 2078-2079.

Li, J., *et al.* (2005). "Methylation Protects miRNAs and siRNAs from a 3′-End Uridylation Activity in Arabidopsis." Current Biology **15**(16): 1501-1507.

Li, Z. and D. A. Brow (1993). "A rapid assay for quantitative detection of specific RNAs." Nucleic acids research **21**(19): 4645-4646.

Li, Z., *et al.* (2012). "Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs." Nucleic Acids Res. **40**(14): 6787-6799.

Li, Z., *et al.* (1998). 3' Exoribonucleolytic trimming is a common feature of the maturation of small, stable RNAs in Escherichia coli. **95:** 2856-2861.

Lim, B., *et al.* (2014). "RNase III Controls mltD mRNA Degradation in Escherichia coli." Current Microbiology **68**(4): 518-523.

Liou, G.-G., *et al.* (2002). "DEAD Box RhlB RNA Helicase Physically Associates with Exoribonuclease PNPase to Degrade Double-stranded RNA Independent of the Degradosome-assembling Region of RNase E." Journal of Biological Chemistry **277**(43): 41157-41162.

Liu, Q., *et al.* (2003). "R2D2, a Bridge Between the Initiation and Effector Steps of the Drosophila RNAi Pathway." Science **301**(5641): 1921-1925.

Liu, X. and C. T. Martin (2009). "Transcription Elongation Complex Stability: The Topological Lock." Journal of Biological Chemistry **284**(52): 36262-36270.

Lodish, H. F. and J. E. Darnell (2000). Molecular cell biology. New York, W.H. Freeman.

Lorenz, C., *et al.* (2010). "Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts." Nucl. Acids Res. **38**(11): 3794-3808.

Lundgren, O., *et al.* (2000). "Role of the Enteric Nervous System in the Fluid and Electrolyte Secretion of Rotavirus Diarrhea." Science **287**(5452): 491-495.

Lye, L. F., *et al.* (2010). "Retention and loss of RNA interference pathways in trypanosomatid protozoans." PLoS.Pathog. **6**(10): e1001161.

Ma, J., *et al.* (2002). "Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures." J Bacteriol **184**(20): 5733-5745.

Magoc, T. and S. L. Salzberg (2011). "FLASH: fast length adjustment of short reads to improve genome assemblies." Bioinformatics. **27**(21): 2957-2963.

Makarova, K. S.*, et al.* (2006). "A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action." Biol Direct **1**: 7.

Makarova, K. S.*, et al.* (2011). "Evolution and classification of the CRISPR-Cas systems." Nat Rev Microbiol **9**(6): 467-477.

Malone, C. D. and G. J. Hannon "Small RNAs as Guardians of the Genome." Cell **136**(4): 656-668.

Mantel, P. Y.*, et al.* (2013). "Malaria-infected erythrocyte-derived microvesicles mediate cellular communication within the parasite population and with the host immune system." Cell Host.Microbe **13**(5): 521-534.

Marquardt, W.*, et al.* (1999). Parasitology and Vector Biology, Academic Press.

Marraffini, L. A. and E. J. Sontheimer (2010). "Self versus non-self discrimination during CRISPR RNA-directed immunity." Nature **463**(7280): 568-571.

Martin, R.*, et al.* (1998). Measurement of Rate of Protein Synthesis In Vitro. Protein Synthesis, Springer New York. **77:** 243-256.

Maslov, D. A. (2010). "Complete set of mitochondrial pan-edited mRNAs in Leishmania mexicana amazonensis LV78." Mol.Biochem.Parasitol. **173**(2): 107-114.

Matranga, C.*, et al.* (2005). "Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes." Cell **123**(4): 607-620.

Matthews, K. S. (1996). "The whole lactose repressor." Science **271**(5253): 1245-1246.

Mattick, J. S. (2001). "Non-coding RNAs: the architects of eukaryotic complexity." EMBO Rep **2**(11): 986-991.

Mauri, M. and S. Klumpp (2014). "A Model for Sigma Factor Competition in Bacterial Cells." PLoS Computational Biology **10**(10): e1003845.

Maute, R. L.*, et al.* (2013). "tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma." Proc.Natl.Acad.Sci.U.S.A **110**(4): 1404-1409.

McAdam, M. E.*, et al.* (1977). "A pulse-radiolysis study of the manganese-containing superoxide dismutase from Bacillus stearothermophilus. A kinetic model for the enzyme action." Biochem J **165**(1): 71-79.

McConville, M. J.*, et al.* (2007). "Living in a phagolysosome; metabolism of Leishmania amastigotes." Trends Parasitol. **23**(8): 368-375.

McCulloch, S. D. and T. A. Kunkel (2008). "The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases." Cell research **18**(1): 148-161.

Michael, A.*, et al.* (2010). "Exosomes from human saliva as a source of microRNA biomarkers." Oral Dis. **16**(1): 34-38.

Michaux, C.*, et al.* (2014). "Physiological roles of small RNA molecules." Microbiology **160**(Pt 6): 1007-1019.

Miranda, K. C.*, et al.* (2014). "Massively parallel sequencing of human urinary exosome/microvesicle RNA reveals a predominance of non-coding RNA." PLoS.One. **9**(5): e96094.

Morin, R. D.*, et al.* (2008). "Comparative analysis of the small RNA transcriptomes of Pinus contorta and Oryza sativa." Genome Res **18**(4): 571-584.

Morita, M.*, et al.* (1999). Heat-Induced Synthesis of Ï‚32 inEscherichia coli: Structural and Functional Dissection ofrpoH mRNA Secondary Structure. **181:** 401-410.

Murakami, K. S. (2013). "The X-ray Crystal Structure of Escherichia Coli RNA Polymerase Sigma70 Holoenzyme." Journal of Biological Chemistry.

Murakami, K. S. and S. A. Darst (2003). "Bacterial RNA polymerases: the wholo story." Curr Opin Struct Biol **13**(1): 31-39.

Murray, H. W.*, et al.* (2005). "Advances in leishmaniasis." The Lancet **366**(9496): 1561-1577.

Nagano, T.*, et al.* (2008). "The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin." Science **322**(5908): 1717-1720.

Nandan, D.*, et al.* (1999). "Activation of phosphotyrosine phosphatase activity attenuates mitogen-activated protein kinase signaling and inhibits c-FOS and nitric oxide synthase expression in macrophages infected with Leishmania donovani." Infect.Immun. **67**(8): 4055-4063.

Nandan, D.*, et al.* (2002). "Leishmania EF-1alpha activates the Src homology 2 domain containing tyrosine phosphatase SHP-1 leading to macrophage deactivation." J.Biol.Chem. **277**(51): 50190-50197.

Nashimoto, H. and H. Uchida (1985). "DNA sequencing of the Escherichia coli ribonuclease III gene and its mutations." Molecular and General Genetics MGG **201**(1): 25-29.

Neusser, T.*, et al.* (2010). "Depletion of the non-coding regulatory 6S RNA in E. coli causes a surprising reduction in the expression of the translation machinery." BMC Genomics **11**: 165.

Ngo, H.*, et al.* (1998). "Double-stranded RNA induces mRNA degradation in Trypanosoma brucei." Proc Natl Acad Sci U S A **95**(25): 14687-14692.

Nicholson, A. W. (1999). "Function, mechanism and regulation of bacterial ribonucleases." FEMS Microbiology Reviews **23**(3): 371-390.

Nishida, K. M. (2007). "Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad." RNA **13**: 1911-1922.

Nissen, P.*, et al.* (2000). "The structural basis of ribosome activity in peptide bond synthesis." Science **289**(5481): 920-930.

Nitzan, M.*, et al.* (2014). "Global Regulation of Transcription by a Small RNA: A Quantitative View." Biophysical Journal **106**(5): 1205-1214.

Nogi, Y.*, et al.* (1991). "Synthesis of large rRNAs by RNA polymerase II in mutants of Saccharomyces cerevisiae defective in RNA polymerase I." Proc Natl Acad Sci U S A **88**(9): 3962-3966.

Noller, H. F.*, et al.* (1992). "Unusual resistance of peptidyl transferase to protein extraction procedures." Science **256**(5062): 1416-1419.

Nolte-'t Hoen, E. N.*, et al.* (2012). "Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions." Nucleic Acids Res. **40**(18): 9272-9285.

Novina, C. D. and P. A. Sharp (2004). "The RNAi revolution." Nature **430**(6996): 161-164.

Doudna, J. A., *et al.* (2015). " Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity." Nature **519**: 193–198.

Obbard, D. J.*, et al.* (2006). "Natural Selection Drives Extremely Rapid Evolution in Antiviral RNAi Genes." Current Biology **16**(6): 580-585.

Orgel, L. E. (2004). "Prebiotic chemistry and the origin of the RNA world." Crit Rev Biochem Mol Biol **39**(2): 99-123.

Oviedo Ovando, M.*, et al.* (2014). "In vitro characterization of 6S RNA release-defective mutants uncovers features of pRNA-dependent release from RNA polymerase in E. coli." Rna **20**(5): 670-680.

Palade, G. E. (1955). "A small particulate component of the cytoplasm." J Biophys Biochem Cytol **1**(1): 59-68.

Panchapakesan, S. S. and P. J. Unrau (2012). "E. coli 6S RNA release from RNA polymerase requires sigma70 ejection by scrunching and is orchestrated by a conserved RNA hairpin." RNA **18**(12): 2251-2259.

Panepinto, J.*, et al.* (2009). "Sec6-dependent sorting of fungal extracellular exosomes and laccase of Cryptococcus neoformans." Molecular Microbiology **71**(5): 1165-1176.

Panepinto, J.*, et al.* (2009). "Sec6-dependent sorting of fungal extracellular exosomes and laccase of Cryptococcus neoformans." Mol Microbiol **71**(5): 1165-1176.

Patton, J. T. (2008). Segmented double-stranded RNA viruses. Norfolk, UK, Caister Academic Press.

Peacock, C. S.*, et al.* (2007). "Comparative genomic analysis of three Leishmania species that cause diverse human disease." Nat.Genet. **39**(7): 839-847.

Pederson, T. (2010). "Regulatory RNAs derived from transfer RNA?" RNA **16**(10): 1865-1869.

Pegtel, D. M.*, et al.* (2010). "Functional delivery of viral miRNAs via exosomes." Proc Natl Acad Sci U S A **107**(14): 6328-6333.

Pélisson, A.*, et al.* (2007). "A Novel Repeat-Associated Small Interfering RNA-Mediated Silencing Pathway Downregulates Complementary Sense gypsy Transcripts in Somatic Cells of the Drosophila Ovary." Journal of Virology **81**(4): 1951-1960.

Phadtare, S. and M. Inouye (2004). "Genome-Wide Transcriptional Analysis of the Cold Shock Response in Wild-Type and Cold-Sensitive, Quadruple-csp-Deletion Strains of Escherichia coli." Journal of Bacteriology **186**(20): 7007-7014.

Pope, S. M. and C. Lasser (2013). "Toxoplasma gondii infection of fibroblasts causes the production of exosome-like vesicles containing a unique array of mRNA and miRNA transcripts compared to serum starvation." J.Extracell.Vesicles. **2**.

Prados-Rosales, R.*, et al.* (2011). "Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice." J.Clin.Invest **121**(4): 1471-1483.

Py, B.*, et al.* (1996). "A DEAD-box RNA helicase in the Escherichia coli RNA degradosome." Nature **381**(6578): 169-172.

Qiu, Y.*, et al.* (2013). "Characterizing the interplay between multiple levels of organization within bacterial sigma factor regulatory networks." Nat Commun **4**: 1755.

Rabinowits, G.*, et al.* (2009). "Exosomal microRNA: a diagnostic marker for lung cancer." Clin.Lung Cancer **10**(1): 42-46.

Raina, M. and M. Ibba (2014). "tRNAs as regulators of biological processes." Front Genet. **5**: 171.

Rajagopalan, R.*, et al.* (2006). "A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana." Genes & development **20**(24): 3407-3425.

Ramig, R. F. (2004). "Pathogenesis of Intestinal and Systemic Rotavirus Infection." Journal of Virology **78**(19): 10213-10220.

Reifur, L.*, et al.* (2012). "Distinct subcellular localization of tRNA-derived fragments in the infective metacyclic forms of Trypanosoma cruzi." Mem.Inst.Oswaldo Cruz **107**(6): 816-819.

Revelles, O.*, et al.* (2013). "The Carbon Storage Regulator (Csr) System Exerts a Nutrient-Specific Control over Central Metabolism in Escherichia coli Strain Nissle 1917." PLoS ONE **8**(6): e66386.

Revyakin, A.*, et al.* (2006). "Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching." Science **314**(5802): 1139-1143.

Revyakin, A.*, et al.* (2006). "ABORTIVE INITIATION AND PRODUCTIVE INITIATION BY RNA POLYMERASE INVOLVE DNA SCRUNCHING: Single-molecule-nanomanipulation experiments--in which magnetic tweezers are used to monitor RNA-polymerase-induced DNA unwinding with ~1 bp resolution and ~1 s temporal resolution--show that both abortive initiation and productive initiation by RNA polymerase involve DNA â€œscrunching.â€• ." Science (New York, N.Y.) **314**(5802): 1139-1143.

Rhoades, M. W. (2002). "Prediction of plant microRNA targets." Cell **110**: 513-520.

Rice, P.*, et al.* (2000). "EMBOSS: the European Molecular Biology Open Software Suite." Trends Genet. **16**(6): 276-277.

Richardson, J. P. (2002). "Rho-dependent termination and ATPases in transcript termination." Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression **1577**(2): 251-260.

Richmond, T. J. and C. A. Davey (2003). "The structure of DNA in the nucleosome core." Nature **423**(6936): 145-150.

Riggs, A. D. (1975). "X inactivation, differentiation, and DNA methylation." Cytogenetic and Genome Research **14**(1): 9-25.

Rinn, J. L. and H. Y. Chang (2012). "Genome regulation by long noncoding RNAs." Annu Rev Biochem **81**: 145-166.

Rio, D. C. (2010). RNA : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

Roberts, A.*, et al.* (2011). "Identification of novel transcripts in annotated genomes using RNA-Seq." Bioinformatics. **27**(17): 2325-2329.

Robertson, H. D.*, et al.* (1968). Purification and Properties of Ribonuclease III from Escherichia coli. **243:** 82-91.

Rodriguez-Valera, F.*, et al.* (2009). "Explaining microbial population genomics through phage predation." Nat Rev Microbiol **7**(11): 828-836.

Rose, M.*, et al.* (1990). Methods in yeast genetics: a laboratory course manual. NY, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Rose, M. D.*, et al.* (1990). Methods in yeast genetics: a laboratory course manual. Cold Spring Harbor, N.Y, Cold Spring Harbor Laboratory Press.

Ruggero, K.*, et al.* (2014). "Small Noncoding RNAs in Cells Transformed by Human T-Cell Leukemia Virus Type 1: a Role for a tRNA Fragment as a Primer for Reverse Transcriptase." Journal of Virology **88**(7): 3612-3622.

Saenger, W. (1984). Principles of nucleic acid structure. New York, Springer-Verlag.

Saito, K. (2006). "Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome." Genes Dev. **20**: 2214-2222.

Sambrook, J. and D. W. Russell (2001). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

Sano, T.*, et al.* (1992). "Identification of multiple structural domains regulating viroid pathogenicity." Proc Natl Acad Sci U S A **89**(21): 10104-10108.

Schaechter, M.*, et al.* (1958). "DEPENDENCY ON MEDIUM AND TEMPERATURE OF CELL SIZE AND CHEMICAL COMPOSITION DURING BALANCED GROWTH OF SALMONELLA-TYPHIMURIUM." Journal of General Microbiology **19**(3): 592-606.

Schageman, J.*, et al.* (2013). "The complete exosome workflow solution: from isolation to characterization of RNA cargo." Biomed.Res.Int. **2013**: 253957.

Schnolzer, M.*, et al.* (1985). "Correlation between structure and pathogenicity of potato spindle tuber viroid (PSTV)." EMBO J **4**(9): 2181-2190.

Sen, G. L. and H. M. Blau (2006). "A brief history of RNAi: the silence of the genes." Faseb J **20**(9): 1293-1299.

Shaham, S. (2006). "Worming into the cell: Viral reproduction in Caenorhabditiselegans." Proceedings of the National Academy of Sciences of the United States of America **103**(11): 3955-3956.

Sharma, U. K. and D. Chatterji (2010). "Transcriptional switching in Escherichia coli during stress and starvation by modulation of sigma activity." FEMS Microbiol Rev **34**(5): 646-657.

Sharp, P. A. (2009). "The Centrality of RNA." Cell **136**(4): 577-580.

Shephard, L.*, et al.* (2010). "Binding and release of the 6S transcriptional control RNA." Rna **16**(5): 885-892.

Shi, H.*, et al.* (2004). "Argonaute protein in the early divergent eukaryote Trypanosoma brucei: control of small interfering RNA accumulation and retroposon transcript abundance." Mol Cell Biol **24**(1): 420-427.

Shine, J. and L. Dalgarno (1974). "The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites." Proc Natl Acad Sci U S A **71**(4): 1342-1346.

Silverman, J. M.*, et al.* (2008). "Proteomic analysis of the secretome of Leishmania donovani." Genome Biol **9**(2): R35.

Silverman, J. M.*, et al.* (2010). "An exosome-based secretion pathway is responsible for protein export from Leishmania and communication with macrophages." J Cell Sci **123**(Pt 6): 842-852.

Silverman, J. M.*, et al.* (2010). "Leishmania exosomes modulate innate and adaptive immune responses through effects on monocytes and dendritic cells." The Journal of Immunology **185**(9): 5011-5022.

Silverman, J. M. and N. E. Reiner (2011). "Exosomes and other microvesicles in infection biology: organelles with unanticipated phenotypes." Cell Microbiol. **13**(1): 1-9.

Simon, A.*, et al.* (2015). "HTSeq - A Python framework to work with high-throughput sequencing data." Bioinformatics.

Simpson, R. J.*, et al.* (2008). "Proteomic profiling of exosomes: current perspectives." Proteomics **8**(19): 4083-4099.

Skarstad, K. and T. Katayama (2013). Regulating DNA Replication in Bacteria.

Skokos, D.*, et al.* (2003). "Mast cell-derived exosomes induce phenotypic and functional maturation of dendritic cells and elicit specific immune responses in vivo." J Immunol **170**(6): 3037-3045.

Skokos, D.*, et al.* (2001). "Mast cell-dependent B and T lymphocyte activation is mediated by the secretion of immunologically active exosomes." J Immunol **166**(2): 868-876.

Sobala, A. and G. Hutvagner (2011). "Transfer RNA-derived fragments: origins, processing, and functions." Wiley.Interdiscip.Rev.RNA. **2**(6): 853-862.

Sobala, A. and G. Hutvagner (2013). "Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells." RNA.Biol. **10**(4): 553-563.

Sobrero, P. and C. Valverde (2012). "The bacterial protein Hfq: much more than a mere RNA-binding factor." Crit Rev Microbiol **38**(4): 276-299.

Sommerville, J. (1986). "Nucleolar structure and ribosome biogenesis." Trends in Biochemical Sciences **11**(11): 438-442.

Sorensen, M. A.*, et al.* (1998). "Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in Escherichia coli in vivo." Journal of Molecular Biology **280**(4): 561-569.

Srivastava, R. A. K.*, et al.* (1992). "Characterization of the RNA processing enzyme RNase III from wild type and overexpressing Escherichia coli cells in processing natural RNA substrates." International Journal of Biochemistry **24**(5): 737-749.

Srivatsan, A. and J. D. Wang (2008). "Control of bacterial transcription, translation and replication by (p)ppGpp." Current Opinion in Microbiology **11**(2): 100-105.

Stead, M. B.*, et al.* (2011). "Analysis of Escherichia coli RNase E and RNase III activity in vivo using tiling microarrays." Nucl. Acids Res. **39**(8): 3188-3203.

Steuten, B.*, et al.* (2014). "Regulation of transcription by 6S RNAs: insights from the Escherichia coli and Bacillus subtilis model systems." RNA Biol **11**(5): 508-521.

Steuten, B.*, et al.* (2013). "Mapping the Spatial Neighborhood of the Regulatory 6S RNA Bound to Escherichia coli RNA Polymerase Holoenzyme." Journal of Molecular Biology **425**(19): 3649-3661.

Steuten, B. and R. Wagner (2012). "A conformational switch is responsible for the reversal of the 6S RNA-dependent RNA polymerase inhibition in Escherichia coli." Biological Chemistry.

Storz, G. and R. Hengge (2000). Bacterial stress responses. Washington, D.C., ASM Press.

Storz, G.*, et al.* (2004). "Controlling mRNA stability and translation with small, noncoding RNAs." Curr Opin Microbiol **7**(2): 140-144.

Sukhodolets, M. V. and S. Garges (2003). "Interaction of Escherichia coli RNA Polymerase with the Ribosomal Protein S1 and the Sm-like ATPase Hfq." Biochemistry **42**(26): 8022-8034.

Surgucheva, I.*, et al.* (2012). "gamma-Synuclein: seeding of alpha-synuclein aggregation and transmission between cells." Biochemistry **51**(23): 4743-4754.

Sydow, J. F. and P. Cramer (2009). "RNA polymerase fidelity and transcriptional proofreading." Current Opinion in Structural Biology **19**(6): 732-739.

Thibonnier, M.*, et al.* (2008). "Translation in *Helicobacter pylori:* Essentiality of Ribosome Rescue and Requirement of Protein Tagging for Stress Resistance and Competence." PLoS ONE **3**(11): 3810.

Trotochaud, A. E. and K. M. Wassarman (2004). "6S RNA function enhances long-term cell survival." J Bacteriol **186**(15): 4978-4985.

Tsui, H.-C. T.*, et al.* (1994). "Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12." Molecular Microbiology **13**(1): 35-49.

Twu, O.*, et al.* (2013). "Trichomonas vaginalis exosomes deliver cargo to host cells and mediate hostratioparasite interactions." PLoS.Pathog. **9**(7): e1003482.

Ullu, E.*, et al.* (2004). "RNA interference in protozoan parasites." Cell Microbiol **6**(6): 509-519.

Vagin, V. V. (2006). "A distinct small RNA pathway silences selfish genetic elements in the germline." Science **313**: 320-324.

Valadi, H.*, et al.* (2007). "Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells." Nat Cell Biol **9**(6): 654-659.

Van Rompay, A. R.*, et al.* (2000). "Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases." Pharmacology & Therapeutics **87**(2â€“3): 189-198.

van Zandbergen, G.*, et al.* (2004). "Cutting edge: neutrophil granulocyte serves as a vector for Leishmania entry into macrophages." The Journal of Immunology **173**(11): 6521-6525.

Venkataramanan, K. P.*, et al.* (2013). "The Clostridium small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in C. acetobutylicum." BMC Genomics **14**: 849.

Vila, J.*, et al.* (2002). "Are Quinolone-Resistant Uropathogenic Escherichia coli Less Virulent?" Journal of Infectious Diseases **186**(7): 1039-1042.

Vilcek, J. (2006). "Fifty Years of Interferon Research: Aiming at a Moving Target." Immunity **25**(3): 343-348.

Visvader, J. E. and R. H. Symons (1986). "Replication of in vitro constructed viroid mutants: location of the pathogenicity-modulating domain of citrus exocortis viroid." EMBO J **5**(9): 2051-2055.

Voinnet, O. (2009). "Origin, biogenesis, and activity of plant microRNAs." Cell **136**(4): 669-687.

Vojtech, L.*, et al.* (2014). "Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions." Nucleic Acids Res. **42**(11): 7290-7304.

Wang, J. D. and P. A. Levin (2009). "Metabolism, cell growth and the bacterial cell cycle." Nat Rev Micro **7**(11): 822-827.

Wassarman, K. M. (2007). "6S RNA: a small RNA regulator of transcription." Curr Opin Microbiol **10**(2): 164-168.

Wassarman, K. M. and R. M. Saecker (2006). "Synthesis-mediated release of a small RNA inhibitor of RNA polymerase." Science **314**(5805): 1601-1603.

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

Wehner, S.*, et al.* (2014). "Dissemination of 6S RNA among Bacteria." RNA Biology **11**(11): 1467-1478.

Weiberg, A.*, et al.* (2013). "Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways." Science **342**(6154): 118-123.

Weigel, C.*, et al.* (1997). "DnaA protein binding to individual DnaA boxes in the Escherichia coli replication origin, oriC." The EMBO Journal **16**(21): 6574-6583.

Werner, F. and D. Grohmann (2011). "Evolution of multisubunit RNA polymerases in the three domains of life." Nat Rev Microbiol **9**(2): 85-98.

Wiedenheft, B.*, et al.* (2012). "RNA-guided genetic silencing systems in bacteria and archaea." Nature **482**(7385): 331-338.

Wiedenheft, B.*, et al.* (2009). "Structural Basis for DNase Activity of a Conserved Protein Implicated in CRISPR-Mediated Genome Defense." Structure **17**(6): 904-912.

Williams, K. P. and D. P. Bartel (1996). Phylogenetic analysis of tmRNA secondary structure. **2:** 1306-1310.

Windbichler, N.*, et al.* (2008). "Isolation of small RNA-binding proteins from E. coli: Evidence for frequent interaction of RNAs with RNA polymerase." RNA Biology **5**(1): 30-40.

Winkler, W.*, et al.* (2002). "Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression." Nature **419**(6910): 952-956.

Winkler, W. C. and R. R. Breaker (2005). "Regulation Of Bacterial Gene Expression By Riboswitches." Annual Review of Microbiology **59**(1): 487-517.

Wittmann, H. G. (1983). "Architecture of Prokaryotic Ribosomes." Annual Review of Biochemistry **52**(1): 35-65.

Wutz, A. (2011). "Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation." Nat Rev Genet **12**(8): 542-553.

Xie, Z. (2004). "Genetic and functional diversification of small RNA pathways in plants." PLoS Biol. **2**: E104.

Xie, Z.*, et al.* (2003). "Negative feedback regulation of Dicer-like1 in Arabidopsis by microRNA-guided mRNA degradation." Curr. Biol. **13**: 784-789.

Yakovchuk, P.*, et al.* (2006). Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. **34:** 564-574.

Yang, Z.*, et al.* (2006). "HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2[prime] OH of the 3[prime] terminal nucleotide." Nucleic Acids Res. **34**: 667-675.

Yarnell, W. S. and J. W. Roberts (1999). "Mechanism of Intrinsic Transcription Termination and Antitermination." Science **284**(5414): 611-615.

Yekta, S.*, et al.* (2004). "MicroRNA-directed cleavage of HOXB8 mRNA." Science **304**: 594-596.

Yeung, M. L.*, et al.* (2009). "Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid." Nucleic Acids Research **37**(19): 6575-6586.

Yi, R.*, et al.* (2003). "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs." Genes Dev. **17**: 3011-3016.

Yin, Y. W. and T. A. Steitz (2004). "The Structural Mechanism of Translocation and Helicase Activity in T7 RNA Polymerase." Cell **116**(3): 393-404.

Yu, B., *et al.* (2005). "Methylation as a Crucial Step in Plant microRNA Biogenesis." Science **307**(5711): 932-935.

Yusupov, M. M., *et al.* (2001). "Crystal structure of the ribosome at 5.5 A resolution." Science **292**(5518): 883-896.

Zamecnik, P. (1966). "The mechanism of protein synthesis and its possible alteration in the presence of oncogenic RNA viruses." Cancer Research **26**(1): 1-6.

Zamore, P. D., *et al.* (2000). "RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals." Cell **101**(1): 25-33.

Zhang, A., *et al.* (2002). "The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs." Mol Cell **9**(1): 11-22.

Zhang, A., *et al.* (2003). "Global analysis of small RNA and mRNA targets of Hfq." Molecular Microbiology **50**(4): 1111-1124.

Zhang, G., *et al.* (1999). "Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 Å Resolution." Cell **98**(6): 811-824.

Zhang, Z., *et al.* (2000). "A greedy algorithm for aligning DNA sequences." J.Comput.Biol. **7**(1-2): 203-214.

Zhou, Y. N., *et al.* (1992). "A mutant sigma 32 with a small deletion in conserved region 3 of sigma has reduced affinity for core RNA polymerase." J Bacteriol **174**(15): 5005-5012.

Zilberman, D., *et al.* (2003). "ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation." Science **299**(5607): 716-719.

# Appendix A.

# Isolation and Biochemical Analysis of Plant Small RNAs

This appendix is largely based on the manuscript; Ebhardt, H. A., M. O. Ovando, et al. (2012). "Isolation and biochemical analysis of plant small RNAs." Methods Mol Biol 894: 223-239.

M.O. O. was in charge of generating Figure A5, and writing the sections on TRIzol Small RNA Extraction, RNA Quantification and Quality Control, Gel Sample Loading, Denaturing Polyacrylamide Gel Electrophoresis, and Sodium Periodate Assay.


### *Abstract*

Small RNAs, defined as non-coding 20-30 nt long RNAs, are instrumental regulators of cellular processes in most eukaryotes. In this chapter we describe three methods for extracting small RNA from cells: a general method, one plant specific and a third particular to conifers. Further, protocols are given for the analysis and quantification of small RNAs using polyacrylamide gel-based approaches. A native streptavidin gel-shift assay, useful for measuring the relative amounts of multiple small RNAs simultaneously, is presented. To further characterize small RNAs biochemically, a sodium periodate assay probing for 2',3' hydroxyl groups on the 3' terminus of small RNAs is outlined.


### *1. Introduction*

Hamilton and Baulcombe first linked small RNAs to post-transcriptional gene silencing in plants in 1999 (Hamilton and Baulcombe 1999). Ever since, the physiological roles and biochemical properties of small RNAs have been actively studied in both plants and animals. Small 21-24 nt long RNAs are the affecter molecules for RNA silencing that in plants are responsible for a broad range of epigenetic and antiviral phenomena (Voinnet 2009). In many cases, studying this small RNA dependent gene regulation involves the isolation, biochemical characterization and sequencing (Lagos-Quintana, Rauhut *et al.* 2001, Lau, Lim *et al.* 2001, Lee and Ambros 2001) of small RNAs from specific tissues or whole organisms. Such approaches have revealed the presence of a ubiquitous 3′ terminal 2′-O-methyl group in plant small RNAs (Ebhardt, Thi *et al.* 2005, Yang, Ebright *et al.* 2006), uncovered the conservation of important micro RNAs (Rajagopalan, Vaucheret *et al.* 2006), and suggest unique patterns of small RNA processing in particular branches of the land plants (Dolgosheina, Morin *et al.* 2008, Morin, Aksay *et al.* 2008). Protocols for isolating small RNAs from various plant types plays a fundamental role in all of these endevours.

In this chapter, we outline three small RNA extraction protocols together with a variety of biochemical methods that can be used to simply characterize extracted small RNA populations. A general TRIzol extraction protocol, suitable for many plant and animal tissues, is outlined together with more dedicated protocols for plant and in particular conifer small RNA extractions. The biochemical analysis of small RNAs includes protocols on native and denaturing polyacrylamide gel electrophoresis (PAGE) analysis,

a streptavidin gel shift assay useful for measuring relative small RNA concentrations and a sodium periodate assay for studying the terminal 3′ modification of small RNA.

## *2. Materials*

Generally, all reagents prepared in this section should be handled so as to prevent RNase contamination. Gloves should be worn, ideally disposable plastic ware should be used and all solutions should be prepared with MilliQ water and then sterile filtered using 0.22 µm filters (i.e. Nalgene Filter unit, 115 ml, 0.22 µm, Cat. No. 121-0020). Handle organic compounds carefully, taking care to use a fumehood where appropriate. Tris buffers are most easily generated by first mixing 1 *M* stocks of Tris base and Tris-HCl according to pre-calculated tables (Sambrook and Russell 2001). All pH measurements are at 25 °C unless otherwise noted.

### *2.1. TRIzol Small RNA Extraction*

1. TRIzol Reagent (Invitrogen), 1-bromo-3-chloropropane (Sigma).
2. 75% EtOH solution.
3. 2-propanol.
4. Prepare and sterile filter the following stock solutions: 5 M NaCl, 1 M KCl, 1 M $Na_2HPO_4$ and 1 M $KH_2PO_4$.
5. Phosphate-buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$ and 2 mM $KH_2PO_4$. Before taking the solution to final volume, adjust pH to 7.4 with HCl (12). Sterile filter and dispense the solution into aliquots. Store the buffer at room temperature or freeze.

### *2.2. RNA Extraction from Plants*

1. Prepare and sterile filter: 10 *M* LiCl, 10% SDS, 500 m*M* EDTA, 3 *M* sodium acetate pH 5.2 and 1 *M* Tris-HCl (pH 9).
2. RNA extraction buffer A: 100 m*M* LiCl, 1% SDS, 10 m*M* EDTA, 100 m*M* Tris (pH 9), sterile filter and freeze.
3. Buffer saturated Phenol (Invitrogen).
4. Chloroform (Sigma).

### *2.3. Conifer RNA Extraction*

1. RNA extraction buffer B: 200 mM Tris-HCl (pH 8.5), 1.5% lithium dodecylsulfate, 300 mM LiCl, 10 mM disodium salt EDTA, 1% (w/v) NP40m (Tergitol Nonidet® P-40, Sigma NP40S). The solution is either sterile filtered or autoclaved for 20 min at 121 ºC. Before use add 5 mM thiourea, 1 mM aurintricarboxylic acid, and 10 mM DTT.
2. TE Buffer: 10 m*M* Tris-HCl (pH 8.0), 1 m*M* EDTA.
3. 10% (w/v) cetyltrimethylammonium bromide (CTAB), Sigma H-6269.
4. Poly (vinylpolypyrrolidone) (PVP) (Sigma P-6755).
5. 100% EtOH, 2-propanol.
6. Chloroform/isoamyl alcohol (24:1) (Sigma C-0549).
7. 3.3 *M* sodium acetate (pH 6.1).

### *2.4. RNA Quantification and Quality Control*

1. Gel staining: SYBR® GreenER™ (Invitrogen).
2. Radioactive labeling: Prepare and sterile filter: 1 *M* Tris-HCl (pH 7.6), 1 *M* $MgCl_2$, 1 *M* DTT.
3. 10 x PNK buffer: 700 m*M* Tris-HCl (pH 7.6), 100 m*M* $MgCl_2$, sterile filter, aliquot into 1 ml fractions and freeze.

4. Polynucleotide kinase Enzyme (New England Biolabs, M0201S; use 1 U of enzyme per 2.5 pmole of RNA).
5. [$\gamma$-$^{32}$P]-ATP (3,000 Ci/mmole) (PerkinElmer, Boston, MA).

### *2.5. Gel Sample Loading*

1. 2x denaturing gel loading buffer: formamide supplemented with 0.03% bromophenol blue, 0.03% Xylene Cyanol and 0.5 m*M* EDTA.
2. 6x native gel loading buffer: 10 m*M* Tris-HCl (pH 7.6), 0.03% bromophenol blue, 0.03% Xylene Cyanol FF, 60% Glycerol.
3. 1 *M* NaOH stock.

### *2.6. Denaturing Polyacrylamide Gel Electrophoresis*

1. Solution A: 25% 19:1 acrylamide:bisacrylamide in 6.667 *M* urea: 312.5 ml of 40% 19:1 acrylamide:bisacrylamide stock, 200.2 g of urea (FW: 60.06 g/mol), and add ~45 ml $dH_2O$ to make up to a final volume of 500 ml. Store the solution at 4°C (*see* Note 1).
2. Solution B: 6.667 *M* urea (400.4 g ultra-pure urea per L of Solution B).
3. Solution C: 10x TBE: 0.9 *M* Tris base, 0.9 *M* boric acid, 25 m*M* EDTA (108 g Tris base, 55 g boric acid, 50 ml of 0.5 *M* EDTA, pH 8 per L of 10x TBE) (Sambrook and Russell 2001).
4. 10% (w/v) ammonium persulfate (APS) (*see* Note 3), tetramethylethylenediamine (TEMED), Sigmacote (Sigma-Aldrich SL2) (*see* Note 4).

### *2.7. Small RNA Gel Shift*

1. 40% acrylamide:bisacrylamide 19:1 liquid stock (Omni Pur®) (see Note 1).
2. 5x TBN buffer: 450 m*M* Tris-base, 225 m*M* boric acid, 250 m*M* NaCl, pH 8.5, prepare by mixing (see Note 2).
3. 10% (w/v) fresh ammonium persulfate (APS) (see Note 3).
4. Tetramethylethylenediamine (TEMED, Sigma T7024) (see Note 4).
5. 10x PNK buffer (*see* Section 2.4 Quantification and Quality Control).
6. Biotinylated DNA oligonucleotides synthesized using 3'-BiotinTEG-CPG (Glen Research, Sterling, VA). Alternatively, DNA synthesis companies will normally offer 3'-Biotin labeled oligonucleotide synthesis. The biotinylated DNA oligos are complementary to the small RNA to be studied. For example, 5'-TGT GCT CAC TCT CTT CTG TCA TTT-3' is complementary to ath-MIR156, with two additional T residues added to the 3` terminus, which are not encoded in ath-MIR156. We rationalize that these additional T residues on the biotinylated 3' could facilitate the binding of streptavidin to the DNA/RNA duplex.
7. 1 mg Streptavidin (Sigma-Aldrich, S0677) was resuspended in 250 µl Binding Buffer (140 m*M* NaCl, 2.7 m*M* KCl, 10 m*M* $Na_2HPO_4$, 1.8 m*M* $KH_2PO_4$, 10% Glycerol, 2 m*M* DTT pH 7.5) and stored at -20°C.
8. 6x native gel loading buffer: *see* Section 2.5.

### *2.8. Sodium Periodate Assay*

1. 2x denaturing gel loading buffer (formamide) (*see* Section 2.5.).
2. 100 m*M* HEPES buffer, adjusted to pH 7.0 using HCl.
3. Freshly-prepared 1 *M* sodium periodate ($NaIO_4$).
4. 5 *M* NaCl.
5. 5` radiolabeled RNA (*see* Section 2.4.).

## *3. Methods*

*RNA Extraction*

There are a wide variety of protocols available for RNA extraction. In the present section, we describe some of the methods used in our laboratory to extract RNA. An extraction with TRIzol (*see* Section 3.1.) is probably the best general starting point for obtaining RNA in a short period of time. This method is ideal for cells and tissues that are easily disrupted, and/or have been grown in suspension. When cells and/or tissues have resistant structures, such as thickened cell walls, mechanical disruption of these structures is necessary. This is most easily achieved by grinding the tissue in a mortar in the presence of liquid nitrogen. Another method, in Section 3.2. (Extracting RNA from Plants), is more aggressive and uses hot phenol to maximize the recovery of RNA from more challenging samples that are difficult to mechanically disrupt. Still other samples, contain very high levels of polysaccharides and metabolites (i.e. polyphenolic compounds or terpenoids), which tend to co-precipitate with nucleic acids during the extraction procedure. These compounds can inhibit enzymatic reactions such as reverse transcription, and if UV absorbing, they can interfere with the determination of RNA concentration. The last extraction protocol in Section 3.3., takes into account the complex composition of difficult samples and is based on a protocol developed by Kolosova and colleagues (Kolosova, Miller *et al.* 2004) with some modifications.

*3.1. TRIzol RNA Extraction*

1. Approximately 50-100 mg of fresh tissue is ground into a fine powder with a pestle and mortar in the presence of liquid nitrogen. Transfer the powder to a tube that is precooled on ice and add 1 ml of TRIzol reagent. The resulting slurry is then made homogeneous by repetitive pipetting or vigorous vortexing. Alternatively, the TRIzol can be added directly to the still frozen powder and the resulting mixture ground until thawed. The resulting mixture is carefully transferred to an Eppendorf tube. When the starting material is a cell culture, pellet the cells by centrifugation at 5,000 g for 5-10 min. The pellet is rinsed with 2 to 5 ml of PBS and centrifuged again at 5,000 g for 5-10 min. The supernatant is discarded. It is important to know the cell density of your culture to determine the volume of reagent for the next step. Use 1 ml of TRIzol reagent per $5x10^6$ cells of animal or plant origin (*see* Notes 6 - 7).
2. Incubate the homogenous mixture resulting from step 1 for 5 min. at room temperature. Spin at 5,000 g for 5 min to pellet any particulate, transferring the supernatant to a fresh tube. Add 0.2 ml of 1-bromo-3-chloropropane per ml of TRIzol used above, vortex vigorously for 10 seconds and incubate for a further 10 min at room temperature.
3. Spin at 12,000 g for 15 min at 4°C. The suspension should separate into a lower organic pink phase, a sticky opaque DNA interphase and an upper aqueous phase.
4. Collect the upper aqueous layer avoiding the transfer of ANY interphase. Add NaCl solution to a final concentration of 0.3 *M* and vortex. Add 0.5 ml of 2-propanol per ml of TRIzol used initially (added in step 1 or 2), vortex and incubate for 10 min on ice (*see* Note 5).
5. Spin at 16,000 g for 30 min. at 4°C. Discard the supernatant and wash the pellet with 1 ml of ice-cold 75% EtOH per ml of TRIzol.
6. Spin at 16,000 g for 15 min at 4°C and carefully remove the EtOH wash.

7. Air dry the pellet slightly (at room temperature, for a few min), re-suspend in water or TE (50-100 μL) and determine the OD at 260 nm.

*3.2. Extracting RNA from Plants*

1. Grind approximately 0.5 g of fresh tissue into a fine powder with a pestle and mortar in the presence of liquid nitrogen. Transfer the powder to a tube that is pre-cooled on ice and add 1 ml of RNA extraction buffer A.
2. Add 1.5 ml of phenol preheated in boiling water. Note: that phenol can cause dangerous chemical burns, make sure to wear appropriate eye protection and gloves. Once the tube cools down, add 1.5 ml of chloroform, and invert repeatedly for 20 min using for example a Fisher Scientific Hematology/Chemistry Rotator Mixer Model 346. Centrifuge at 12,000 g for 10 min at 4°C.
3. Transfer the upper aqueous phase to a fresh tube and extract again with an equal volume of chloroform.
4. Nucleic acids are precipitated by addition of 0.1 vol. of 3 *M* sodium acetate, pH 5.2 and 3 volumes of 100% EtOH. Incubate for 20 min. at -80°C or 2 hours at -20°C to completely precipitate nucleic acids or O/N at -20°C, which can increase the RNA recovery.
5. Recover nucleic acids by centrifugation at 16,000 g for 30 min at 4°C. Let the pellet air dry and resuspend in 100-200 μL of dH2O or TE.

*3.3. Isolation of RNA from Conifers*

This protocol was developed by Kolosova and colleagues (Kolosova, Miller *et al.* 2004) and is used to extract high quality RNA from species where a high content of polysaccharides or secondary metabolites, such as polyphenolic compounds or terpenoids, are present. The protocol, with slight modifications, has been used successfully to extract RNA from such species as *P.contorta, P. glauca, and P. sitchensis* in our hands. From our experience, between 50 and 200 μg of RNA can be recovered per gram of conifer tissue.

1. Prior to the start of the extraction, a mortar and pestle should be chilled with liquid nitrogen. One gram of plant tissue together with 0.1 g of PVPP is ground under liquid nitrogen until a very fine powder is obtained.
2. The powder is transferred to a polypropylene tube with a clean spatula. Add 20 ml of extraction buffer B per gram of tissue and vortex vigorously. Snap-freeze this suspension in liquid nitrogen.
3. Thaw the suspension in a 37°C water bath and centrifuge at 3,000 g for 20 min at 4°C. If particles can be observed in the supernatant, filter through one layer of Kimwipe® tissue using a funnel. Keep the resulting liquid on ice. This is a tedious step that can be avoided by grinding the tissue very well at the liquid nitrogen stage.
4. Add one thirtieth (1/30) volume of 3.3 *M* sodium acetate (666 μL per 20 ml of buffer B) and 0.1 volume of 100% EtOH (2 ml). Store the sample on ice for 10 min. Centrifuge at 3,000 g for 30 min. at 4°C to precipitate the polysaccharides. This step can be skipped in spruce tissues due to their lower polysaccharide content (Kolosova, Miller *et al.* 2004).
5. Add one ninth (1/9) volume, or 2.2 ml of 3.3 *M* sodium acetate and 0.6 volume, or 12 ml , of ice cold 2-propanol are then added to the supernatant (~22 ml ) in order to precipitate nucleic acids. Store the tube at -80°C for 30 min. Collect the nucleic acid by centrifugation at 3,000 g for 45 min. at 4°C. Remove the supernatant and resuspend the pellet in 1.6 ml of TE buffer.

Add 1.6 ml of 5 *M* NaCl and keep on ice for approximately 30 min. with occasional vortexing until the pellet is dissolved.

6. Mix the sample with 800 μL of 10% cetyltrimethylammonium bromide (CTAB), and incubate for 5 min. at 65°C. This step removes residual polysaccharides.
7. Extract twice with an equal volume of chloroform/isoamyl alcohol (24:1). Centrifuge at 10,000 g for 10 min at 4°C each time.
8. Recover the supernatant and add one-fourth (1/4) of the volume of 10 *M* LiCl. Mix thoroughly and, in the case of spruce or pine needles, precipitation can be achieved after 2 h at -20°C. For poplar leaves and spruce bark or xylem the sample should be precipitated O/N at 4°C. Overnight precipitation increases the efficiency of the RNA precipitation. Recover the RNA by centrifugation at 3,000 g for 30 min at 4°C. From this step on, all samples originating from tissues that are rich in polysaccharides should be stored at 4°C to avoid co-precipitation of RNA and polysaccharides.
9. Remove the supernatant carefully and re-suspend the RNA pellet in 500 μL of TE buffer. Keep on ice and vortex occasionally until the pellet is dissolved; this process can take up to one hour.
10. Transfer the sample to a 2 ml tube and add 0.9 vol. of ice-cold 2-propanol and 0.1 vol. of 3.3 *M* sodium acetate. Store at -80°C for 30 min.
11. Collect the RNA pellet by centrifugation at 14,000 g for 10 min at 4°C. Allow the pellets to air dry for approximately 10 min. Resuspend the pellets in 100-200 μl of RNase-free water.

### *3.4. RNA Quantification and Quality Control*

1. Total RNA concentration is determined using a spectrophotometer at 260 nm where absorbance $A_{260nm}$ = 1 corresponds to 40 μg RNA/ml for a 1 cm pathlength (Sambrook and Russell 2001).
2. The integrity of the RNA is determined by running 1-2 μg of RNA, either in an agarose or polyacrylamide gel, and staining with SYBR Green. Dilute 10 μL of SYBR Green (assuming a 10,000x stock) in 100 ml of 1x TBE. Place the gel in a suitable container, add the diluted dye solution and agitate gently for approximately 30 min. No de-staining is required before imaging. When RNA integrity has been preserved, larger RNA molecules (ribosomal RNA, rRNA) can be clearly observed as distinct bands and any smearing will be kept to a minimum (Figure A1).
3. RNA quality can also be judged by radiolabeling a small quantity of RNA. The radiolabeling reaction is set up as follows: 2 μg or RNA, 1μL of [$\gamma$-$^{32}$P]-ATP (PerkinElmer, Boston, MA), 1x PNK buffer, 0.5 U/μL of polynucleotide kinase (PNK) (New England Biolabs, Ipswich, MA). Bring the reaction to a final volume of 10 μL with $dH_2O$, and incubate at 37°C for 30 min.
4. The quality of the smaller RNA fragments can be simply evaluated by loading the remaining radio-labeled RNA into a high-percentage (20-22%) polyacrylamide gel under denaturing conditions (*see* Section 3.5. and 3.6.).

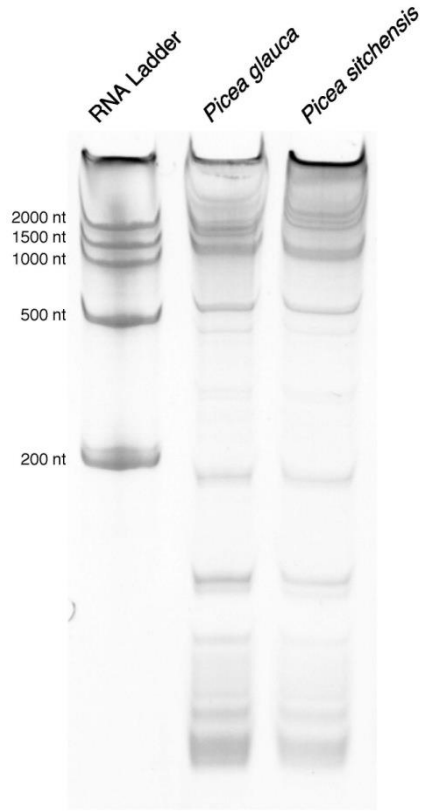



***Figure A1 15 % denaturing polyacrylamide gel loaded with 1 μg of RNA from White Spruce (Picea glauca) and Sitka Spruce (Picea sitchensis)***

The RNA ladder was constructed by transcribing PCR products of defined length with T7 RNA polymerase. The gel was stained using SYBR Green and photographed using an Alpha Imager EC camera system.

*3.5. Gel Casting*

Glass plates for your particular gel electrophoresis apparatus should be selected that are free of substantial chips and cracks. It is important that the glass plates be clean and free of RNases. Scrub the plates with a detergent such as 'Count Off' and a sponge or brush until completely clean. Do a final rinse with $dH_2O$, followed with EtOH and dry with a Kimwipe. If this is the first time the glass plates are being used for RNA work, it is recommended that they be soaked O/N in 1 *M* NaOH prior to use. Rinse them thoroughly the next morning with $dH_2O$ and EtOH. Clean plates will avoid the formation of air bubbles when pouring the polyacrylamide solution. It can also be very useful when preparing thin gels to spread a few drops of Sigmacote onto ONE of the plates using a Kimwipe.

For analytical gels, use thin spacers (0.4 mm), while preparative gel are typically gels are 1 to 2 mm thick (depending on the sample volume requirements). Large paper binder clips are very useful to hold spacers and glasses together during casting. Table A1 indicates the proportions of the different reagents depending on the percentage of gel

needed. As soon as you add the APS and TEMED, mix the solution, pour the mixture between the glass plates and insert the comb. Allow the gel to polymerize for at least 1 hour. Polymerization can be confirmed by checking for a change in index of refraction around the wells of the gel.

***Table 1. Proportions of reagents needed for casting denaturing polyacrylamide gels***

$C_{final}$ is the desired final gel percentage, $V_{final}$ is the desired final gel volume (remember to include some extra volume for wastage), $V_A$ is the volume of Solution A, and $V_B$ is the volume for Solution B, $V_C$ the volume for solution C (see Section: 2.6.).

| Reagent | Relative amounts | example: 20 ml 10 % |
|---|---|---|
| Solution A | $V_A = \frac{C_{final} \cdot V_{final} \cdot 0.9}{22.4}$ | 8.0 mL |
| Solution B | $V_B = V_{final} \cdot 0.9 - V_A$ | 10.0 mL |
| Solution C | $V_c = V_{final} \cdot 0.1$ | 2.0 mL |
| 10 % APS | 0.24 % | 48 μL |
| TEMED | 0.10 % | 20 μL |

Once the polyacrylamide has polymerized (which can be determined by examining the comb area of the gel closely), remove the clips and comb carefully. Rinse the wells thoroughly with 1x TBE running buffer prior to sample loading (*see* Section 3.7. for more details). After pre-running the gel for 20-30 min, a gel ~1 mm thick be can loaded with up to 10 μg of total RNA in a single 6 mm wide lane. Run gels at 0.4-0.5 mW per $cm^2$ of gel surface area. Running at higher wattages may cause glass plates to crack. The running time for the gel will depend on the percentage of the gel and RNA size you are interested in and is best judged by monitoring the mobilities of gel loading dyes while the gel is running (*see* Table A2).

***Table 2. Approximate dye and nucleic acid mobilities for denaturing PAGE***

For example, Bromophenol blue migrates with the mobility of a 12-nt long RNA in a 10 % denaturing PAGE gel.

| denaturing PAGE | Bromophenol blue | Xylene Cyanol FF |
|---|---|---|
| 5 % | 35-nt | 130-nt |
| 6 % | 26-nt | 106-nt |
| 8 % | 19-nt | 70-80-nt |
| 10 % | 12-nt | 55-nt |
| 20 % | 8-nt | 28-nt |

*3.6. Denaturing PAGE: Sample Preparation and Loading*

RNA duplexes are thermodynamically much more stable than a corresponding dsDNA. As a result incomplete denaturation of RNA secondary structures can have a dramatic and undesired effect on gel mobility in denaturing PAGE. While it is important to minimize salt concentration in samples as much as possible and to use formamide gel loading dye, incomplete dehybridization can be a routine problem. One particularly effective mechanism to dehybridize RNA duplex structure is to briefly treat the sample with base immediately prior to gel loading. Figure A2 shows the difficulty associated in dehybridizing a 21-nt radiolabelled RNA from a longer RNA transcript in high salt conditions (in this case RNA transcription buffer). Even through the urea in the polyacrylamide gel is a denaturant, loading the sample with native loading dye fails to completely denature the duplex. Denaturing formamide gel loading buffer performs slightly better and is further improved by heating the sample to 99°C for 2 min immediately prior to loading. Adding 50 m*M* NaOH to regular formamide gel loading buffer immediately prior to loading however results in nearly complete dehybridization. This approach can be particularly important when working with small RNAs from biological samples where many small RNAs might be expected to be hybridized to longer RNA transcripts (*i.e.* virally-derived small RNAs).

Important: RNA is unstable in strong base and degrades quickly at elevated temperatures. Prepare and load samples promptly using this method - do not heat! Upon entering the gel, the RNA is buffered to the pH of the gel system.

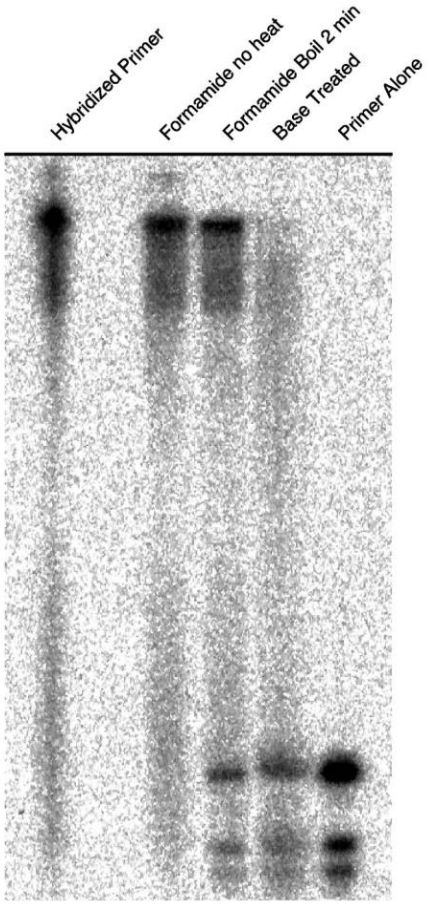



***Figure A2 High salt effectively prevents dehybridization in an 8 % denaturing PAGE***

A radiolabeled 21-nt primer was hybridized to a ~100-nt RNA transcript in the presence of T7 RNA polymerase transcription buffer. This sample was then mixed with native gel loading dye (most left lane) or formamide gel loading buffer, which was also boiled for 2 min. prior to loading. The addition of 50 m*M* NaOH to formamide gel loading buffer does effectively denature the sample. Most right lane shows the primer loaded by itself.

*3.7. Small RNA Streptavidin Gel Shift*

The streptavidin gel-shift assay makes possible the simultaneous measurement of multiple small RNA expression levels in a single gel, in contrast to a northern blot where only a single probe can be used at any one time (Ebhardt and Unrau 2009). Furthermore, shifted RNA bands can be excised from the gel for subsequent biochemical analysis (Alwine, Kemp *et al.* 1977, Sambrook and Russell 2001, Kolosova, Miller *et al.* 2004, Ebhardt, Thi *et al.* 2005, Rajagopalan, Vaucheret *et al.* 2006, Yang, Ebright *et al.* 2006, Dolgosheina, Morin *et al.* 2008, Morin, Aksay *et al.* 2008) (Figure A3). While the Brow gel (Li and Brow 1993) also allows analysis with many probes in

parallel, it does not simply allow an analysis of relative RNA concentrations as here the DNA probe is radiolabeled and not the RNA sample.

This protocol requires total RNA obtained by using one of the extraction protocols found in Sections 3.1 to 3.3:

1. To size fractionate small RNAs 10-30 nts in length, preparative gel extraction is used. A 20% preparative denaturing PAGE is loaded with 3-6 µg of total RNA per square mm of gel well area using formamide denaturing gel loading buffer.
2. To visualize the desired size range, multiple synthetic 5' radiolabeled RNA oligos, *e.g.* 18, 21, 24 and 30 nts, synthesized by Thermo Scientific Dharmacon, Lafayette, are mixed with an equal amount of total RNA and loaded into a separate lane. Loading synthetic size markers, together with an equal amount of total RNA, ensures that the salt concentration in both lanes are identical and the synthetic size markers truly reflect the migration of small RNA in the total RNA lane.
3. Following electrophoresis, the gel is exposed to a phosphorimager screen to visualize the desired size range. An autoradiogram representing a 1:1 scale of the gel (100% printout) is printed and a gel slice corresponding to small RNAs 18-30 nts in length is excised as shown in Figure A3.
4. RNA contained in the gel slice (from step 3) is eluted O/N in 300 m*M* NaCl and 5 m*M* EDTA at room temperature.
5. The resulting RNA is precipitated by the addition of 1 µL 12.5 µg/ml glycogen (Ambion, Austin, TX) and 2.5 volumes of anhydrous EtOH. Samples are placed at minus 80°C for 20 min before pelleting the labeled RNA at 12,000 g for 30 min.

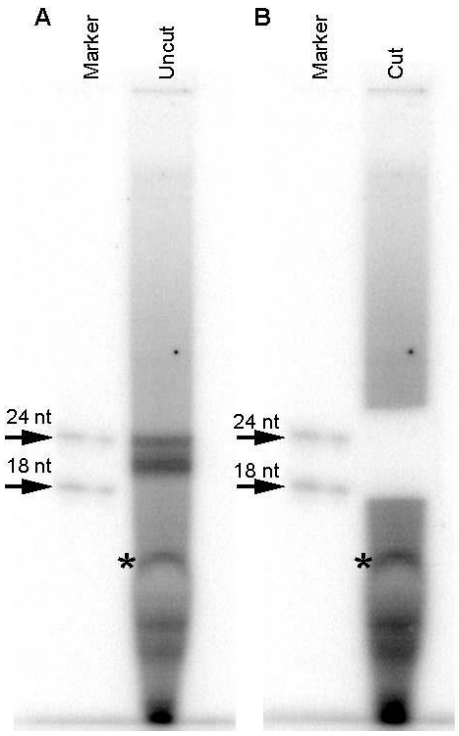



***Figure A3 Gel purification of plant small RNAs***

(A) Autoradiogram of a 20 % denaturing PAGE. Radiolabeled synthetic marker RNA, 18 and 24 nts in length, mixed with unlabeled total RNA. Total RNA from tobacco was radiolabeled and shows a characteristic bimodal distribution of small RNAs: 21/22 and 24 nts. (B) Autoradiogram after cutting out gel fragment corresponding to plant small RNAs. Note the salt front (as indicated by an asterisk) below the 18 nt marker in lanes containing the radiolabeled RNA. If the gel is overloaded, this salt front will shift higher, preventing the proper separation of small RNAs.

6. Radiolabel five pmoles of gel-purified small RNA using an excess of [$\gamma$-$^{32}$P]-ATP (PerkinElmer, Boston, MA) and 5U of T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) in 1 x PNK buffer (*see* Note 8).
7. Incubate the reaction at 37°C for 15 min., followed by 65°C for 25 min. to heat inactivate the T4 polynucleotide kinase.
8. For annealing the RNA to the biotinylated DNA probe, add 1 µl of 5x TBN buffer, 2 µl of radiolabeled RNA (from step 2), and 1 µl of 10 µ*M* complementary 3'-biotin-labeled oligonucleotide in a 600 µL centrifuge tube. Incubate the tube containing RNA and biotin-labeled DNA oligonucleotide for 90 sec. at 90°C in a thermocycler (*e.g*. PerkinElmer PTC-100).
9. Slowly cool down the reaction to 30°C, at which point 1 µl of streptavidin (4 mg/ml) is added and incubated for 5 min. at room temperature.
10. After completion of step 9, 1 µl of 6x native gel loading dye is added and the samples are directly loaded onto a 10% native PAGE using flat end loading tip (Costar Gel Loading Tips Flat, CS004854) and run for 2 hours, or longer, at 15 W in an environment-controlled room set at 4°C (*see* Note 9).
11. The final buffer concentration within the gel and running buffer is 1x TBN.

12. The native PAGE is usually performed in a cold room at lower power (gel volume ~20 ml polyacrylamide 0.4 mm thick spacers, 15 W). If a cold room is not available, gels can be run at lower power at room temperature. Check the glass plates periodically throughout the run to ensure that they are not warm to the touch. This is particularly important if potential structures within the gel might become denatured due to heat.
13. Native gels run quite slowly, typically 2 hours of electrophoresis at 15 W will result in migration of the Bromophenol Blue front by 2-3 inches. The single-stranded RNA will have migrated the furthest, while the supershifted streptavidin-biotin-DNA-RNA complex migrated minimally in the gel.
14. Once the gel run is complete, while the gel is still assembled, dry the glass plates and wells thoroughly using Kimwipes or paper towels. Once all liquid is removed, which could potentially interfere with the autoradiogram, take off one glass plate, leaving the native gel on the other glass plate. Wrap the gel and glass plate in saran wrap (any brand that does not allow leakage) and expose the gel to a digital phosphorimager screen at 4°C (*see* Note 9 for further details). The gel's autoradiogram is scanned in from the phosphorimager screen using a phosphorimager, *e.g*. Typhoon Storm 820, with an example of a native gel shift assay given in Figure A4.

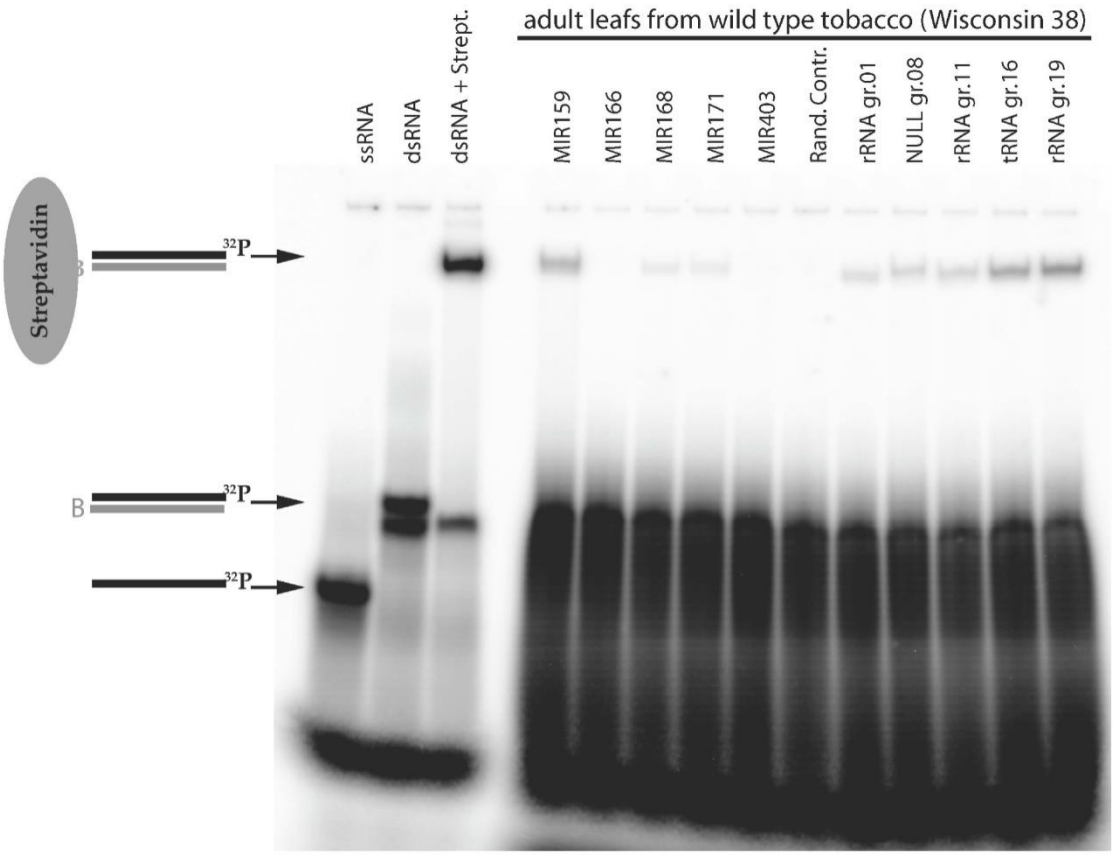


***Figure A4 Native streptavidin gel shift. The lanes on the left-hand side demonstrate the three steps of the gel shift assay***

From single-stranded RNA, to double-stranded DNA-RNA hybrid to the super shift with streptavidin. Complementary biotinylated DNA probes were designed against various small RNAs and incubated with radiolabeled total small RNA from tobacco. The shifted material represents the relative abundance of each RNA to the total population of small RNA in the tobacco sample.

*3.8. Sodium Periodate Assay*

The Sodium Periodate Assay test the 3' ends of RNA for the existence of unmodified 2' and 3' hydroxyls. Oxidization by periodate results in a labile dialdehyde, which upon heating, results in the β-elimination of the terminal nucleoside. In a high-percentage denaturing PAGE, the cleaved RNA will run ~2 nts faster than its original RNA. If however, either or both 2' and 3' hydroxyls are modified, *e.g*. with a 2'-O-methyl group, the sodium periodate reaction will not proceed and the radio-labeled RNA will run at the same mobility as the untreated RNA (*see* Figure A5).

1. Radiolabel 5 pmoles of gel-purified small RNA using an excess of [γ-$^{32}$P]-ATP (PerkinElmer, Boston, MA) and 5 U of T4 polynucleotide kinase in 1x PNK buffer (*see* Note 8).
2. Incubate the reaction at 37°C for 15 min., followed by 65°C for 25 min. to heat inactivate the T4 polynucleotide kinase.
3. Take 1 µl of phosphorylated RNA and add 9 µl 100 m*M* HEPES, pH 7.0. Take 5 µL and add 5 µL 200 m*M* NaCl together with 10 µL 2x denaturing gel loading buffer (the no periodate control).
4. To the remaining 5 µL of sample add 5 µL of 200 m*M* $NaIO_4$. The reaction mixture from is incubated at 22 °C in the dark for 10 min. (*e.g*. in a PCR machine) after which time 10 µL of denaturing gel loading buffer (formamide) is added.
5. Both samples +/- periodate (steps 4 and 3 respectively) are heated to 99 °C for 30 min. *(e.g*. in a PCR machine).
6. All samples are then loaded and resolved on a high-percentage denaturing PAGE, *e.g*. 20-22%.

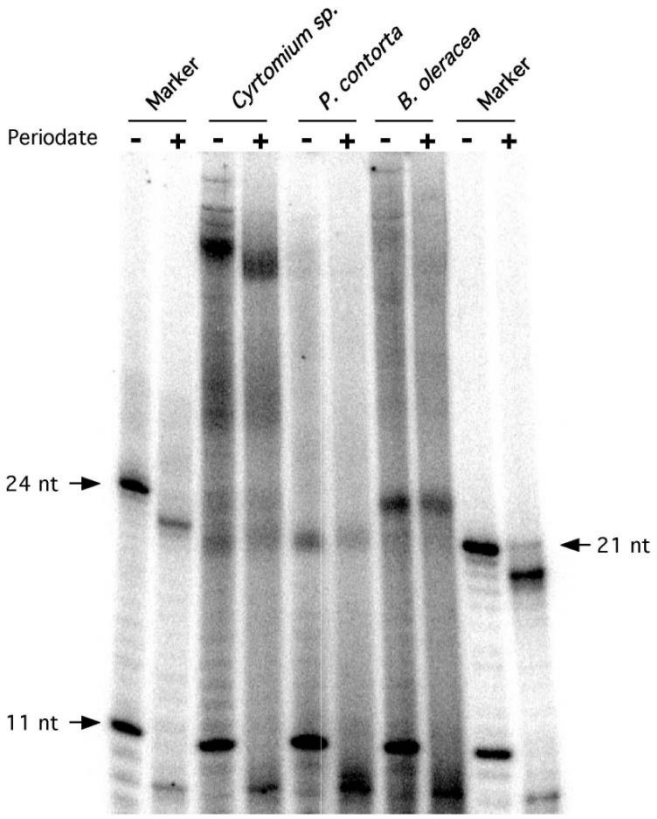



***Figure A5 Periodate cleavage of radiolabeled plant small RNA***

Total RNA extracted from the indicated species was radiolabeled and then subjected to periodate as described. The small RNAs produced by these species are substantially uncleaved as judged by this gel. An 11 nt synthetic RNA was added to each sample as an internal control for periodate-dependent cleavage.

## *4. Notes*

1. Acrylamide is a neurotoxin. Avoid use of the powdered form and take care to avoid ingestion or contact with skin.
2. 5x TBN buffer: This buffer is stored as a 5x stock at room temperature for 1-2 months only.
3. 10% APS: This can be stored for several weeks at 4°C. Fresh APS crystals should be granular and free flowing; APS that is clumped, should be discarded.
4. Sigmacoat contains heptane: avoid inhalation, ingestion and eye or skin contact. Wear gloves and use only in well-ventilated areas, *e.g*. in a fume hood).
5. General precaution: Keep the RNA on ice at all times.
6. Whenever possible, the cells should be cultivated in axenic media to avoid contamination with foreign nucleic acid. If you are working with large volumes of cell culture and TRIzol, use disposable polypropylene tubes.
7. If you are not sure about the number of cells in the culture, add excess of TRIzol to be confident that all enzymes and RNases are inactivated.

8. After radiolabeling an RNA sample, it is generally a good practice to load several different dilutions on a test gel to determine the optimal contrast for the experiment at hand.
9. Freezing of native gels during exposure will lead to cracks, while storing the gel at room temperature during exposure might cause unwanted diffusion.

## Acknowledgments

Cancer Research Fellowship from the Alberta Cancer Research Institute [to H.A.E.]. Operating grants from the Natural Sciences and Engineering Research Council of Canada and the Michael Smith Foundation [to P.J.U.].

# Appendix B.

# Supplementary Data Files for Chapter 4

***Supplementary Figure 1 – Leishmania braziliensis exosomes contain RNA***

Filename: Supplementary_Figure_1.tif

Description: Exosomes were purified from *L. braziliensis* axenic amastigote culture supernatant as described in the Materials and Methods. RNA was extracted from exosomes with phenol-chloroform and then analyzed. A. Agilent Bioanalyzer RNA length profiles of exosome RNA alongside total RNA, B. RNA inside exosomes is resistant to degradation. Prior to RNA extraction, intact exosomes were left untreated or treated with either RNase A or TritonX-100 or both. Samples were then subjected to RNA extraction and run on the Agilent Bioanalyzer. Arrowhead indicates internal 25 nt marker. nt = nucleotides.

***Supplementary Figure 2 – Leishmania donovani exosomes can be efficiently stained with green fluorescent SYTO RNASelect dye***

Filename: Supplementary_Figure_2.tif

Description: Exosomes were purified from 400 ml supernatant of *L. donovani* axenic amastigotes and stained with a membrane permeant, green fluorescent RNA-specific dye (as described in Methods). A sample each of stained and unstained *L. donovani* exosomes were then examined by microscopy using an Axioplan II epifluorescence microscope equipped with 63×/1.4 Plan-Apochromat objective (Carl Zeiss Inc). Images were recorded using an AxioCam MRm Camera coupled to the AxioVision software Version 4.8.2 (Carl Zeiss Inc.).

***Supplementary Figure 3 – Length histograms of reads mapping to rRNA genes***

Filename: Supplementary_Figure_3.tif

***Supplementary Table B1A: sequences found in L. donovani exosomes***

Filename: Supplementary_table_1A.zip

Description:

Sheet 1: Collapsed Reads including full nucleotide sequences

Sheet 2: Results of the Bowtie2 alignment of the *L. donovani* exosome sequencing library against the LdBPK reference genome

Sheet 3: Results of the Bowtie2 alignment of the *L. donovani* exosome sequencing library against the LmjF reference genome

*"cRead ID" refers to the identity of the collapsed read, which is composed of the cRead number, followed by the number of copies. "Strand" refers to the strand the reads were aligning with, plus = top strand, minus = bottom strand.*

***Supplementary Table B1B: sequences found in L. braziliensis exosomes***

Filename: Supplementary_table_1B.zip

Description:

Sheet 1: Collapsed Reads including full nucleotide sequences

Sheet 2: Results of the Bowtie2 alignment of the *L. braziliensis* exosome sequencing library against the LbrM reference genome

Sheet 3: Results of the Bowtie2 alignment of the *L. braziliensis* exosome sequencing library against the LmjF reference genome

*"cRead ID" refers to the identity of the collapsed read, which is composed of the cRead number, followed by the number of copies. "Strand" refers to the strand the reads were aligning with, plus = top strand, minus = bottom strand.*

***Supplementary Table 2A***

Filename: Supplementary_table_2A.zip

Description: Results of the Bowtie2 alignment of the unaligned reads from the *L. donovani* exosome sequencing library against the NCBI nucleotide collection database

***Supplementary Table 2B***

Filename: Supplementary_table_2B.zip

Description: Results of the Bowtie2 alignment of the unaligned reads from the *L. braziliensis* exosome sequencing library against the NCBI nucleotide collection database

***Supplementary Table 3: Overall alignment statistics***

Filename: Supplementary_table_3.docx

***Supplementary Table 4***

Filename: Supplementary_table_4.xlsx

Description: Mapping of reads to rRNA genes based on Bowtie 2 alignments with the LmjF reference genome.

***Supplementary Table 5***

Filename: Supplementary_table_5.xlsx

Description: Putative ORFs found in the novel transcripts.

***Supplementary Table 6A***

Filename: Supplementary_table_6A.xlsx

Description: Results of blast search of all reads from *L. donovani* exosome library against human and vector genomes

***Supplementary Table 6B***

Filename: Supplementary_table_6B.xlsx

Description: Results of blast search of all reads from *L. braziliensis* exosome library against human and vector genomes.

***Supplementary Table 7***

Filename: Supplementary_table_7.xlsx

Description: Results of blast search of all reads from *L. donovani* and *L. braziliensis* exosome library against SLACS and TATEs database.

***Supplementary Table 8***

Filename: Supplementary_table_8.xlsx

Description: Hits against miRbase: Top 1000 most abundant reads in both libraries were blast searched against human and mouse miRNAs.