Genome wide analyses of the *Escherichia coli* primary and secondary transcriptomes

David Romero A.

Submitted in accordance with the requirement for the degree of PhD

The University of Leeds, Faculty of Biological Sciences

March 2014

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to thank my supervisor, Kenny McDowall, who saw that I had the potential needed to do a PhD. He was always there to guide and support me; not only in an academic, but also in a personal way. I am grateful as he never gave up on me and even when I was beginning to doubt myself; he still had confidence in me and pushed me forward. In addition to this, his intelligence and academic knowledge kept me in check and helped me think more before talking. I will not forget the effort he put into my scientific formation, and the cost he paid for me having me in his lab as since my arrival his hair has thinned and greyed considerably.

I would also like to thank the truly good friends that I have made during this part of my life, especially: Justin; Ayad; Tom; Louise, who didn't go out much with us, but when she did was great to have around and Mia. Their support and long talks, both scientific and not, along with the pub times are something that will always bring a smile to my face, with an occasional frown, especially when thinking back at some of Justin's jokes which, disturbingly enough I usually find humour in, when I look back at the times spent together.

I want to thank the O'Neill group as since we started sharing a lab and office it has been a nice experience and has led to a good work environment and good nights out. I also want to thank the Stockley's for the time during which we shared a lab.

Special thanks to our collaborators Lira, Kathleen, Cecilia and the Kaberdin group: Vladimir and Olatz. In addition, I would like to thank David Westhead and Vijay for helping us write the scripts needed to process and map the massive sequencing files generated by the sequencing runs.

Massive thanks goes to all my family, especially my parents who have always supported me and despite the distance have always found a way to be close and supportive. I can say that without their unconditional help, guidance and support I would be in a completely different path in life. Once again thanks for all the support through the years. I am sure Kenny differs and holds them responsible for shipping me to his lab and costing him his "younger looks".

I would also like to thank, last but not least, my girlfriend, Victoria who in addition to cooking and feeding me when I forgot to do so myself, which was pretty much every day, has been very understanding and supportive during my write-up period as well as taken a fair deal of moodiness from me.

Abstract

Escherichia coli K12 serves as an important model for studying systems that are important to bacteria in their own right as well as those that are conserved in 'higher' organisms, which are more difficult and costly to study. Like many model organisms, the genome of K12 has been sequenced, producing a catalogue of protein-coding and stable-RNA genes that enabled study using 'omic' approaches. This has led to a rapid expansion of our knowledge of patterns of gene expression and their dependency on growth conditions, cell physiology and individual genes. However, the underlying networks of gene regulation are less well understood, but are known to involve the control of steps in RNA processing and degradation as well as transcription and translation. With this in mind, this thesis describes the development of an approach based on RNA sequencing that produces nucleotide-resolution transcriptome maps that distinguish sites that correspond to RNA processing and steps in degradation from those of transcription initiation, while incorporating all classes of RNA. Comparison with results obtained previously validated the approach, which has been applied already to the study of other bacterial species. Within the *E. coli* map, many new features were identified, such as previously undetected small RNAs and processing at a site associated with the production of specialised ribosomes, which may ensure the translation of leaderless mRNAs, which were also mapped. The approach also showed the benefit of incorporating steps that can differentiate the 5' status of transcripts in assigning sites of transcription initiation. RNA sequencing was also used to map sites of cleavage by RNase E, an essential endoribonuclease that is central to both the processing and degradation of RNA in bacteria and plant plastids. This aspect of the thesis has advanced from pilot studies to the point where the 'code' that determines one form of substrate recognition by RNase E is beginning to emerge. As a result of this success, equivalent data has been collected for other ribonucleases involved in RNA processing and degradation. Continuing analysis of the primary and secondary transcriptomes, consisting of native, unprocessed transcripts and of transcripts that have been modified from their native form via processing and/or degradation respectively, with the tools presented here promises to broaden and deepen our understanding of an important model organism.

Table of Contents

List of Figures

List of Tables

Abbreviations

ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
cm	centimetre
cDNA	complementary deoxyribonucleic acid
CTH	C-terminal half
Ci	curie
°C	degree Celsius
DNA	deoxyribonucleic acid
dRNA-seq	differential RNA-sequencing
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FRT-seq	flowcell reverse transcription sequencing
GEO	Gene Expression Omnibus
gRNA-seq	global RNA-sequencing
h	hour
kb	kilobase
kDa	kilodalton
LB	Luria-Bertani
Mbp	mega base pair
mRNA	messenger ribonucleic acid
µCi	microcurie
µg	microgram
µl	microlitre
mA	milliampere
mCi	millicurie
ml	millilitre
mM	millimolar
mmol	millimole
min	minute
M	molar
MEME	Multiple Em for Motif Elucidation
nm	nanometre
nM	nanomolar
NTH	N-terminal half
nt	nucleotide
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PNK	polynucleotide kinase
psi	pound per square inch
REP	repetitive extragenic palindromic
RT	reverse transcription
rpm	revolutions per minute

RNA	ribonucleic acid
rNTP	ribonucleotide triphosphate
rRNA	ribosomal ribonucleic acid
RBS	ribosome binding site
RBD	RNA binding domain
RLM-RT-PCR	RNA ligase-mediated reverse transcription PCR
sec	second
SD	Shine-Dalgarno
sRNA	small ribonucleic acid
SDS	sodium dodecyl sulphate
g	standard gravity
TEX	TerminatorTM 5’ phosphate-dependent exonuclease
TAP	tobacco acid pyrophosphatase
TA	toxin anti-toxin
TSS	transcription start site
TF	transcription factor
tRNA	transfer ribonucleic acid
tmRNA	transfer-messenger ribonucleic acid
Tris	tris(hydroxymethyl)aminomethane
TBE	Tris-borate-EDTA
TE	Tris-EDTA
TB	tuberculosis
UV	ultraviolet
U	unit
UCSC	University of California, Santa Cruz
V	volt
Vh	volt hour
v/v	volume/volume
w/v	weight/volume

Chapter 1

1 Introduction to RNA transcription and degradation

1.1 Overview

The Central “Dogma” of molecular biology, which is an explanation of the flow of genetic information (Crick, 1958; Crick, 1970), is often described as "DNA makes RNA makes protein" (Nirenberg, 2004). This particular description, which emphasises the role of transcription and translation, is an over simplification. It not only fails to include the flow of information from RNA to DNA through reverse transcription (Ahlquist, 2002), but also fails to describe the essential role of RNA processing in, for example, producing the RNA components of the translational machinery (Deutscher, 2009) and the critical role of mRNA degradation, which ensures that translation has to follow programming at the level of transcription (Dreyfus, 2009).

Over the last decade, the study of gene regulation has made extensive use of microarrays to provide information on transcript levels on a genome-wide scale (Schena et al., 1995; Bier and Kleinjung, 2001; Goldsmith and Dhanasekaran, 2004). Such transcriptome data when available for different growth conditions, stages of development (for organisms that change morphology), and altered genetic backgrounds has allowed the identification of co-expressed genes, which in turn has enhanced the identification of co-regulation (De Smet and Marchal, 2010; Storms et al., 2010). For example, numerous algorithms are available to identify sequences that are enriched upstream of co-regulated genes, and represent candidates for the binding of a shared transcription factor. Information on such cis-regulatory sites is being collated in databases for organisms ranging from *E. coli* (RegulonDB) to humans (ENCODE) (The_Encode_Project_Consortium, 2004; Salgado et al., 2006). With the obvious exception of splicing in eukaryotes, microarrays have provided scant information on the steps that control gene expression at the level of RNA processing and degradation. There are several main reasons for this, including insufficient resolution and sensitivity to detect processing and degradation intermediates and the inability to distinguish the pathways by which particular species are generated. However, these limitations no longer exist with the advent of RNA sequencing.

This chapter provides an introduction to the control of gene expression in *E. coli*, with a particular emphasis on those processes that register an effect at the level of the RNA. It

starts by providing an overview of transcription initiation and its control, then moves on to describe RNA degradation and the various components involved in this process, placing specific attention on RNase E. It finally presents the sequencing approach that inspired the work presented in this thesis alongside the objectives and aims. The direct control of translation is covered, but only in relation to antisense RNA regulators. An overview of RNA sequencing is also provided.

1.2 Transcription initiation

As indicated above, gene expression is tightly regulated at several levels in all living organisms. One of these levels of regulation is the process of transcription. *E. coli* codes for a single RNA polymerase that is composed of four sub-units (β, β' and two α subunits; $\alpha_2\beta\beta'$). In order to recognise specific promoter sequences, a sigma factor (σ) must be associated with this $\alpha_2\beta\beta$ RNA polymerase core to yield a holoenzyme (Eσ) which can initiate transcription (Browning and Busby, 2004).

1.2.1 **Sigma factors**

E. coli has 7 sigma factors (Table 1.1) that are grouped into two main families, based on their evolutionary relatedness and their gene targets (Wösten, 1998; Buck et al., 2000; Browning and Busby, 2004; Sharma and Chatterji, 2010). These sigma factors work in conjunction within the cell and compete to bind RNA polymerase, with which they bind with a range of affinities. The first family belong to the *E. coli* σ^{70}–type sigma factors (the number denotes mass in kDa), which are evolutionarily conserved in all bacteria and regulate most of the housekeeping genes throughout the bacterial life cycle (Gruber and Gross, 2003) as well as most genes during bacterial exponential growth (Buck et al., 2000; Browning and Busby, 2004; Sharma and Chatterji, 2010). In addition, extra cytoplasmic function (ECF) sigma factors, which are responsible for switching on the expression of genes involved in environment stress responses also belong to this family (Helmann, 2002; Brooks and Buchanan, 2008; Ho and Ellermeier, 2012). The second family corresponds to the σ^{54}–type sigma factors, which share no homology with σ^{70} and in addition to the core $\alpha_2\beta\beta'$ complex require ATP and enhancers to form a holoenzyme (Gruber and Gross, 2003). The latter group is involved mainly in the regulation of genes related to nitrogen metabolism (Buck et al., 2000). Whilst there are multiple representatives of the σ^{70} family per organism, the presence of more than one σ^{54} is rare (Wösten, 1998; Buck et al., 2000; Sharma and Chatterji, 2010). Also, worth mentioning is that an additional level of

regulation is achieved by sigma factor antagonists known as anti-sigma factors (Hughes and Mathee, 1998).

Name	Consensus sequence			Binding affinity
	−35	**Spacer**	**−10**	**K_d (nM)**
σ70-family				
σ^{70}, (σ^{D})	TTGACA	16–18	TATAAT	0.26
σ^{38}, (σ^{S})	N/D	N/D	CTATACT	4.26
σ^{32}, (σ^{H})	CTTGAAA	11–16	CCCATNT	1.24
σ^{28}, (σ^{F})	TAAA	N/D	GCCGATAA	0.74
σ^{24}, (σ^{E})*	GAACTT	16–17	TCTRA	2.43
σ^{19}, (σ^{FECI})*	N/D	N/D	N/D	2.43
σ54-family				
	−24	**Spacer**	**−12**	
σ^{54}, (σ^{N})	TGGCAC	N/D	TTGCW	0.3

Table 1.1 - Sigma factors in *E. coli.* Table presenting the sigma factors of *E. coli* alongside the consensus sequences they recognize relative to the site of transcription initiation, the spacer length between the consensus sequences and the binding affinity of each sigma factor to RNA polymerase. ECFs are indicated with an *. N = any base, R = A or G and W = A or T N/D=Not determined. This table was adapted from two publications (Wösten, 1998; Sharma and Chatterji, 2010).

1.2.2 **Consensus sequences involved in transcription initiation**

Sigma factors are essential for basal transcription initiation; thus, knowledge of the specific sequences, known as consensus sequences, that each of them recognise when assembled into a holo-RNA polymerase is essential if the regulation of gene expression is to be understood and modelled. Consensus sequences have been determined for sigma factors in *E. coli*. For example, the consensus sequences for σ^{70} are TATAAT, centred at position -10 (relative to site of transcription initiation) and TTGACA at position -35 (Hook-Barnard et al., 2006) whilst the consensus sequences for σ^{54} are TGGCA, centred at position -24 and TTGC at position -12. The particular sequence bound by a sigma factor has a major impact on the rate of transcription initiation. Indeed, the better the match between a promoter sequence and the consensus sequence for the sites recognised by a sigma factor, the stronger the promoter (Ishihama, 1990; Kobayashi et al., 1990).

1.2.3 **Role of transcription factors**

In addition, the rate of transcription initiation can be further regulated by a set of proteins referred to by bacteriologists as transcription factors (TFs). In general TFs, bind to specific

sites in the vicinity of promoters and make contacts with RNA polymerase, thereby altering its ability to initiate transcription. Activators promote the binding of RNA polymerase and, as a result, the initiation of transcription. On the other hand, repressors prevent the binding or progression of RNA polymerase, thus inhibiting transcription. Overall, in *E. coli* there are over 270 genes that have been suggested to code TFs (Madan Babu and Teichmann, 2003). Only about half have been characterized experimentally, and of these TFs the consensus sequences (or knowledge of the contribution of individual nucleotide positions to binding) of only 172 are currently logged in RegulonDB (Salgado et al., 2006).

1.3 RNA degradation

1.3.1 **Introduction**

Another level of gene regulation occurs when RNA is processed and degraded, which is the main area of research in the McDowall laboratory and of central importance for the development of the project here presented. The degradation of RNA provides the counterbalance to transcription and enables the translational machinery to follow closely programs of transcription. Furthermore, there is a close interplay between translation and mRNA decay as the translational machinery is known to protect mRNA from decay (Iost and Dreyfus, 1995; Joyce and Dreyfus, 1998; Dreyfus, 2009). This means that the effects of translation will affect mRNA decay and register at the RNA level.

In *E. coli,* which is the major model system for studying post-transcriptional control in bacteria, RNA is degraded by the combined action of endonucleases and 3' to 5' exonucleases, which function in concert with helicases, poly(A) polymerase, an RNA 5' pyrophosphohydrolase (or 'decapping' enzyme) and protein regulators (Carpousis et al., 2009). Endonucleolytic cleavage is thought to accelerate the degradation of *E. coli* mRNA (or at least segments) by allowing 3' exonuclease access either prior to the termination of transcription or transcriptional terminator structures at the 3'. The generally accepted model for RNA degradation is presented in Figure 1.1.

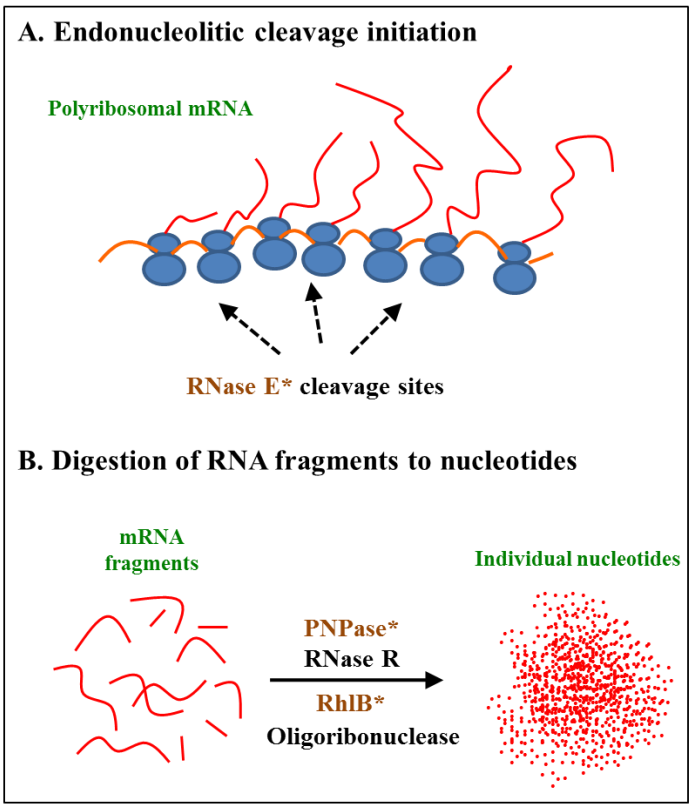

Figure 1.1 - Endonucleolytic RNA degradation model. (A) Polyribosomal RNA is cleaved at AU rich areas by RNase E (Mcdowall et al., 1994; Mcdowall and Cohen, 1996)**.** (B) The resulting RNA fragments are digested into single nucleotides by the combined action of exonucleolytic enzymes such as PNPase, RNase R and helicase RhlB (Carpousis et al., 2009). Oligoribonuclease is a 3' to 5' processive RNase and degrades partially degraded RNAs to monoribonucleotides. The RNA degradosome, a multi-component complex in *E. coli*, is composed of RNase E, PNPase, RhlB and enolase (Taghbalout and Rothfield, 2008; Carpousis et al., 2009); enzymes that form part of the degradosome are highlighted in brown and marked with an asterisk. Taken from (Carpousis et al., 2009).

1.3.2 **Ribonucleases**

As mentioned above, RNA in *E. coli* is degraded and matured by the concerted action of endoribonucleases and 3' to 5' exoribonucleases. In *E. coli*, there is no evidence for a 5' to 3' exoribonuclease, although one does exist in *B. subtilis* and other bacteria (Even et al., 2005; Madhugiri and Evguenieva-Hackenberg, 2009; Bugrysheva and Scott, 2010). Endoribonucleases cleave within the RNA strand, similar to the cleavage of DNA by restriction enzymes, but without the same stringent sequence specificity. Much work has focused on RNase E as its inactivation leads to the stabilisation of most transcripts in *E. coli*, as well as blocking processing of tRNA (Ow and Kushner, 2002) and rRNA (Apirion and Lassar, 1978; Ono and Kuwano, 1979). Other important endoribonucleases are RNase III, RNase P, RNase G and MazF. On the other hand, exoribonucleases digest the RNA strand from the 3' to the 5' end of the transcript and remove one nucleotide at a time. The main

representatives of this group with regard to the degradation of mRNA are RNase II, RNase R and PNPase. Exonuclease RNase Z is also worth mentioning, as it is responsible for the processing of the 3' end of tRNAs. For a full list of exo- and endo-ribonucleases, along with their role in RNA processing and degradation, refer to Table 1.2. Exoribonucleases can either be distributive (will dissociate from the RNA strand after each individual nucleotide is removed) or processive (remain associated with the substrate until the phosphodiester bond upstream of the last removable nucleotide has been cleaved). PNPase is a phosphorolytic enzyme that produces nucleotide diphosphates, while RNase II and RNase R are hydrolytic and produce nucleotide monophosphates (Carpousis et al., 2009). The fact that PNPase activity generates nucleotide diphosphates and RNase II and RNase R generate nucleotide monophosphates is important for energy conservation and the generation of ATP. Ribonuclease activity is not the only function that aids in the degradation of RNA; indeed, polyadenylation, catalysed by poly(A) polymerase, causes the 3' terminus of RNA to become more susceptible to 3' exonucleolytic attack with as little as 5 adenosine residues added (Blum et al., 1999). Furthermore, PNPase is also known to function in reverse in order to add heterooligomeric tails which could promote 3' to 5' exonucleolytic attack (Mohanty and Kushner, 2000).

Enzyme	Exo or Endo-ribonuclease	Role in RNA metabolism	Sigma factor*	Reference
PNPase	Exo	3’ tRNA processing and mRNA decay	70	(Soreq and Littauer, 1977; Littauer and Soreq, 1982)
RNase R	Exo	rRNA maturation, mRNA degradation	ND	(Cheng et al., 1998)
RNase PH	Exo	3' tRNA processing	ND	(Deutscher et al., 1988; Mian, 1997; Zhou and Deutscher, 1997)
RNase II	Exo	3' tRNA processing	70	(Gupta et al., 1977)
RNase D	Exo	3' tRNA processing	70	(Cudny et al., 1981; Mian, 1997)
RNase T	Exo	3' tRNA processing and 5S rRNA 3' maturation	24	(Deutscher and Marlor, 1985)
Oligo-RNase	Exo	Processive 3' to 5' RNase	70	(Niyogi and Datta, 1975; Zhang et al., 1998)
RNase III	Endo	5S, 16S and 23S rRNA precursor formation	70	(Robertson et al., 1967; Conrad et al., 2002)
RNase P	Endo	5' tRNA processing	70	(Pace and Brown, 1995)
RNase Z	Endo	tRNA processing	70	(Callahan et al., 2000; Ezraty et al., 2005)
MazF	Endo	mRNA processing and degradation	70	(Christensen et al., 2003; Zhang et al., 2003; Munoz-Gomez et al., 2004)
RNase E	Endo	5S, 16 S and 23S rRNA maturation and mRNA decay	70	(Ghora and Apirion, 1978; Ono and Kuwano, 1979)
RNase G	Endo	16S rRNA 5' maturation and mRNA decay	N/D	(Mcdowall et al., 1993; Wachi et al., 1999; Tock et al., 2000)
RNase I	Endo	23S RNA and total mRNA degradation	N/D	(Cannistraro and Kennell, 1989; Meador and Kennell, 1990; Cannistraro and Kennell, 1991; Subbarayan and Deutscher, 2001)
RNase LS	Endo	Bacteriophage T4 and cellular mRNA degradation	70/32	(Otsuka et al., 2003; Otsuka and Yonesaki, 2005)
RNase HI	Endo	RNA degradation in RNA-DNA hybrids	24	(Mian, 1997)
RNase HII	Endo	RNA degradation in RNA-DNA hybrids	24	(Mian, 1997)

Table 1.2 - List of exo and endo-ribonucleases in *E. coli.* This table presents all of the known *E. coli* ribonucleases along with their role in RNA metabolism and the sigma factors that regulate their transcription. Note that σ70 is responsible for mediating the transcription of most ribonucleases. Adapted from (Condon, 2007) and (Gutgsell and Jain, 2010). *Data from RegulonDB if present. RNases which are part of toxin-antitoxin systems are not included with the exception of MazF (N/D= not determined).

1.3.3 **RNase E**

In *E. coli,* the rapid degradation of mRNA is known to be highly dependent on RNase E. This endoribonuclease is required for the normal, rapid degradation of many, if not most transcripts (for recent reviews, see (Deutscher, 2006; Carpousis et al., 2009)) , including RNAI, the antisense RNA regulator of ColEI-type plasmid replication (Tomcsanyi and Apirion, 1985; Lin-Chao and Cohen, 1991). It also has a role in the processing of precursors of ribosomal RNA (Ghora and Apirion, 1978; Misra and Apirion, 1979; Li et al., 1999b) and transfer RNA (Li and Deutscher, 2002; Ow and Kushner, 2002) as well as several other small, non-protein-coding RNAs (Lundberg and Altman, 1995; Lin-Chao et al., 1999). Concordant with its central role in RNA processing and degradation, RNase E is essential for *E. coli* viability (Apirion and Lassar, 1978; Ono and Kuwano, 1979). It is known that the inactivation of *rne* leads to the accumulation of a large rRNA precursor (unprocessed 17 S rRNA) (Li et al., 1999a)and the 9S precursor of 5S rRNA (Ghora and Apirion, 1978). RNase E is assisted in the processing of 16S rRNA by a paralogue called RNase G (Li et al., 1999b; Wachi et al., 1999); which is also required for the normal degradation of several functional transcripts (Wachi, 2001; Lee et al., 2002; Jourdan and Mcdowall, 2008).

1.3.3.1 Regulation and role in RNA degradation

RNase E plays a central role in the widely accepted RNA degradation model (Figure 1.1). Despite this model being widely acknowledged, inactivation of *rne* in *E. coli* does not abolish mRNA degradation but only slows it down (Ono and Kuwano, 1979; Lee et al., 2002). This suggests that, despite being essential for cell viability, the role of RNase E in mRNA processing and degradation might be fulfilled by other enzymes; for example, its paralogue RNase G. RNase E shares with RNase G a high N-terminal catalytic domain sequence similarity and the ability to process 16S rRNA, as well as high specificity for 5'-monophosphorylated RNA, due to both enzymes having a 5'-monophosphate sensing pocket (Mcdowall and Cohen, 1996). Furthermore, the activity of RNase E is tightly regulated in the cell by a variety of mechanisms. RNase E can auto regulate the level of its own transcript by controlling the decay of its mRNA and adjusts its synthesis according to the levels of its substrates (Mudd and Higgins, 1993; Jain and Belasco, 1995; Sousa et al., 2001) In addition, RNase E can be inhibited by direct binding of the inhibitor proteins RraA and RraB (Gao et al., 2006; Gorna et al., 2010). Antisense RNAs are capable of blocking or enhancing translation and mRNA degradation. This means, RNase E activity can be affected by sRNAs, which can act passively or actively on specific mRNA targets. In the first case,

translation is silenced or sites are made available for RNase E cleavage by the direct binding of sRNAs to the mRNA (Lease et al., 1998; Morita et al., 2006; Darfeuille et al., 2007; Maki et al., 2008; Pfeiffer et al., 2009). In the second case, sRNAs generate complexes with the chaperone Hfq in order to recruit specific mRNAs to RNase E and thus stimulate their degradation (Massé et al., 2003; Morita et al., 2005; Pfeiffer et al., 2009; Vogel and Luisi, 2011). An example of the latter is the MicC-Hfq RNase E mediated degradation of *ompD* (Pfeiffer et al., 2009). Other components of the degradosome are also regulated, for example PNPase activity is known to be modulated by citrate (Nurmohamed et al., 2011).

Models that do not depend on endonucleolytic cleavage of the RNA strand by enzymes such as RNase E as the initial step of degradation should be further explored. Indeed, the existence of RppH, a 5' pyrophosphatase, suggests that there might be a complementary pathway for RNA degradation in which RppH converts triphosphate groups at the 5' end of nascent transcripts (which protect RNA from degradation), into 5'-monophosphate groups. This would render the RNA susceptible to 5'-monophosphate dependent ribonucleases, including RNase E (Deana et al., 2008).

1.3.3.2 Structure of RNase E

RNase E is an enzyme of two halves (Figure 1.2). The N-terminal half (NTH) provides the endonucleolytic activity (Mcdowall and Cohen, 1996; Callaghan et al., 2005; Caruthers et al., 2006), while the C- terminal half (CTH) provides contacts that allow the formation (Vanzo et al., 1998; Callaghan et al., 2004; Carpousis et al., 2009) and cellular localisation (Khemici et al., 2008) of the degradosome complex. RNase G is related in sequence to only the NTH of RNase E (Mcdowall et al., 1993). The NTH of RNase E forms a tetramer (Callaghan et al., 2003), which is a dimer of the dimeric unit that forms the active sites (Callaghan et al., 2005). The CTH of RNase E contains ancillary RNA-binding sites (Taraseviciene et al., 1995; Mcdowall and Cohen, 1996; Kaberdin et al., 2000) and segments that facilitate interaction with the inner membrane (Miczak et al., 1991; Liou et al., 2001; Khemici et al., 2008) and with RhlB, enolase and PNPase which come together to form an RNA processing and degradation complex called the RNA degradosome (Figure 1.3) (for reviews, see (Carpousis et al., 2001; Carpousis, 2002; Marcaida et al., 2006; Carpousis et al., 2009). *E. coli* strains that have the CTH of RNase E have a competitive growth advantage over those where this region has been deleted. The CTH also aids in the degradation of highly structured RNAs, which accumulate in CTH deletion strains.(Mackie, 2013b). Furthermore, the CTH of RNase E plays a vital role in processing and degradation *in vivo* as

strains that contain RNase E where the 5'-monophosphate sensing pocket has been inactivated or deletion of RppH together with truncation of the CTH are not viable (Anupama et al., 2011). Nevertheless, only the NTH of *E. coli* RNase E is required for cell viability (Apirion and Lassar, 1978; Ono and Kuwano, 1979).

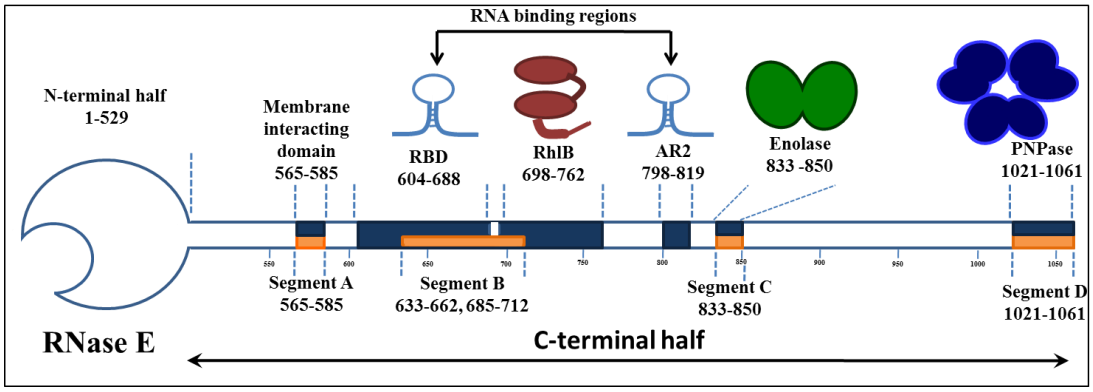

Figure 1.2 – Schematic representation of the primary structure of RNase E and components of the degradosome. The catalytic region of RNase E (NTH) spans residues 1-529. Residues 530-1061 correspond to the C- terminal half of the enzyme, which has several functional segments that enable RNase E to engage in protein-protein interactions (Carpousis et al., 2009) including with RhlB (residues 698-762). The segments labelled as A, B, C and D are regions that are known to: i) localize the degradosome to the cellular membrane, ii) form a coil-coil motif and an arginine rich RNA binding domain (RBD) iii) bind enolase and iv) bind PNPase, respectively.(Carpousis et al., 2009). Regions of the CTH involved in intermolecular interactions are labelled dark blue whilst highly structured regions are presented in orange. Adapted from (Carpousis et al., 2009).

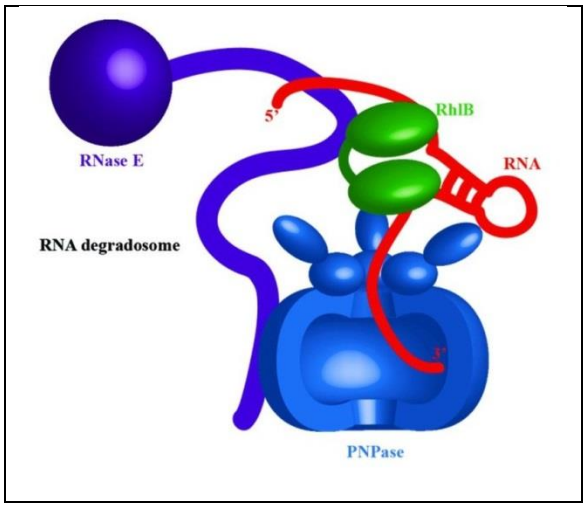

Figure 1.3 – Schematic representation of the RNA degradosome in *E. coli.* This figure represents the degradosome; PNPase (blue) has been cross-sectioned to illustrate the active site. It can be seen that the RNA is presented to PNPase in an unwound conformation as RhlB (green) unwinds the RNA's stem loop (red) which would otherwise provide the RNA with ribonucleolytic protection. Both PNPase and RhlB are associated with RNase E (dark blue). Enolase and other components of the degradosome were removed for simplicity. Adapted from (Bandyra and Luisi, 2013).

1.3.3.3 5’ monophosphate sensing by RNase E

The ability of RNase E to sense RNA 5’ end structures *in vivo* (Bouvet and Belasco, 1992) results in a repertoire of transcripts incorporating secondary structures to their 5’ end in order to grant them increased stability. Indeed, two mechanisms, which provide RNA transcripts with increased protection against RNA degradation, have been identified. Firstly, the concealment of the 5’ ends by either adding a 5’ stem-loop or circularizing the RNA. The presence of a 5’ stem-loop has been shown to give *ompA* mRNA (Emory et al., 1992) great stability. This transcript has a half-life of 15-20 min and has been shown to be amongst the most stable mRNAs in *E. coli* (Von Gabain et al., 1983); this feature is also shared with RNAI (Bouvet and Belasco, 1992). RNA circularization has also been shown to stabilize *rpsT* mRNA (Mackie, 2000; Baker and Mackie, 2003). Secondly, the 5’-triphosphate group or ‘cap’ incorporated during transcription has also been shown to stabilize transcripts. Initial observations showed that RNase E cleavage rates for *rpsT* mRNA and the 9S precursor of 5S rRNA *in vitro* were significantly greater when the 5’ end possessed a 5’-monophosphate group in contrast to when they had a 5’-triphosphate group (Mackie, 1998).

Following the studies described above, the X-ray crystal structure of the N-terminal half of RNase E complexed with oligonucleotide substrates lacking any 5’ end protective features (corresponding to the segment cleaved at the 5’ end of RNAI) was solved. This structure revealed a binding pocket for 5’ ends that terminate in a 5’-monophosphate group and have at least 3-5 unpaired nucleotides (Figure 1.4). Transcripts which have a 5’-monophosphate group are contacted by a semi-circular ring of hydrogen bonding donors created by the side chain and peptide amide of Thr170 and the guanidino group of Arg169; the terminal base is contacted via a hydrophobic interaction with the side chain of Val128 (Callaghan et al., 2005). The interaction of Arg169 is consolidated by a hydrogen bond to the peptide backbone of Gly124 in the neighbouring strand of the sensor. As there is insufficient room in the pocket to accommodate a 5’ end that has a triphosphate group or is base-paired, the pocket can only bind a 5’-monophosphorylated, single stranded end. Residues in this pocket have been mutated in both *E. coli* RNase E and RNase G (Jourdan et al., 2009; Kime et al., 2010). The RNase E mutant, T170V, shows a notable reduction in its ability to cleave 5’-monophosphorylated oligonucleotide substrates, but still retains its basal activity for the cleavage of 5’-hydroxylated oligonucleotide substrates (Kime et al., 2010).

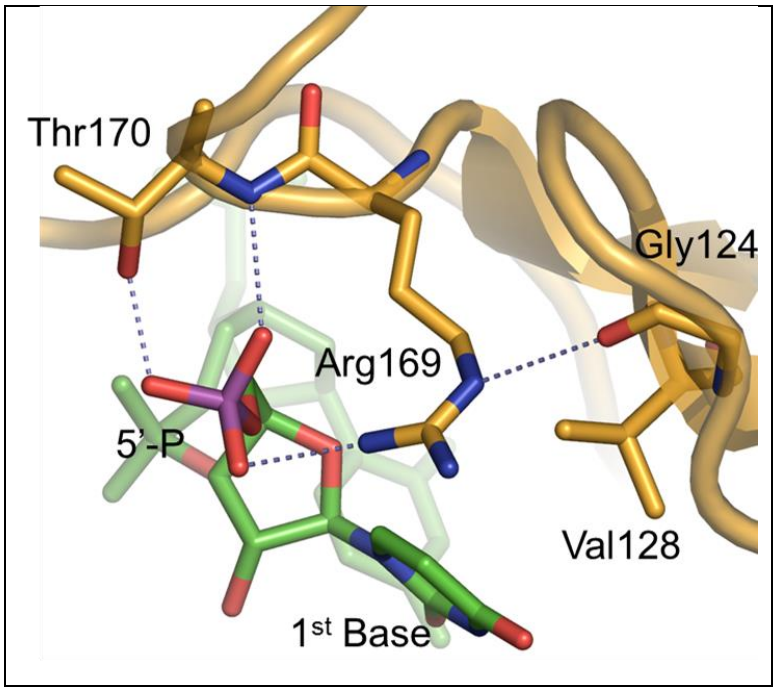

Figure 1.4 - The binding pocket for 5'-monophosphorylated ends in RNase E. This figure shows the interactions that take place between the 5'-monophosphate group in the RNA and RNase E. The monophosphate group is held in place by three hydrogen bonds (represented as dotted lines). An additional hydrogen bond is formed between R169 and G124 within RNase E in order to support the interaction between R169 and the 5'-monophosphate group. Another important interaction that stabilizes this complex is that of V128 and the aromatic ring in the RNA terminal base; an interaction that is modified in the T170V mutant, making it insensitive to the presence of 5'-monophosphorylated ends. Adapted from (Callaghan et al., 2005).

1.3.3.4 Stimulation of RNase E activity via RppH mediated 'decapping'

Structural studies aided in the elucidation of the preference of RNase E for 5'-monophosphorylated substrates (Tock et al., 2000; Callaghan et al., 2005). Nevertheless, the mechanism by which 5' stem-loops provided protection against RNase E-mediated degradation *in vivo* (Bouvet and Belasco, 1992; Baker and Mackie, 2003) was still unclear as primary transcripts are synthesised with a 5' triphosphate group. This was recently elucidated with the discovery, in *E. coli,* of an RNA pyrophosphohydrolase (RppH). RppH converts the primary 5' triphosphate group to a 5'-monophosphate group (Celesnik et al., 2007; Deana et al., 2008).

Upon inactivation of RppH *in vivo*, the half-lives of a range of mRNAs were studied and were found to increase considerably; up to fourfold for *rpsT*. This observation confirmed the importance of RppH in the generation of 5'-monophosphate single stranded ends under

physiological conditions (Deana et al., 2008). Nevertheless, when assessed on a genome-wide scale, the majority of transcripts in *E. coli* seemed to be largely unaffected by inactivation of RppH (Deana et al., 2008). Thus 5' 'decapping' does not appear to be essential for rapid mRNA degradation. This observation is possibly due to RNase E not requiring an interaction with 5'-monophosphate ends to in order to make initial cuts in the RNA (see section 1.3.3.5).

1.3.3.5 Direct entry by RNase E

Consistent with the idea that RNase E can cleave independently of the 5' end it has been found recently that the NTH of both wild-type RNase E and the T170V mutant can rapidly cleave *cspA* mRNA *in vitro* without requiring a 5'-monophosphorylated end (Kime et al., 2010). Moreover, using a well-characterised oligonucleotide substrate, a minimum requirement for cleavage was determined in which the presentation of multiple single-stranded segments allows their simultaneous interaction with RNase E and enhanced cleavage (Kime et al., 2010). In addition, RNase E has been shown to process a tRNA precursor *in vitro* by binding single stranded regions that are adjacent but not contiguous to the site that is cleaved (Kime et al., 2014). Single-stranded segments are abundant in RNA, thus direct entry could represent a major mechanism by which mRNA degradation and processing is initiated by RNase E in *E. coli* and other RNase E homologue containing organisms. Furthermore, the above observations suggest that direct entry might be the preferred mechanism of 5'-triphosphorylated RNA cleavage, a mechanism which is enhanced by the ability of RNase E which is a tetramer to bind multiple single stranded regions.

1.3.4 **RNA stability**

RNA can be categorized into two groups based upon its turnover rate; "stable" RNA, such as tRNA and rRNA, which have half-lives in the order of tens of minutes (Curnow et al., 1993) and "unstable" RNA, such as the majority of mRNA. The half-life of total mRNA has been estimated to be 6.8 min, with some mRNAs having a half-life of under 40 sec (Selinger et al., 2003). In the case of the latter, it is likely that translation is limited to the passage of only a single ribosome per transcript. It has been reported that there is some correlation between half-lives and the function encoded by the transcript, e.g. transcripts encoding enzymes of central metabolism tend to have longer half-lives than those encoding products of unknown function (Selinger et al., 2003); these short half-lives make it difficult to

characterize and study such mRNAs. These findings reiterate the fact that mRNA decay is a crucial step in regulating the expression of specific genes and of the need for high throughput methods to better understand the global code required for the degradation of mRNA.

1.4 Previous RNA sequencing approach

Microarrays have been extensively used in order to study the expression of thousands of genes in parallel in a variety of organisms (Bier and Kleinjung, 2001). They have also been implemented to elucidate the response and regulation of the transcriptome to environmental stresses. They have provided a methodology reaching further than that which was achievable by previous technologies such as nuclease mapping (Weaver and Weissmann, 1979) and primer extension (Shelness and Williams, 1985). Despite the density of probes increasing in recent years to give a coverage of up to 3 Mbp (approximately every 10 bp), this approach is still incapable of providing nucleotide resolution maps of primary or secondary transcriptomes (Goldsmith and Dhanasekaran, 2004). At present, the only way to achieve single-nucleotide resolution maps is through global RNA sequencing (gRNA-seq); often referred to as transcriptome shotgun sequencing (Mamanova and Turner, 2011).

Recently, the primary transcriptome of the major human pathogen *Helicobacter pylori* has been sequenced with an RNA-seq approach described in Figure 1.5 (Sharma et al., 2010). This approach was a genome-wide RNA-ligase mediated, reverse transcription PCR. The method depleted the transcriptome of 5'-monophosphorylated RNA by treating total RNA with Terminator™ 5' phosphate-dependent exonuclease (TEX). The remaining 5'-triphosphorylated RNA was treated with a pyrophosphatase in order to generate 5'-monophosphorylated ends and these were then cloned, sequenced and mapped onto the genome to identify transcription start sites. This technology had two major limitations. Firstly, it removed the secondary transcriptome and any information regarding processing; secondly, it used the Roche-454 sequencing technology that provides a lower number of reads than the Illumina technology.

Whilst the work of Sharma et al. (2010) inspired the work here presented, modifications were made in order to adapt and enhance the capability of this approach. These modifications were made in order to increase the sequencing depth and secondly, detect and map the secondary transcriptome of *E. coli* (Figure 1.5).

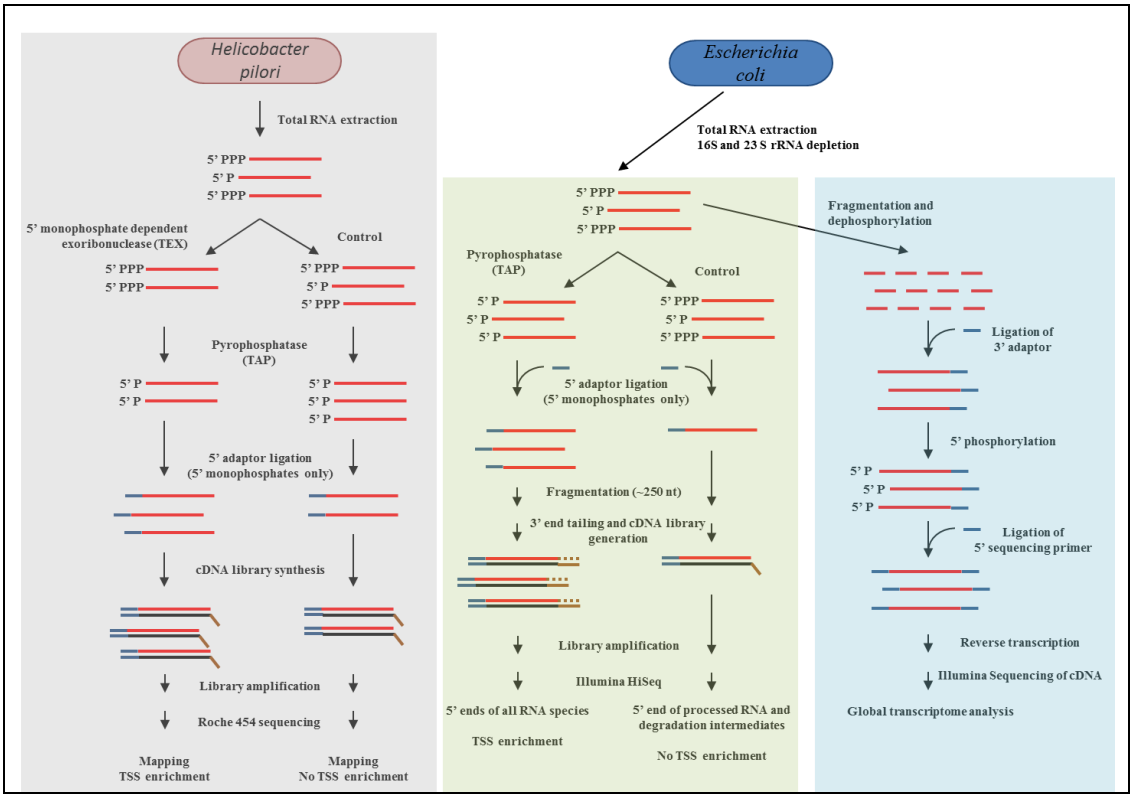

Figure 1.5 - The sequencing approach of Sharma et al. (2010) and our sequencing approach. The sequencing approach undertaken by Sharma et al. (2012) is highlighted grey. Total RNA was extracted from *H. pylori* and separated into two aliquots; a control and a sample which was depleted of 5'-monophosphate ends (processing sites) with TEX. The 5'-triphosphorylated RNA in both samples was then treated with a pyrophosphatase in order to generate 5'-monophosphorylated ends and then ligated to a 5' adaptor. This was followed by the generation of cDNA libraries and sequencing, performed using a Roche 454 platform (Sharma et al., 2010). The sample treated with TEX was enriched for TSSs and the control sample contained both processing sites and TSSs. Our sequencing approach made use of dRNA-seq (highlighted green) and gRNA-seq (highlighted blue). The dRNA-seq was performed in a similar way to Sharma et al. (2010) with three main differences. Firstly TAP was used to treat the non-control sample, instead of TEX. This enriched TSSs and, unlike TEX, did not remove products of processing and degradation. Secondly, RNA was fragmented prior to 3' end adaptor ligation in order to allow the sequencing of both long and short transcripts. The third difference was that Illumina sequencing was used instead of Roche 454 sequencing, as it is more adequate for the sequencing of shorter transcripts. In addition to these differences, our approach also included gRNA-seq (Mamanova et al., 2010) in which we fragmented and sequenced every transcript present in the cells.

1.5 Broad objective and specific aims

The broad objective of this thesis was to develop approaches that would provide, for the first time, a transcriptome-wide view of steps in RNA processing and decay, with focus on the role of direct-entry cleavage by RNase E. The specific aims were to (i) confirm that treatment with TAP could be used to provide efficient discrimination of the 5' phosphorylation status of transcripts (Chapter 3), (ii) confirm that the T170V substitution of RNase E reduces the efficiency of 5'-monophosphate-dependent cleavage, without affecting the efficiency of 5'-monophosphate-independent cleavage (Chapter 3), (iii) prepare samples of total RNA digested with T170V and perform RNA–sequencing to identify sites of cleavage *in vitro* (Chapter 3 and Chapter 4), (iv) in parallel with (iii), analyse the transcriptional landscape of *E. coli* using FRT RNA-seq, an improved global approach, in conjunction with a differential RNA-seq approach that retains products and intermediates of RNA processing and decay (Chapter 4), (v) identify sites of direct entry cleavage that occur *in vivo* as well as *in vitro* (Chapter 4), and (vi) validate experimentally any novel features identified within the transcriptional landscape of *E. coli* (Chapter 5). Overall, the specific aims were met, but further experiments by others were required to confirm that cleavages produced by T170V were indeed independent of 5'-monophosphate sensing.

Chapter 2

2 Materials and Methods

2.1 Bacterial strains and media

Strains and plasmids used are listed in Table 2.1. *E. coli* cells were grown at various temperatures, which are indicated as appropriate. Cells were grown in Luria Bertani medium: 1% (w/v) tryptone, 0.5% (w/v) NaCl, and 0.5% (w/v) yeast extract (Sigma Aldrich). When needed, selective growth conditions were achieved by adding carbenicillin or kanamycin to a final concentration of 50 µg/ml. All media were autoclaved for 20 min at 121 psi prior to use. Heat-sensitive components such as antibiotics were sterilized by filtration through 0.2 µm disposable filters (Sarstedt GmbH) and added after heat sterilization of the other components. Plates were prepared by adding 1.5% (w/v) of agar (Sigma Aldrich) to liquid medium prior to autoclaving. Aeration of 50 ml liquid cultures was achieved by continuous shaking at 220 rpm in 250 ml Erlenmeyer flasks. Strains were stored at –80°C as glycerol stocks: 50% (v/v) glycerol and 50% (v/v) overnight culture.

Strains	Description	Source
BW25113	F^-, *Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), λ^-, *rph-1*, Δ*(rhaD-rhaB)568*, *hsdR514*	Keio collection (Baba et al., 2006)
JW2753	As BW25113, but *ΔchpA781::kan*	Keio collection (Baba et al., 2006)
N3433	Hfr(PO1), *lacZ43* (FS), λ^-, *relA1*, *spoT1*, *thiE1*	(Browning and Busby, 2004)
N3431	As N3433, but *rne-3071*(ts)	(Browning and Busby, 2004)
JM109	*end*A1, *rec*A1, *gyr*A96, *thi, hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F′ *tra*D36, *pro*AB, $laqI^qZ\Delta M15$]	Promega (Sharma and Chatterji, 2010)

Table 2.1 - Strains. This table lists all strains used in this thesis alongside a description and indication of their source.

2.2 Gel electrophoresis

2.2.1 **Agarose gel electrophoresis**

DNA was separated according to its size using agarose gel electrophoresis. Protocols were based on (Sambrook, 2001). The agarose concentration (w/v) varied depending on the size range of the fragments that were analysed and is stated within the corresponding figure legends. Agarose (Melford) was dissolved in 1 x TBE: 89 mM Tris-HCl (pH 8.3), 89 mM boric acid, and 2 mM EDTA. Samples were loaded in 1 x DNA loading dye: 10 mM EDTA (pH 8), 10% (v/v) glycerol, 0.06% (w/v) Bromophenol Blue (Ambion). DNA markers (Fermentas) were used as size standards. Electrophoresis was performed at $10Vcm^{-1}$ of gel for 60 min. Gels were stained by soaking for 30 min in 1 x TBE containing 1 µg/ml ethidium bromide and imaged with the use of a GeneGenius UV transilluminator (Syngene) set at an excitation wavelength of 302 nm.

2.2.2 **Polyacrylamide gel electrophoresis**

Polyacrylamide gel electrophoresis was used to separate both RNA and DNA samples by size. In order to separate RNA, denaturing polyacrylamide gels composed of 19:1 polyacrylamide: *bis*-acrylamide in 7 M urea and 1 x TBE were used. The concentrations of acrylamide are stated in the corresponding figure legends. DNA was similarly separated, except urea was omitted from the gels. For RNA gels, markers and 2 x denaturing loading dye from Ambion were used, while for DNA gels, the marker and loading dyes were those used for agarose gel electrophoresis. Gels were typically run for ~2200 Vh.

2.2.3 **SDS gel electrophoresis**

SDS-PAGE was performed as previously described (Laemmli, 1970). Gels contained a stacking and a running layer of 4 % (w/v) and 18 % (w/v) acrylamide: *bis*-acrylamide (29:1, Severn Biotech), respectively. Gels were run in 1 x electrophoresis buffer: 0.1 % SDS (w/v), 25 mM Tris base, 192 mM glycine (Sigma Aldrich) at 25 mA for 1-2 h. Gels were stained by soaking in Coomassie stain: Coomassie Brilliant Blue R-250 (150 mg l^{-1}) in 25 % (v/v) ethanol, 10 % (v/v) acetic acid (Sigma Aldrich). Gels were destained by soaking in destain solution: 40 % (v/v) methanol, 10 % (v/v) acetic acid (Sigma Aldrich). Imaging was performed in a GeneGenius UV transilluminator (Syngene) with lower white light settings. Samples were denatured in 2 x Laemmli buffer: (4% SDS, 20% glycerol, 10% 2-

mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris HCl, pH approx. 6.8) (Sigma Aldrich) for 5 min at 99 °C prior to loading.

2.3 Radio-labelled probe generation

Specific transcripts were probed using complementary oligonucleotides (see Table 2.2) labelled at their 5' ends with ^{32}P using T4 polynucleotide kinase (Thermo Scientific) and γ-^{32}P-ATP (3000 Ci mmol^{-1} 10 mCi mL^{-1}, 250 µCi, Perkin Elmer). The labelling reaction was carried at 37°C for 30 min and stopped by the addition of EDTA, both as described by the vendor of the enzyme (Thermo Scientific). The radioactively labelled probes were ethanol precipitated, and resuspended in 20 µl of RNase-free water.

2.4 DNA methods

2.4.1 **DNA purification and restriction digest**

Genomic DNA used as a template for PCR was obtained as a by-product of the RNA isolation procedure (see Section 2.5.1). Following the addition of ethanol to precipitate the nucleic acid, the chromosomal DNA, which formed a stringy mass, was removed using the end of a sterile tip, washed with 70% (v/v) ethanol, pelleted by brief centrifugation (16,000 x *g* for 30 s), and resuspended in TE buffer (pH 8.0) (Sigma). Restriction enzymes, used on PCR products, were obtained from New England BioLabs. DNA restrictions were carried out as per the supplier's protocol.

2.4.2 **DNA quantification**

DNA samples with a relatively high concentration (> 5 µg/ml) were quantitated by spectrophotometry by measuring the absorbance at 260 nm with a Geneflow Perl nanophotometer. At this wavelength, an absorbance of 1 corresponds to approximately 50 µg/ml of double-stranded DNA. For a low DNA concentration (< 5 µg/ml), the amount was estimated by running a sample on an agarose gel and comparing the intensity of the ethidium-bromide stained bands with those of marker bands of similar size and known concentration.

2.4.3 **DNA gel extraction**

Following electrophoresis (see Section 2.2.1) DNA fragments were extracted from slabs excised from agarose gels using the QIAquick® Gel Extraction Kit, as per the manufacturer's instructions (Qiagen).

2.4.4 **Polymerase chain reaction (PCR)**

PCRs were carried out as described in Sambrook *et al.*, (1989). In general, annealing temperatures 5°C below the estimated T_m of the primers were used. An extension time of 1 min was allowed for every 1 kb to be amplified. Cycles were repeated 35 times for all reactions unless otherwise stated. When a downstream reaction was to follow PCR, products were purified using the PCR purification kit (Quiagen). All primers used are presented in Table 2.2.

Table 2.2 - Oligonucleotides used in this study.

Name	Sequence (5' to 3')
Used for RLM-RT-PCR	
16S	GATCCAACCGCAGGTTCC
ftsT	GGCTTCTCAACAGGTGGTGT
pgpA	TGCCACGGATTACTCATCTTC
rhlB	TGGGCGTACAGTTATGAAACC
RLM1	CGAAGACAACAAAGAAGTTCAACTC
RLM2	CATGAGGATTACCCATGTCG
ymfK	GGCCTGTACCCATGATATGAC
Used for PCR and *in vitro* transcription	
5SF	ATCCTAATACGACTCACTATAGGGTGGCGGATTGAGAGAAGATT
5SR	ATGCCTGGCAGTTCCCTACT
9S F	ATCCTAATACGACTCACTATAGGGAAGCTGTTTTGGCGGATGAG
9S R	ACGAAAGGCCCAGTCTTT
argXF	CGCGTAATACGACTCACTATAGGGAGGGGTGGGAAGTCCGTATTA
argXR	AGGGTGACGAAATGCACAGAA
cspA F	ATCCTAATACGACTCACTATAGGGTTTGACGTACAGACC
cspA R2	AAAATCCCCGCCAAATGGCAGGG
dnaKF	ATCCTAATACGACTCACTATAGGGAAGACAAAGCCCTATCGAA
dnaKR	CTTTGTCACCCTGGTTACGG
fumAF	ATCCTAATACGACTCACTATAGGGCCGGACGGATGGATTCTTAT
fumAR	AAAAAAACCGCCCCGAAG
glyWF	TGCGTAATACGACTCACTATAGGGACTCATCGCGCCAGGTAAGTA
glyWR	AGGTGGTTTCACGACACTGCT

Table 2.2 (continued)

manXF	ATCCTAATACGACTCACTATAGGGCCTCTGTTAACGTCGGTGGT
manXR	ACCAGGGTACACGCGATTAG
pheUF	ATATTAATACGACTCACTATAGGGCGAGATGTGCAGATTACGGTTT
pheUR	GCACGACATTTCACGTCAGTT
RNAI F	ATCCTAATACGACTCACTATAGGGACAGTATTTG
RNAI R	AACAAAAAAACCACCGCTACC
Used to generate ^{32}P labelled oligoprobes	
ecr0174(u+)	CCACGGCATATCTGACCTTATAAAGCCAAC
ecr3777(d+)	ACCTTTTCGGCTGTCTCTTCTCTCGTACTG
ecr2775(u+)	CTCAAGGGGAGAAAACTTAGGGCCTCTATG
ecr4051(u+)	CTGTATGTAGGGTACAGCACGATGAATCTG
ecr1743(d+)	GTATTACCGTAGTAATGCAAGCGCGTCTCAG
AgrB	ACTTTCCAGCCCTGAGTTGGTGGCTCTG

Table 2.2 - Oligonucleotides used in this study. This table presents the name, orientation and 5' to 3' sequence of all primers and oligonucleotides used in this study and groups them based on what they were used for.

2.4.5 **Cloning and sequencing of fragments produced by PCR**

PCR products purified as described above (Section 2.4.4) were ligated into the pGEM-T® Easy vector as per the vendor's instructions (Promega) and then the ligation mix was introduced by transformation into JM109 cells purchased in a competent state (Promega). The transformation mixture was spread on the surface of selective Luria Bertani agar (Sigma) that contained 50 µg/ml carbenicillin. Following transformations, colonies were picked and grown overnight in 10 ml Luria Bertani media containing 50 µg/ml carbenicillin. Plasmid DNA was extracted from these overnight cultures using the Wizard® Mini-prep DNA Purification system as per manufacturer's instructions (Promega). Inserts of purified plasmids were sequenced by GATC-biotech (Germany) using universal primers SP6 and T7.

2.5 RNA methods

2.5.1 **Total RNA extraction**

RNA was isolated from various *E. coli* strains (Section 2.1) following a protocol outlined previously (Kime et al., 2008a). Briefly, *E. coli* culture pellets corresponding to 5-8 OD_{600} units were resuspended in 1 ml resuspension buffer; 10 mM Tris-HCl (pH 8.0), 1 mM EDTA. Cells were then lysed by adding 1 ml lysis buffer; 20 mM Tris-HCl (pH 8.0), 40 mM EDTA, 300 mM NaCl, and 0.5% (w/v) SDS, which was preheated to 100°C. This mix was then

incubated in a boiling water bath for 30 s or until it turned translucent in appearance. After cooling in ice, an equal volume of acid phenol saturated with 100 mM citrate buffer (pH 4.3) was added. This mix was briefly centrifuged (14,000 x *g* for 3 min) and the aqueous phase was retained. Total nucleic acid in the aqueous phase was precipitated by adding 2.5 volumes of absolute ethanol. Precipitation was achieved by centrifugation at 7,000 x *g* for 15 min at 4 °C. The pellet was thoroughly resuspended in 200 µl molecular biology-grade TE buffer (pH 8.0) (Sigma) containing 150 mM NaCl and re-precipitated by adding 500 µl absolute ethanol. Stringy chromosomal DNA was removed using a sterile tip and the remaining nucleic acid, largely RNA, pelleted by centrifugation at 14,000 x *g* for 20 min at 4 °C. The RNA pellet was then resuspended in 400 µl resuspension buffer containing 150 mM NaCl and extracted further with an equal volume of acidic phenol (pH 4.3), an equal volume of acidic phenol: chloroform: isoamyl alcohol (25:24:1), and chloroform: isoamyl alcohol (49:1). RNA was reprecipitated by adding 2.5 volumes of absolute ethanol, pelleted by centrifugation at 14,000 *g* for 15 min at 4°C and washed twice in 70% (v/v) ethanol; a 5 min centrifugation at 14,000 x *g* at 4 °C was performed after each wash. The resulting pellet was air-dried and resuspended in 100 µl molecular biology-grade water (Sigma).

In order to achieve DNA-free RNA, the harvested RNA was treated with RNase-free DNase I (Promega). 1 U DNase I was used to treat 100 µg of nucleic acid in 400 µl of 40 mM Tris-HCl (pH 8.0), 10 mM $MgSO_4$, and 1 mM $CaCl_2$ for 1 h at 37 °C. RNA was then extracted with an equal volume of acidic phenol: chloroform: isoamyl alcohol (25:24:1), NaCl was added to a final concentration of 150 mM, and the RNA precipitated by adding 2.5 volumes of absolute ethanol. The RNA was harvested by centrifugation at 14,000 x *g* for 15 min at 4 °C. Pellets were washed twice as above and resuspended in either water or TE buffer (pH 8.0) of molecular biology grade (Sigma).

2.5.2 **Total RNA enrichment**

Total RNA was enriched for mRNA by removing 16S and 23S rRNA using the MICROBExpress rRNA removal kit as per the manufacturer's instructions (Ambion). The average yield from 10 µg of total RNA was 3 to 4 µg of enriched RNA. Successful removal of excess 16S and 23S rRNA was assessed by analysing 1 µg samples by gel electrophoresis using 2% (w/v) agarose gels. RNA samples were loaded using 2 x RNA loading dye; 95% (v/v) formamide, 0.025% (w/v) bromophenol blue, 0.025% (w/v), xylene cyanole and 0.025% (w/v) SDS (Ambion).

2.5.3 **Northern blotting**

RNA for northern blotting was isolated as described previously (see Section 2.5.1), with the exception that the mRNA was not enriched, from a second batch of cultures. *E. coli* K-12 MG1655 (seq) was grown in Luria-Bertani (Amresco) as well as M9 minimal media (Sigma) supplemented with glucose (0.4%,w/v) at 37°C with shaking (100 rpm) until an OD_{600} of ~0.5 at which point RNA was isolated as described previously (Kime et al., 2008b). For each sample, an aliquot of 5 µg was mixed with an equal volume of 2 x RNA-loading dye (New England BioLabs), denatured by incubation at 90°C for 90 sec, chilled on ice, and analysed along with other samples by denaturing electrophoreses using a 6% sequencing-type gel (acrylamide: *bis*-acrylamide [29:1], 1 x TBE, 7 M urea). Fractionated RNA was electro-transferred to a Hybond-N^+ membrane (Amersham) using 20 x saline-sodium citrate (SSC) buffer at 11 V for 1 h, and subsequently fixed to the membrane by UV-crosslinking. The membrane was pre-hybridized with 3 ml of ULTRAhyb-Oligo Hybridization Buffer (Ambion) at 42^{o}C for 30 min.

Radiolabelled probe, which had been denatured by incubation at 90°C for 90 sec and chilled on ice, was added to the hybridization tube. Hybridization was carried at 42°C overnight. The membrane was washed twice with 20 ml of preheated washing buffer (5 x SSC containing 0.5% [w/v] SDS) at 49°C for 30 min and exposed to Imaging Screen-K (Bio-Rad). The image was captured by Molecular Imager FX (Bio-Rad), and further processed using Quantity One (Bio-Rad) and GeneSys (Syngene) software.

2.5.4 **Global RNA sequencing**

Global Transcriptome sequencing was performed by Dr. Lira Mamanova (Wellcome Trust, Sanger Institute, Cambridge, UK) using a published methodology (Mamanova et al., 2010). Sequencing was performed using an Illumina Genome Analyzer (HiSeq 2000). RNA sequences were processed in-house using Galaxy (Goecks et al., 2010) and mapped onto the *E. coli* reference genome (NCBI accession number U00096.2) using Bowtie 2.0 (Langmead and Salzberg, 2012) with custom parameters: -l 28 for read, -l 20 for read2, and -y -a --best --strata. For every position in the genome, the number of times it was read irrespective of its position in the individual RNA fragments was determined using a combination of BEDtools (Quinlan and Hall, 2010) and bash scripts (this script was written and run by Vijaya Mahalingam Shanmugiah). These counts were then recorded in the format of a bedGraph (part of the UCSC format options) and visualized using the UCSC

genome browser (Schneider et al., 2006; Chan et al., 2012). The above sequencing data was submitted to the GEO repository (Edgar et al., 2002) and is stored under accession number GSE46232.

2.5.5 **Differential RNA sequencing**

Prior to differential RNA-seq, samples were enriched for mRNA using MICROBExpress-Bacteria beads, as described previously (see Section 2.5.2). Differential RNA-seq was performed by Vertis Biotechnologie AG (Germany) as part of a service that included the construction of cDNA libraries before and after treatment with TAP, sequencing of libraries using an Illumina HiSeq platform (single end, 50-bp read length), and the alignment of sequences to the *E. coli* K12 sub-strain MG1655 genome (NCBI, accession number U00096.2). The 5'-sequencing adaptor was ligated to transcripts prior to fragmentation, thereby allowing the 5' ends of both long and short transcripts to be detected. RNA was fragmented using a Bioruptor® Next Gen UCD-300 ™ sonication system (Diagenode), then tailed at the 3' end using poly(A) polymerase (New England BioLabs), copied into cDNA using M-MLV reverse transcriptase (RNase H minus, AffinityScript, Agilent) and an oligo-dT primer, amplified by PCR and fractionated using gel electrophoresis. Fragments of 250 - 500 bp were selected for Illumina sequencing. Reads were trimmed off the 5' adapter and poly(A) sequences and mapped using the CLC Genomics Workbench and standard settings. For each library, the number of times each genome position was the first nucleotide in sequence reads (*i.e.* associated with a 5' end *in vivo*) was counted using a combination of BEDtools (Quinlan and Hall, 2010) and bash scripts (this script was written and run by Vijaya Mahalingam Shanmugiah). The reads corresponding to minus and plus TAP treatment were then compared using M-A scatterplots. The counts were also recorded in bedGraph format (part of the UCSC format options) to allow visualisation using the UCSC genome browser (Schneider et al., 2006; Chan et al., 2012). This dRNA-seq data was also submitted to the GEO repository (Edgar et al., 2002) and stored under accession number GSE46232.

2.5.6 **RNA synthesis by *in vitro* transcription.**

Transcription reactions were carried out as 40 µl reactions containing 80 U RNaseOUT™ (Invitrogen), 5 mM DTT, 1 U yeast inorganic pyrophosphatase (Sigma), 100-150 nM template, 2.5 mM rNTP mix (GE Healthcare) and 100 U T7 RNA polymerase (Invitrogen) in 1

x T7 RNA polymerase buffer: 0.04 M Tris-HCl (pH 8.0), 8 mM $MgCl_2$, 2 mM spermidine-$(HCl)_3$, 25 mM NaCl.

Reactions were incubated at 37 °C for 3 h. Templates were generated via PCR. Following incubation, 2 U of DNase I (Promega) and 40 U RNAseOUT™ (Invitrogen) were added to the reaction which was incubated at 37 °C for a further 1 h.

Following DNase I treatment, reactions were extracted with an equal volume of phenol: chloroform and then phenol: chloroform: isoamyl alcohol (25:25:1). The RNA was precipitated by adding 0.1 volumes of 3 M sodium acetate (pH 5.5) and 2.5 volumes of 100% (v/v) ethanol. The RNA was then harvested by centrifugation at 16,000 x *g* for 10 min at 4°C. The pellets were washed twice with 500 µl ice-cold 70% (v/v) ethanol. Pellets were then air dried and re-suspended in 40 µl of molecular biology-grade water (Sigma). Samples were mixed with an equal amount of 2 x RNA loading dye (Ambion) and separated on a denaturing polyacrylamide gel (Section 2.2.2). RNA was visualized using a GeneGenius UV transilluminator (Syngene) set at an excitation wavelength of 302 nm after incubating the gels in 1 x TBE buffer containing 1 µg/ml ethidium bromide for 5 min.

2.5.7 **Discontinuous cleavage assays**

The discontinuous cleavage assay used here was adapted from a previously published method (Redko et al., 2003). The oligoribonucleotide substrates used during this thesis were 3' fluorescein-labelled versions of BR13 (5'-GGGACAGUAUUUG-3'; (Mcdowall et al., 1995)). Two versions were used; one with a 5'-monophosphate ($_{P}$BR13-Fl) and the other with a 5'-hydroxyl group ($_{HO}$BR13-Fl). These substrates were synthesized and purified by Dharmacon (USA). Reactions were carried out in 1 x reaction buffer (RB): 25 mM bis-Tris-propane (pH 8. 3), 100 mM NaCl, 15 mM $MgCl_2$, 0.1% (v/v) Triton X-100, and 1 mM DTT. Unless otherwise indicated, reactions with oligonucleotides contained substrate and enzyme at concentrations of 250 and 5 nM, respectively, as well as 0.32 U/µL of RNaseOUT™ (Invitrogen). Before starting the reaction, all components were pre-warmed at 37°C for 15 min. Samples removed for analysis were quenched immediately by mixing with an equal volume of 2 x RNA loading dye: 95% (v/v) formamide, 0.025% (w/v) bromophenol blue, 0.025% (w/v), xylene cyanole, and 0.025% (w/v) SDS (Ambion). All reactions were incubated at 37°C. Samples were separated in denaturing polyacrylamide gels (Section 2.2.2). Reaction products were detected using a Fuji FLA 5000 scanner (Fuji) set to fluorescein detection (excitation wavelength 473 nm, filter 523 ±35 nm). Substrate and

product were quantified using the AIDA Image Analyzer software package (Raytest Isotopenmessgeräte GmbH). Initial rates of reaction were calculated by establishing the slope representing the percentage of product produced over time during the initial, linear phase of the reaction. Rates were then calculated taking into account concentrations of enzyme and substrate in the reaction mix.

Total RNA cleavage assays were performed in the same way as oligoribonucleotide cleavage assays; only substrate and enzyme concentrations were varied to 10 µg and 250 nM, respectively. Furthermore when samples were taken, the reaction was quenched and the RNA extracted by increasing the volume to 400 µl with TE buffer (pH 8.0) (Sigma), adjusting the salt concentration to 150 mM NaCl and immediately extracting the RNA as described in Section 2.5.1.

2.5.8 **Enzymatic modifications of RNA**

RNA was subjected to a repertoire of enzyme treatments in order to modify the 5' and 3' end. Terminator™ 5' phosphate-dependent exonuclease (TEX) treatments were typically performed in 40 µl reactions containing 40 U RNaseOUT™ (Invitrogen) and 1 U TEX (Epicentre Biotechnologies) per 1 µg of RNA in 1 x TEX buffer A (Epicentre Biotechnologies). Reactions were incubated at 30°C for 1.5 h. Tobacco acid pyrophosphatase (TAP) treatments were performed in 40 µl reactions containing 3 U of TAP (Epicentre Biotechnologies) per 1 µg RNA in 1 x TAP buffer: 0.5 M sodium acetate (pH 6.0), 10 mM EDTA, 1% β-mercaptoethanol, and 0.1% Triton® X-100. Reactions were incubated at 37°C for 2.5 h. T4 polynucleotide kinase (PNK) treatments were performed in 30 µl reactions containing 40 U RNaseOUT™ (Invitrogen), 100 mM ATP and 13 U of PNK (New England BioLabs) per 1 µg RNA in 1 x PNK buffer: 70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, and 5 mM dithiothreitol. Reactions were incubated at 37°C for 1.5 h. Poly(A) polymerase treatments were performed in 30 µl reactions containing 40 U RNaseOUT™ (Invitrogen), 1 mM ATP and 4 U of Poly(A) polymerase (New England BioLabs) per 1 µg of RNA in 1 x Poly(A) polymerase buffer: 50 mM Tris-HCl (pH 7.9), 250mM NaCl, and 10mM $MgCl_2$. Reactions were incubated at 37 °C for 10 min.

Following all treatments, volumes were adjusted to 400 µl, NaCl added to 150 mM NaCl and phenol: chlorophorm extracted as in Section 2.5.1. Samples were ethanol precipitated and resuspended in molecular biology grade water (Sigma).

2.5.9 **RNA-ligase mediated, reverse-transcription PCR assay**

This method was performed as described previously (Kime et al., 2008a). Briefly, this method consists of three reactions: an RNA ligation, followed by a reverse transcription, and finally PCR.

For the ligation reaction, 20 pmol of an RNA adapter was mixed with 1.2 µg of total RNA, 5 U of T4 RNA ligase (New England BioLabs), 20 U of RNaseOUT™, 1 mM ATP in 10 µl RNA ligase buffer: 50 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT, and 1 mM ATP. The reaction was then incubated at 37°C for 1 h. The adapter used was a 54-mer, with sequence:
5'-ACAUGAGGAUUACCCAUGUCGAAGACAACAAAGAAGUUCAACUCUUUAUGUAUU-3' which was synthesized by Dharmacon (USA). For the reverse transcription reaction, 50 ng of random hexamers (Amersham Biosciences) were annealed to 600 ng of ligated RNA in a volume of 20 µl molecular biology grade water (Sigma). This mix was heated at 65°C for 5 min and snap chilled on ice and added to a solution containing 250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$, 50 mM DTT, 4 µl of a 10 mM stock of dNTPs, 80 U of RNaseOUT™ (Invitrogen) and molecular biology-grade water (Sigma) to a final volume of 40 µl. Following this, two 19 µl aliquots were processed further. 200 U (1 µl) of MMLV reverse transcriptase RNase H minus (Promega) was added to one, and 1 µl water to the other. These mixes when then incubated at 25°C for 15 min, 42°C for 1 h, and finally 70°C for 15 min. At the end of the reaction, 80 µl of 10 ng/ml yeast tRNA (Ambion) was added. For the PCR reaction, the same conditions as in Section 2.4.4 were followed using RLM-2 (Table 2.2), which is complementary to the 5' adaptor and a gene-specific primer (Table 2.2).

Chapter 3

3 Enzyme characterization and workflow set up

3.1 Introduction

RNA sequencing technologies have improved to a point where it is now possible to sequence entire bacterial genomes and transcriptomes in a single sequencing run (Cho et al., 2009; Sharma et al., 2010; Lin et al., 2013b; Wang et al., 2013; Al Rashdi et al., 2014; Benahmed et al., 2014; Hasman et al., 2014). Recently, by removing the products of processing and the intermediates of mRNA degradation (the 'secondary' transcriptome) using Terminator™ Exonuclease (TEX), a 5'-3' enzyme that is specific for 5'-monophosphorylated ends, it was possible to generate more than enough reads using 454 sequencing (2 to 4.5 million) to map transcriptional start sites (TSSs) on a genome-wide scale (see Section 1.4). This involved a second enzymatic treatment with tobacco acid pyrophosphatase (TAP) to produce a 5'-monophosphate on the nascent ends to permit their cloning and subsequent sequencing. This differential RNA (dRNA)-seq approach, described initially for *Helicobacter pylori* (Sharma et al., 2010), is being used extensively to map TSSs in other prokaryotes including archaea (Butcher and Stintzi, 2013; Cortes et al., 2013; Lin et al., 2013b; Pfeifer-Sancar et al., 2013; Rumbo-Feal et al., 2013).

Since the initial description of dRNA-sequencing, Illumina sequencing has been replacing 454 sequencing in the mapping of the transcriptional landscapes of bacteria. Illumina technology is less expensive and can generate 10 fold more reads than its 454 equivalent. Moreover, although Illumina reads are shorter, they are sufficiently long (30 to 50 bp) to allow unambiguous assignment to positions within bacterial chromosomes. This suggested that differential RNA sequencing could be performed without erasing the secondary transcriptome, provided that TAP could be shown to remove pyrophosphates from 5' nascent ends efficiently. If it could, the subtraction of 5' end fragments before treatment from those detected after treatment with TAP would identify those associated with nascent ends. There was also a concern that TEX might be less efficient at degrading RNA from *S. coelicolor*, another bacterial species of interest to the laboratory (Uguru et al., 2005; Hong et al., 2007; Van Wezel and Mcdowall, 2011), as it has a high GC-content, which in turn can introduce more stable secondary structures into RNA. This chapter describes the development of the front end of a pipeline that allowed the nature of 5' ends to be

differentiated using TAP rather than TEX. In addition, it describes biochemical confirmation that the T170V mutant of RNase E reduces the efficiency of cleavages dependent on, but not those independent of the presence of a 5'-monophosphate, and the preparation of samples of total RNA digested with T170V for analysis by RNA–sequencing. The work relates to the broad objective of developing an approach that provides, for the first time, a transcriptome-wide view of steps in RNA processing and decay, with a focus on the role of direct-entry cleavage by RNase E. As described (Chapter 1), the laboratory has produced biochemical evidence that suggests many of the cleavages produced by RNase E might be independent of interaction with a 5'-monophosphorylated end. Moreover, the T170V mutant has been used in prior studies of RNase E cleavage (Kime et al., 2010; Kime et al., 2014). The threonine substitution for valine removes one of three hydrogen bond donors that can engage a 5'-monophosphorylated end (see Figure 1.4). A major role for direct entry would explain, at least in part, why the normal rapid degradation of only a proportion of the mRNAs in *E. coli* is highly dependent on 5' pyrophosphate removal by RppH (Deana et al., 2008; Kime et al., 2010).

3.2 Results

3.2.1 Removal of 23S and 16S rRNA from total RNA

To reduce the number of reads that would correspond to rRNA, a MICROBExpress kit (Ambion) was used to deplete rRNA from total RNA isolated from *E. coli* after it was treated with DNase I (Figure 3.1). MICROBExpress (Ambion) is a commercially available kit which removes 16S and 23S rRNA from a sample by using oligonucleotides that are complementary to these rRNAs and are immobilised on magnetic beads. RNA from an *rppH* deletion strain (JW2798*Δkan*) as well as a congenic wild-type strain (BW25113) were included to provide samples for the analysis of direct-entry cleavage (see below). The analyses confirmed that the DNase I treatment removed all visible signs of contaminating chromosomal DNA and the MICROBExpress kit removed the vast majority of 23S and 16S rRNA (>95%, as specified by the manufacturer) without causing any degradation of the samples. Interestingly, a band of small RNA and a much larger RNA species of unknown identity (marked sRNA and by a double asterisk respectively) were detected in the *ΔrppH* strain. The degradation or processing of these species appears to be dependent on 5' pyrophosphate removal.

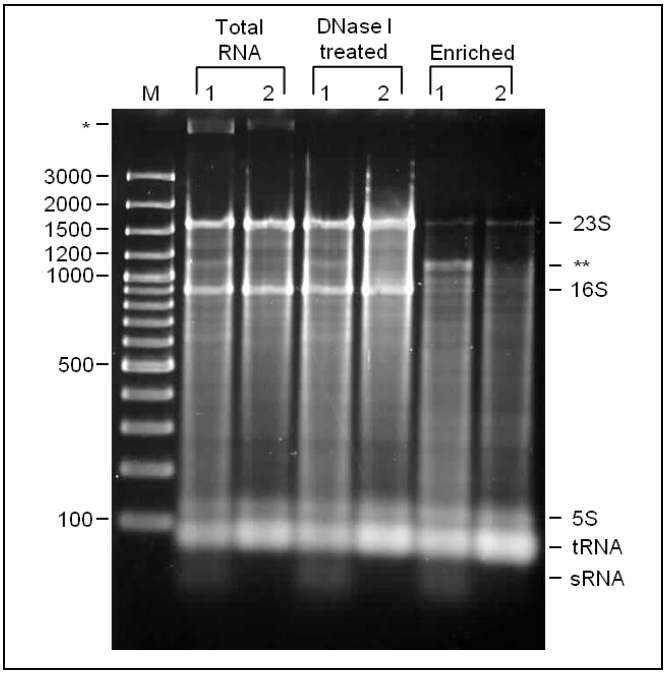

Figure 3.1 - Removal of large ribosomal RNA species from *E. coli* total RNA samples. Total RNA isolated from *E. coli ΔrppH* (lane 1) and its congenic wild-type, BW25113, (lane 2) was treated with DNase I in order to remove chromosomal DNA (labelled *) and then enriched for mRNA with the MICROBExpress kit (Ambion). The marker lane (labelled M) contains a GeneRuler 100 bp Plus DNA Ladder (Fermentas). Numbers at the left hand side of the panel indicate the sizes (bp) of selected markers, whereas the labelling at the right hand side indicates the position of 23S, 16S, 5S rRNA, tRNA, and smaller RNA species (sRNA). A large species that is more abundant in the *ΔrppH* strain is indicated by double asterisks (labelled **). The amount loaded was 1 µg of total RNA or the equivalent amount of enriched RNA. RNA was extracted from exponential cell cultures (0.6 OD_{600} units) grown in LB media. Electrophoresis was performed using a 2% (w/v) agarose gel.

3.2.2 **Characterisation of TAP**

Having found a straightforward method for removing the bulk of the large rRNA in *E. coli*, the next step was to confirm that TAP removes 5' pyrophosphates efficiently. This required the synthesis of 5'-triphosphorylated substrate. The template for T7 polymerase mediated *in vitro* transcription of *cspA* mRNA, which was the subject of previous studies (Hankins et al., 2007; Hankins et al., 2010; Kime et al., 2010), was generated by PCR, as described previously (Kime et al., 2010). The amplicon was of expected size (444 bp) and could be transcribed to produce a single RNA species of the expected size (423 nt) (Figure 3.2).

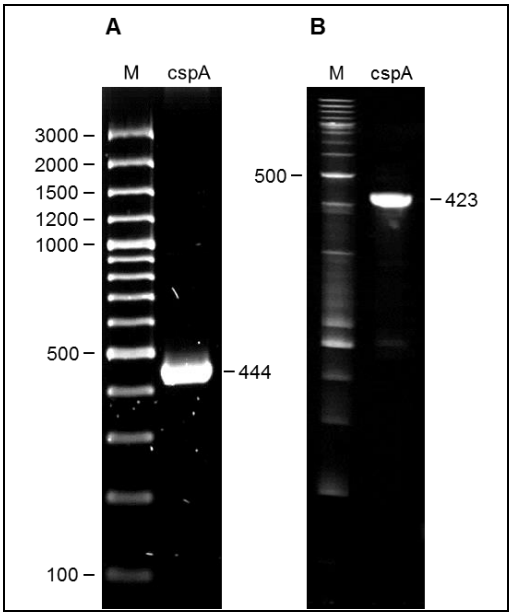

Figure 3.2 - Preparation of 5' triphosphorylated *cspA* mRNA. A template for production of *cspA* mRNA *in vitro* was generated by PCR using primers cspA F and cspA R2 (Table 2.2) (panel A) and then transcribed *in vitro* using T7 RNA polymerase (for details, see Section 2.5.6) (panel B). Amplicon and transcript sizes are indicated on the right. The source of markers in both panels was the GeneRuler 100 bp Plus DNA Ladder (Fermentas) which in panel B is denatured and was used to give a rough estimate of size. The PCR amplicon (~1.3 µg) was analysed by agarose gel electrophoresis (panel A), whereas, the *in vitro*-transcribed RNA was analysed under denaturing conditions using a 7 M urea, 7% (w/v) polyacrylamide (19:1) gel (panel B) and both were stained using ethidium bromide.

RNA generated by *in vitro* transcription is 5'-triphosphorylated (Golomb and Chamberlin, 1974). Incubation with TAP and TEX individually did not produce a detectable change in the migration of the *cspA* transcript (Figure 3.3). This was the expected result as pyrophosphate removal does not change the length of a transcript significantly and TEX is reported to be specific for 5'-monophosphorylated substrates. Incubation with TEX following treatment with TAP resulted in the efficient degradation of the *cspA* transcript. Indeed, residual substrate was not detected. This indicated that the TAP treatment was efficient. Moreover, the TAP treatment on its own did not result in any decrease in the intensity of the major *cspA* species nor the appearance of additional bands. This result, which was obtained independently by other members of the laboratory, indicated that the preparation of TAP was free of RNase contamination. It was found, in our laboratory, that some commercial preparations of enzymes used to treat the 5' end of RNA can be contaminated with RNases. Calf-intestinal (alkaline) phosphatase (CIP) was trialled in combination with T4 polynucleotide kinase (PNK) to monophosphorylate the 5' ends of nascent transcripts. However, a source of CIP that was not contaminated with RNase activity could not be found (data not shown).

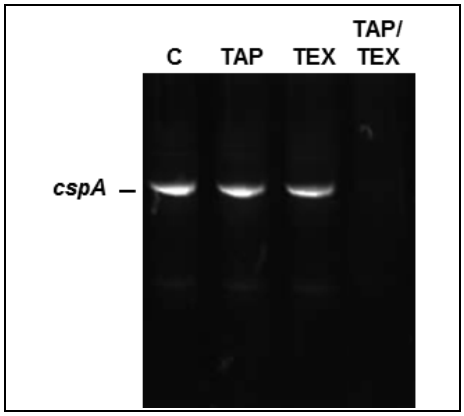

Figure 3.3 – Characterisation of TAP and TEX treatments. *cspA* mRNA was incubated in the absence of enzyme as a negative control (labelled C) or with TAP, TEX, or TAP followed by TEX (labelled at top of panel). The reactions were performed as outlined in Section 2.5.8. Aliquots of RNA (~500 ng) of each treatment were analysed by denaturing, polyacrylamide gel electrophoresis. The 10% (w/v) polyacrylamide (19:1) gel and was stained using ethidium bromide.

In parallel, the specificity of TEX for 5'-monophosphorylated ends was checked by co-incubating 5'-triphosphorylated *cspA* with an RNA oligoribonucleotide, R1, that had been monophosphorylated at its 5' end using T4 polynucleotide kinase (Figure 3.4). Only the oligonucleotide was degraded by TEX. This confirmed that the efficient degradation of *cspA* mRNA following TAP treatment was due to the generation of 5'-monophosphorylated ends. 5'-hydroxylated R1 was not degraded by TEX (Lin, unpub. result).

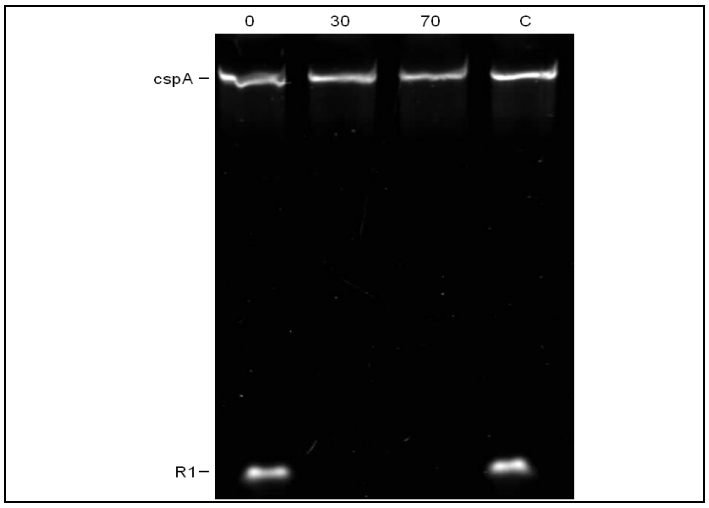

Figure 3.4 – Assay of the specificity of TEX towards 5' phosphorylated ends. 3 µg of an equimass mix of a 5'-monophosphorylated RNA oligonucleotide (labelled R1) (5'-CCUUUUCAAGACAUGCAACAAUGCACACAG-3') and 5'-triphosphorylated *cspA* mRNA (labelled *cspA*) was incubated with 3 U TEX. Samples of the reaction were taken after 0, 30 and 70 min and electrophoresed, along with a control which was incubated without TEX for 70 min (labelled C), on a 13% denaturing polyacrylamide gel with staining using ethidium bromide.

3.2.3 **Characterisation of the RNase E T170V substitution**

3.2.3.1 Determination of the enzyme stock concentration

Prior to conducting enzymatic comparisons, the concentrations of laboratory stocks of preparations of NTH-RNase E and the T170V mutant were confirmed by comparing aliquots against a dilution series of BSA standards using SDS-PAGE (Figure 3.5). Both NTH-RNase E and T170V migrated with an apparent mass of ~ 65 kDa, as previously described (Callaghan et al., 2003; Callaghan et al., 2005; Chandran and Luisi, 2006). Previously, each of the two stocks had been adjusted to 2 µM (tetramer concentration) on the basis of spectrophotometric analysis (absorbance at 280 nm); this corresponds to 500 ng/ µl. In agreement, the intensity of the bands corresponding to 1 µl aliquots of RNase E were similar to the band of BSA corresponding to 500 ng. The results of this analysis were also consistent with the preparations being described as >95% pure. The preparations had been purified as described previously (Kime et al., 2008a).

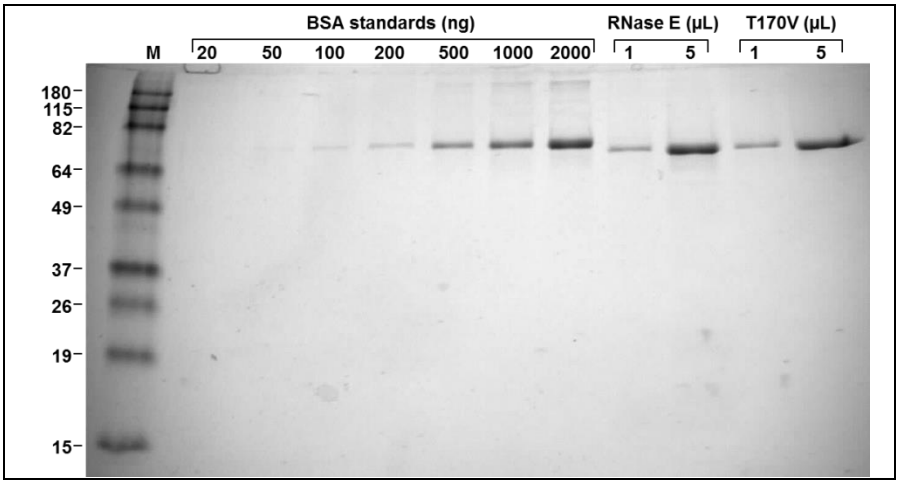

Figure 3.5 - Determination of NTH-RNase E and T170V mutant stock concentration. A serial dilution of BSA was prepared and aliquots containing amounts ranging from 20 to 2000 ng were compared against 1 and 5 µl volumes of NTH-RNase E and T170V. The intensity produced by loading 1 µl aliquots of preparations of NTH-RNase E and the T170V mutant correlated to that of the 2000 ng BSA standard marker. Markers (labelled M) are the Benchmark pre-stained protein ladder (Invitrogen) and the size, in kDa, of selected markers are labelled on the left. The gel was 18% (w/v) acrylamide (19:1).

3.2.3.2 Assessment of enzyme stock activity

Next the activity of the stocks described above was confirmed using derivatives of the oligonucleotide substrate BR13 that were labelled with fluorescein at the 3' end and had either a monophosphate or hydroxyl group at the 5' end. These substrates have been used previously to characterise mutants in the 5' sensor domain of RNase E and RNase G

(Jourdan and Mcdowall, 2008; Kime et al., 2008a). As reported previously, the 5'-monophosphorylated substrate was cleaved less efficiently by the T170V mutant than wild-type NTH-RNase E (Figure 3.6). The 5'-hydroxylated substrate was cleavage inefficiently by both enzymes. The differences in the initial rates were also in accordance with what had been reported previously (Jourdan and Mcdowall, 2008). T170V cleaved 5'-monophosphorylated BR13 11-fold slower than NTH-RNase E (1 vs 11.7 nM/min, respectively). The initial rates for cleavage of 5'-hydroxylated substrate by T170V and NTH-RNase E were similar to each other but slower than 5' –monophosphorylated BR13 (0.4 and 0.27 nM/min respectively).

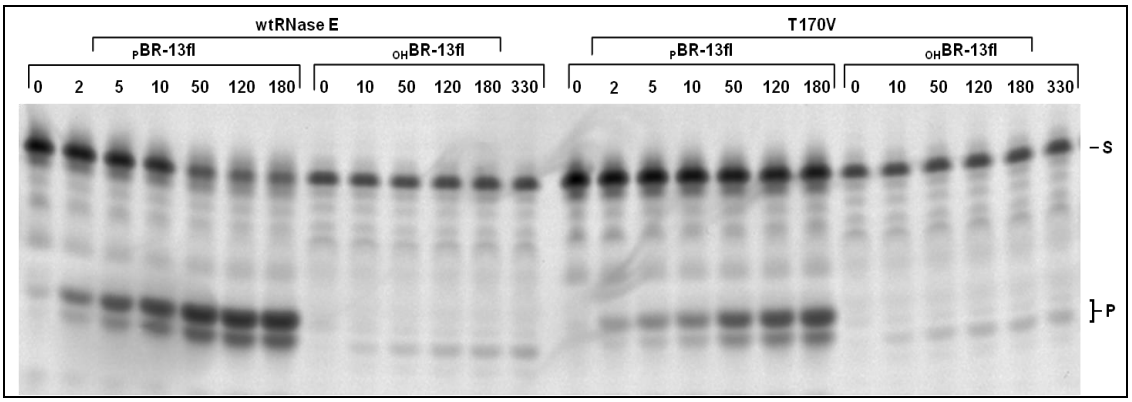

Figure 3.6 - Characterisation of the RNase E T170V mutant. A discontinuous cleavage assay containing derivatives of BR13 and either T170V or wild-type RNase E was electrophoresed on a 7M urea, 13% (w/v) polyacrylamide (19:1) gel. The concentrations of substrate and enzyme at the start of the reaction were 250 nM and 5 nm, respectively. For $_{P}$BR13, time points were taken at 0, 2, 5, 10, 50, 120 and 180 min following addition of substrate to the reaction; time points for $_{HO}$BR13 were taken at 0, 10, 50, 120, 180 and 330 min. All reactants were pre-warmed at 37 ^{o}C for 30 min (Kime et al., 2010) prior to starting the reactions. Substrate bands are labelled S whilst products are labelled P, indicated on the right.

The activity of these enzyme stocks were also confirmed using *cspA in vitro* transcribed mRNA (Kime et al., 2010). Unlike the BR13 oligonucleotide substrate, *cspA* mRNA is cleaved efficiently in the absence of a 5'-monophosphate (Kime et al., 2010). To provide a control, RNAI, the antisense RNA regulator of ColE1 replication was included. It is well established that cleavage of this RNA requires interaction with a 5'-monophosphate (Jiang et al., 2000; Kime et al., 2010). As described above for *cspA* mRNA (Section 3.2.2), RNAI RNA was generated by *in vitro* transcription from templates produced by PCR. Both the amplicon and its transcriptional product migrated as expected for their predicted sizes, which were 132 bp and 111 nt (Figure 3.7). Following *in vitro* transcription, *cspA* and RNAI transcripts

where incubated with NTH-RNase E and T170V (Figure 3.8). It was shown that, as reported recently (Kime et al., 2010), 5'-triphosphorylated *cspA* mRNA was cleaved efficiently by both NTH-RNase E and T170V. The site of cleavage is located close to the 3' end (Hankins et al., 2007; Kime et al., 2010) producing a detectable upstream product (labelled p) that migrates slightly faster than the full-length substrate. As expected, RNAI was not cleaved by either NTH-RNase E or T170V.

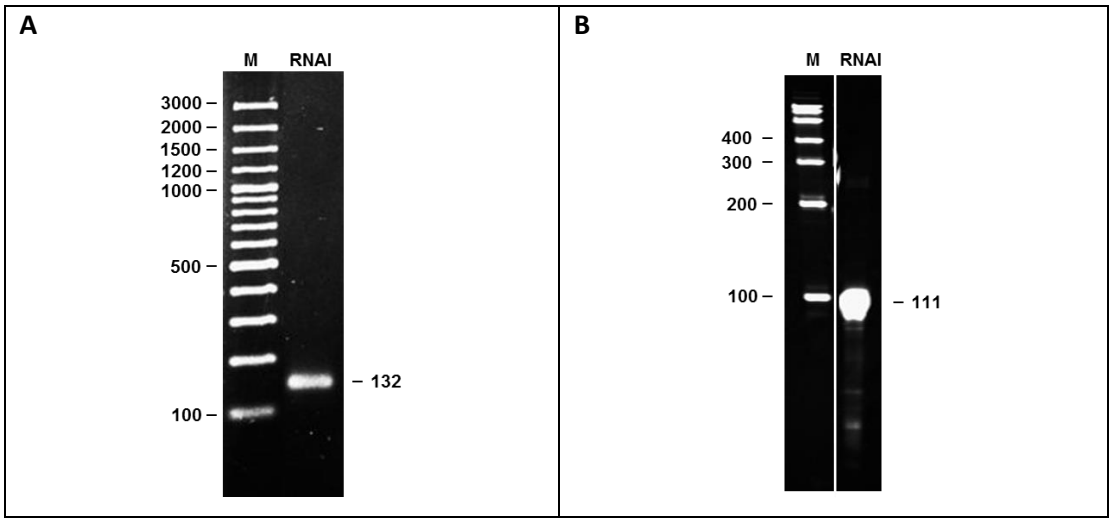

Figure 3.7 - Preparation of 5'-triphosphorylated RNAI RNA. The template for the *in vitro* synthesis of RNAI RNA was generated by PCR using primers RNAI F and RNAI R (Table 2.2) (panel A). The template was then transcribed *in vitro* using T7 RNA polymerase, (see Section 2.5.6) (panel B). Amplicon and transcript sizes are indicated on the right. The marker (lane M) in panel A was the GeneRuler 100 bp Plus DNA Ladder (Fermentas) and for panel B the marker was the RiboRuler™ low range RNA ladder (Fermentas). The sizes of the markers are indicated on the right of each panel. The PCR amplicon (~1.3 µg) was analysed by agarose gel electrophoresis, whereas, the *in vitro*-transcribed RNA was analysed under denaturing conditions using a 7 M urea, 7% (w/v) polyacrylamide (19:1) gel and both were stained using ethidium bromide.

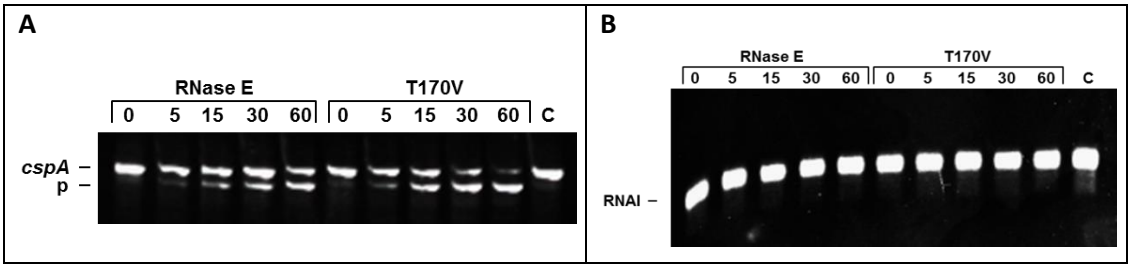

Figure 3.8 – Cleavage of *cspA* and RNAI mRNA by the N-terminal half of RNase E. The concentrations of substrate and enzyme at the start of the reaction were 0.64 µM of *cspA* mRNA (panel A) or RNAI (panel B) and 25 nM of wild-type NTH-RNase E or the T170V mutant, respectively (see Section 2.5.7). Samples were taken at 0, 5, 15, 30 and 60 min following addition of substrate to the reaction. A control of each mRNA which was incubated in the absence of enzyme for 60 min was also electrophoresed (lanes labelled C). The samples were analysed under denaturing conditions using a 7 M urea, 7% (w/v) polyacrylamide (19:1) gel and were stained using ethidium bromide.

3.2.4 **Determination of conditions for the incubation of total RNA with RNase E and T170V**

Following empirical testing, conditions were obtained that resulted in detectable cleavage of total RNA samples (Figure 3.9). Incubation of NTH-RNase E (300 nM) with total RNA (10 µg/180 µl reaction) from wild-type *E. coli* (BW25113) or the congenic *ΔrppH* strain (JW2798*Δkan*) for 15 min was sufficient to detect significant amounts of cleavage as evidenced by the appearance of discrete product bands and a reduction in the intensity of the large rRNA species, in particular 16S rRNA. Total RNA from the *ΔrppH* strain was included to evaluate the overall contribution of 5' pyrophosphate removal on cleavage by RNase E. Interestingly, cleavage in terms of level and pattern was broadly similar, although not identical, to that of the equivalent RNA from the wild-type strain. This result suggested that the absence of 5'-monophosphorylated ends that would normally be generated by RppH does not have a major effect on RNase E cleavage of total RNA, at least *in vitro*. However, absence of *rppH* did not prevent enriched mRNA being susceptible to TEX (Figure 3.10). One explanation for this observation, which is discussed in Chapter 1 is that the majority of the RNAs in total RNA do not correspond to nascent transcripts and are in fact degradation intermediates. Regardless of the composition of visible species in total RNA, it appears that direct entry may be a major pathway of RNase E cleavage. For both RNA samples, incubation with T170V and NTH-RNase E produced similar levels and patterns of cleavage (Figure 3.9). The actual sites of cleavage were then mapped using an RNA sequencing approach that is described in Chapter 1.

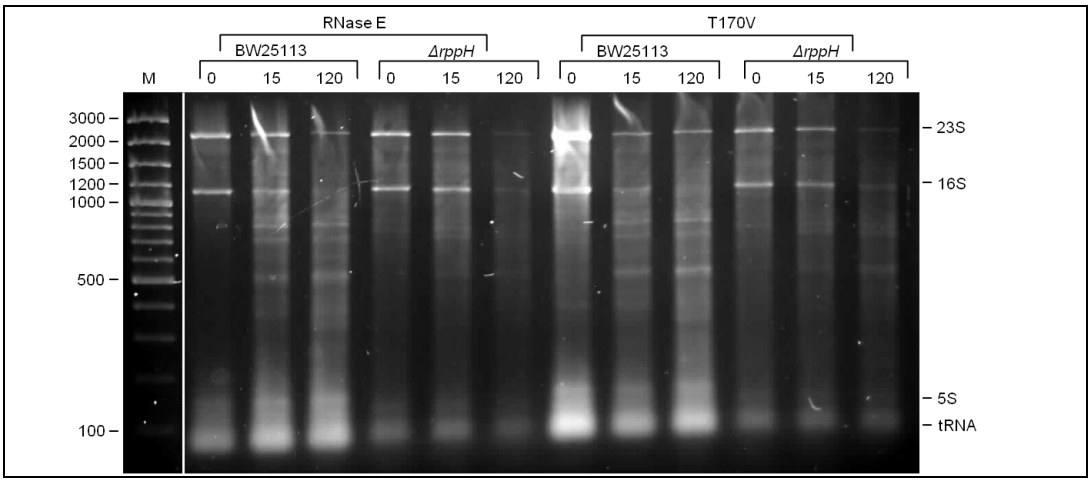

Figure 3.9 - Cleavage of total RNA by the N-terminal half of RNase E and the T170V mutant. Total RNA was isolated from a *ΔrppH* mutant of *E. coli* and BW25113, a congenic wild-type strain. The amount of total RNA in each 180 µl of reaction was 10 µg. The concentration of enzyme at the start of the reaction was 300 nM. Samples were removed prior to the addition of enzyme (corresponding to lane 0) and after 15 and 120 min of incubation. The reaction conditions were as described (Section 2.5.7). The samples were extracted with phenol chloroform and precipitated with ethanol and 1 µg of each sample was then resuspended in loading buffer and analysed by agarose gel electrophoresis. Markers, labelled on the left of the panel, and gel electrophoresis are as Figure 3.1. The gel was 2% agarose (w/v). Numbers at the top of the panel indicate the time (min) at which samples were removed from each reaction.

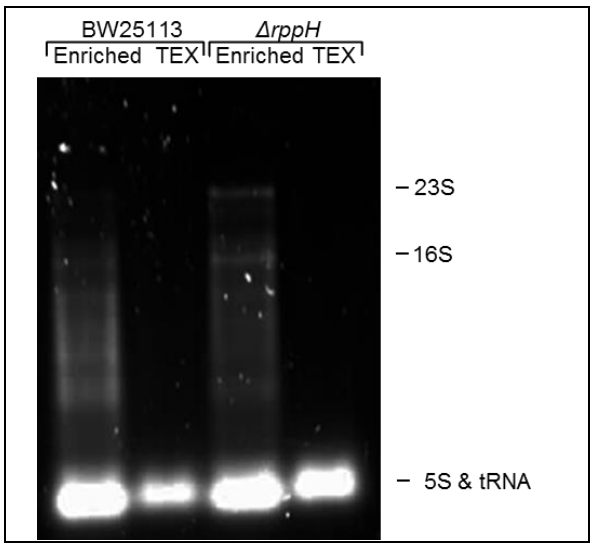

Figure 3.10 - TEX treatment of *E. coli* total RNA samples. Enriched mRNA (labelled Enriched) samples from *E. coli* strains BW25113 and its congenic *ΔrppH* mutant were treated with TEX (labelled TEX). Reaction conditions were as outlined in Section 2.5.8. The amount of RNA loaded was 1 µg of total RNA or the same amount of TEX treated RNA. The gel was 1.5% (w/v) agarose and stained with ethidium bromide.

3.3 Discussion

The work described in this chapter successfully: i) showed that the majority of 23S and 16S rRNA could be removed using the MICROBExpress system, with the aim of enabling better coverage of non-ribosomal species by RNA sequencing (Figure 3.1); ii) demonstrated that tobacco acid pyrophosphohydrolase (TAP) provides an effective means of discriminating the phosphorylation status at the 5' ends of transcripts (Figure 3.3 and Figure 3.4); iii) confirmed that the T170V mutant of RNase E is deficient in 5'-end sensing (Figure 3.6 and Figure 3.8); and iv) established the conditions for the cleavage of total RNA by NTH-RNase E *in vitro* (Figure 3.9). These results largely represent the preparatory work for RNA-sequencing analysis that ultimately proved to be successful (Chapter 4).

A finding that was initially unexpected was generated when preparing total RNA. In RNA prepared from the *ΔrppH* mutant there are small RNAs that migrate faster than tRNA that appear to accumulate (Figure 3.1). This RNA species could be sRNAs that are stabilized as a result of not being 'decapped' by RppH or mRNA transcripts with structured 5' UTRs, which can only be degraded via interaction with the 5'-monophosphate sensing pocket of RNase E.

Results of assaying the cleavage of total RNA also indicated that the majority of the cleavages produced by NTH-RNase were also produced by T170V. There was no substantial difference in the rate or pattern of cleavage (Figure 3.9). This result is consistent with the notion that direct-entry cleavage might, contrary to what has been suggested recently (Garrey et al., 2009), not be the exception. Our laboratory has proposed that the tetrameric structure of RNase E allows it to bind to substrates with high affinity provided it can contact simultaneously a single-stranded segment(s) in addition to the one in which cleavage occurs. The apparent widespread nature of direct entry, as found here (Figure 3.9), is consistent with such a simple requirement.

Chapter 4

4 Mapping of sites of RNase E cleavage within the transcriptional landscape of *Escherichia coli* as determined using a combination of global and differential RNA-seq

4.1 Introduction

Since its discovery in 1885 by Theodor Escherich, a German paediatrician, (Escherich, 1885) *Escherichia coli* has become an important model system for studying molecular and bacterial biology and establishing biotechnological platforms. *E. coli* is perhaps best known for being the organism that hosts the majority of recombinant DNA work. The first complete DNA sequence of an *E. coli* genome, laboratory strain K-12 derivative MG1655 (seq), was published in 1997 (Blattner et al., 1997). It consists of 4.6 Mbp and is currently annotated as encoding 4,284 protein-coding genes (Keseler et al., 2013) organised as 2,642 operons (Salgado et al., 2006), seven rRNA operons, and 86 transfer RNA genes (Keseler et al., 2013). The genome also contains a significant number of transposable genetic elements, repeat elements, cryptic prophages, and the evolutionary remnants of bacteriophages (Blattner et al., 1997).

The availability of the DNA sequence of MG1655 (seq) (Blattner et al., 1997), and now many other strains, permits the study of *E. coli* on a genome-wide scale. For example, there have been hundreds of studies on the control of gene expression at the RNA level using microarrays (Schena et al., 1995; Bier and Kleinjung, 2001; Goldsmith and Dhanasekaran, 2004) and more recently global RNA sequencing approaches (Butcher and Stintzi, 2013; Cortes et al., 2013; Lin et al., 2013b; Pfeifer-Sancar et al., 2013; Rumbo-Feal et al., 2013) alongside studies at the protein level using 2D-gel electrophoresis, a method which was used to assemble a 2D PAGE database (Appel et al., 1996) and has now shifted to the use of mass spectrometry approaches (Arifuzzaman et al., 2006; Cho et al., 2009). High-throughput approaches are also being developed to map macromolecular interactions (protein-protein and protein-nucleic acid) and the flux through major metabolic pathways (Rajagopala et al., 2014). This system level understanding provides an envelope in which to model the underlying regulation. *E. coli* offers an attractive model for systems biology, not only because of the relative ease with which it can be grown, but because much more is probably known about the mechanisms that regulate gene expression in *E. coli* than any

other organism. However, as is true for all model organisms, knowledge of the mechanism involved in mediating and regulating RNA degradation lags behind that of transcription and translation.

As outlined in Chapter 1, with the advent of microarrays it was possible to begin to identify transcripts whose degradation was dependent on particular mRNA decay factors (e.g. ribonucleases) and growth conditions. However, microarray-based studies on their own do not identify the site of action of ribonucleases. This would have required further analysis using techniques such as RNA-ligase mediated reverse transcription PCR, which identify differences in the pattern of 5' ends. A 'game changer' was the implementation of RNA sequencing. Although used initially to map transcriptional landscapes via the sequencing of fragments generated by shearing of RNA following its isolation, it was evident that RNA-seq could also be used to map 5' end fragments by incorporating an RNA ligation step before the shearing of the RNA (Figure 1.5). In such a way, Cynthia Sharma and co-workers mapped the transcriptional starts sites of *Helicobacter pylori*, but after removing degradation and processing intermediates by treating with TEX (Sharma et al., 2010). RNA ligation to the 5'-triphosphorylated ends of nascent transcripts was facilitated by treating with TAP. TAP treatments were performed by Vertis Biotechnologie AG (Germany), using the same enzyme supplier as the one used for the enzyme characterizations in Chapter 3. As indicated in the Introduction (Section 1.4), this paper provided the inspiration for the work described in this chapter, which describes the development of an alternative approach that, while still providing nucleotide resolution and differentiating sites of transcription initiation, simplifies the identification of the sites of processing and degradation. Comparison with results obtained previously validated this approach, which has been applied already to the study of other bacterial species. Within the *E. coli* genetic map, many new features were identified, such as previously undetected small RNAs and processing at a site associated with the production of specialised ribosomes, which may ensure the translation of leaderless mRNAs, which were also mapped. The approach also showed the benefit of incorporating steps that can differentiate the 5' status of transcripts in assigning sites of transcription initiation. RNA sequencing was also used to map sites of cleavage by RNase E (see Chapter 3). Finally, just as topographic maps are only of value with the inclusion of contour lines, a meaningful context for the analysis of sites of processing and degradation was provided by also determining the profile of global transcription.

4.2 Results

4.2.1 **Identification of transcription start sites**

Prior to analysing the data collected for the cleavage of total RNA by the T170V mutant of RNase E (see Section 3.2.4), sites of transcription initiation and RNA cleavage were identified and differentiated on a genome-wide scale using a published approach (Lin et al., 2013b) that was trialled as part of this work. These sites were then mapped against a transcriptional landscape map of *E. coli* generated from the same RNA sample using FRT-seq (flowcell reverse transcription), a global mapping approach that is strand specific and free of amplification prior to sequencing (Mamanova et al., 2010). The RNA sample for these differential and global approaches was isolated from a wild-type strain, MG1655 (seq) (Blattner et al., 1997), grown exponentially in Luria-Bertani medium. For dRNA-seq, total RNA was enriched for mRNA and then one half of the sample was treated with tobacco acid pyrophosphatase (TAP). Treatment with this enzyme removes pyrophosphate from nascent 5' ends leaving a monophosphate group (Breter and Rhoads, 1979) to which a sequencing adaptor can be ligated in a subsequent step (Figure 4.1). Samples treated with or without TAP were then sent for sequencing as part of a service that also included the construction of the cDNA libraries. It should be noted that the RNA was fragmented after the addition of the 5' adaptor to improve the efficient cloning of 5' ends from large transcripts. Sequencing of each of the two libraries produced 3 to 9 million reads, which were mapped against the *E. coli* genome. For each position in the genome, the number of times it corresponded with the first nucleotide in a sequence read was counted. A significant increase in the number of sequencing reads at any position following TAP treatment provides an identifier of a transcriptional start site (TSS). For further details of this differential approach, see Section 2.5.5.

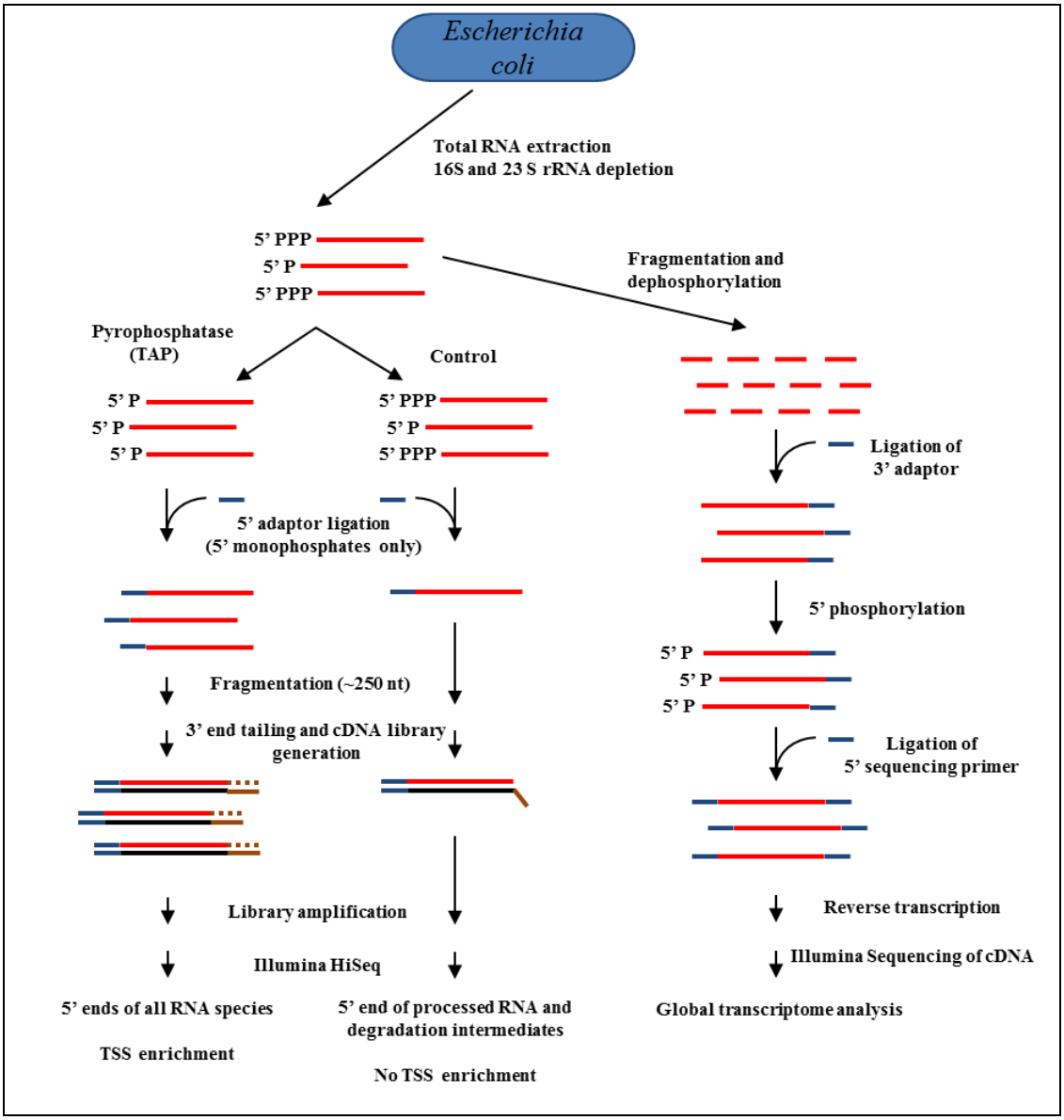

Figure 4.1 – RNA-seq pipeline used in this study. This figure presents the dRNA-seq and gRNA-seq approaches used in this study. On the left the dRNA-seq approach is illustrated with one sample being treated with tobacco acid pyrophosphatase (TAP) and the other not. On the right the gRNA-seq approach is illustrated. For a more detailed description refer to Figure 1.5.

Sites enriched following TAP treatment were identified by analysing M-A (ratio-intensity) scatterplots (Figure 4.2). Two populations of values were found, as described previously for *Propionibacterium acnes* (Lin et al., 2013b). The largest population corresponds to sites of processing and degradation and it centres on a value of M close to 0, while the smaller population corresponds to TSSs, which are associated with higher M values. Nucleotide positions with M values above what was judged to be the upper boundary of the population corresponding to processing and degradation sites were designated as possible TSSs. Positions within 8 nt of each other were assigned to the same TSS, as it is known that

many promoters initiate transcription from a cluster of nucleotide positions (Salgado et al., 2006). These sites were then mapped against leading edges of transcription, which were determined independently via manual inspection of the global RNA (FRT)-seq data (see Section 0), which had been processed at Leeds by Vijaya Mahalingam Shanmugiah. The reads from the global RNA-seq were trimmed and mapped to the genome, and for every position in the genome, the number of times it was read, irrespective of its position in individual RNA fragments, was counted and recorded in the format that allowed visualisation using the UCSC genome browser (Schneider et al., 2006; Chan et al., 2012). The upper range of the gRNA-seq reads was limited to make it easier to decipher their 5' and 3' boundaries, the transcription units becoming block-like in appearance. The main gene(s) of interest in any view are generally depicted left to right, and for genes on the reverse strand, the RNA-seq data are given negative values and are shown in red instead of black (for example, see Figure 4.3).

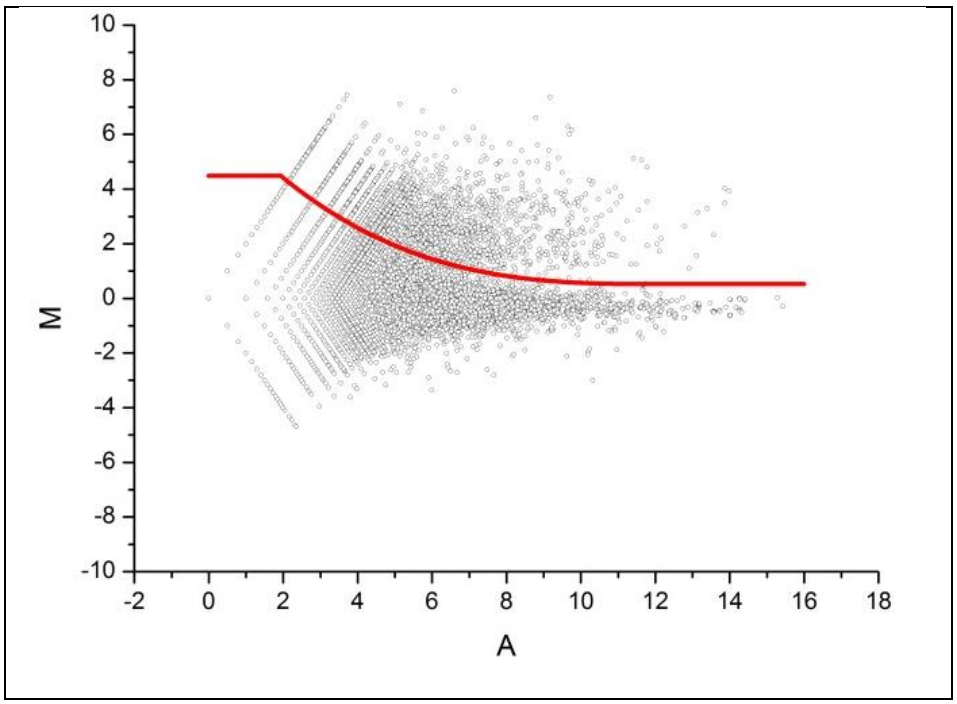

Figure 4.2 - M-A scatterplots of values from the differential RNA-seq analysis. The M values correspond to Log_2 (plus/minus) and A values to (Log_2 plus + Log_2 minus)/2, where minus and plus refer to the number of reads before and after treatment with TAP. For further details, see Materials and Methods (Section 2.5.5). The red line represents the upper boundary of the population of values corresponding to site of processing and degradation. The upper boundaries were placed manually to enclose the majority of the lower population, while taking into consideration the spread of M values scattered around 0. The boundary was described by the polynomial equation $M = -0.003A^3 + 0.13A^2 - 1.57A + 7.08$.

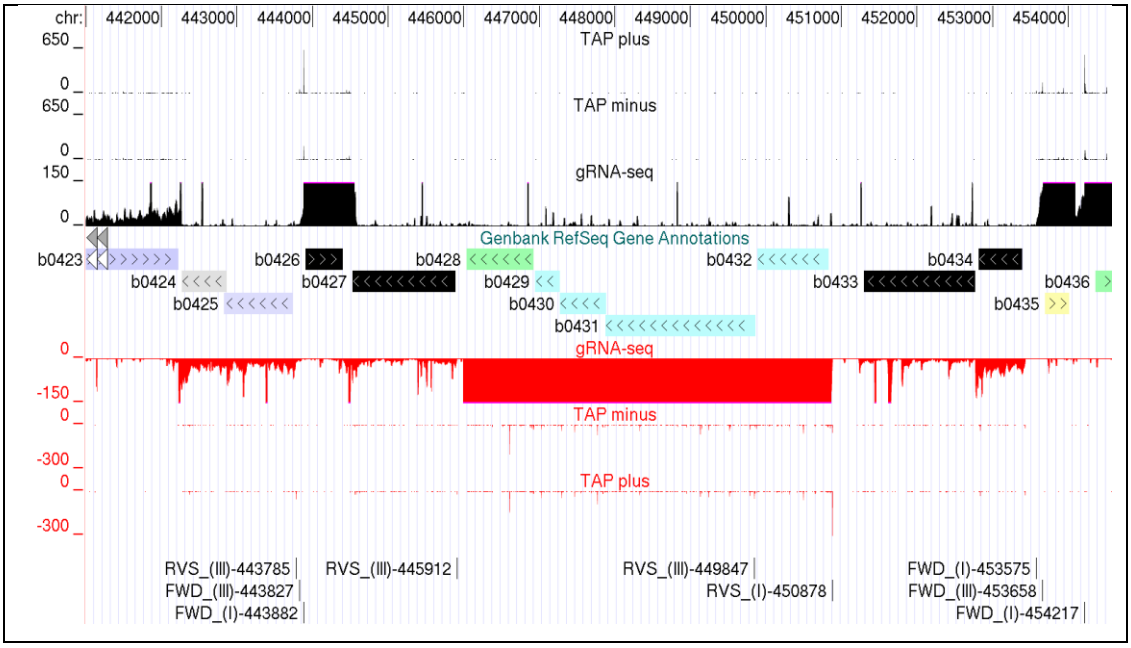

Figure 4.3 - Example of sequencing visualization in the UCSC Microbial Genome Browser. This figure presents a screenshot of the UCSC genome browser (Schneider et al., 2006; Chan et al., 2012). Tracks depict, from top to bottom, the number of times each nucleotide position was the first in sequence reads after and before treatment with TAP (dRNA-seq data), the number of times each position was sequenced following fragmentation of the transcriptome (gRNA-seq), the position of annotated genes (protein and RNA coding, as appropriate). The bottom three tracks, which are represented in red, depict the same results as their forward strand (black) equivalents for the reverse strand. The numbers at the left of the RNA-seq tracks indicate the sequencing reads, whilst the topmost numbers indicate genome position. The labels at the very bottom represent the candidate TSSs which are marked by short vertical lines that are labelled to indicate the strand of DNA to which they correspond (FWD or RVS), the class to which they belong (I, II or III) in parentheses, and the position of the first nucleotide in the site.

A total of 709 sites that were associated with leading edges of transcription as well as being enriched following TAP treatment (Class I), 311 that were associated with leading edges, but not enriched (Class II) and 1554 that were enriched, but not associated with leading edges of transcription (Class III) were identified. Although, Class I sites were assigned with the most confidence, being based on two criteria, all three classes contained TSSs that have been identified previously by others and recorded in RegulonDB (Salgado et al., 2006; Mendoza-Vargas et al., 2009a) (Table 4.1). This information has been included in the annotation of *E. coli* sites (Table S.1). The majority of the 5' ends identified by dRNA-seq do not correspond to TSSs, consistent with a major role of endoribonucleases in the degradation of mRNA (see Introduction, Table 1.2). Examples of each class of TSS are shown in Figure 4.4. The probable basis of the different classes, which has been described previously by us (Lin et al., 2013a), is outlined in Section 4.3.

The genes for which TSSs are presented in Figure 4.4 are: *thrL* which is the leader of the thrLABC operon that encodes four of the five enzymes involved in the threonine biosynthesis pathway (Gardner, 1982); *talB* which encodes transaldolase B (Sprenger et al., 1995) and *gapA* which encodes glyceraldehyde 3-phosphate dehydrogenase A monomer (Charpentier and Branlant, 1994). As discussed later, enrichment in the absence of an obvious step increase in transcript abundance (Class III) and a step increase in transcript abundance in the absence of enrichment (Class II) are a result of promoters nested downstream of strong promoters and efficient 5' pyrophosphate removal *in vivo*, respectively (Celesnik et al., 2007; Deana et al., 2008; Richards et al., 2011). All of the TSSs identified here have been annotated to indicate whether or not they have been identified previously using experimental approaches.

TSS Class	Number of TSSs identified	Overlap with RegulonDB TSSs	Percentage overlap
Class I	709	298	42
Class II	311	79	25.4
Clas III	1554	197	12.7
Total	2574	574	22.3

Table 4.1 - Overlap of TSSs identified with TSSs recorded in RegulonDB. Number of class I, II, and III TSSs identified by our approach are presented along with and their percent overlap with RegulonDB TSSs which have experimental evidence to support them.

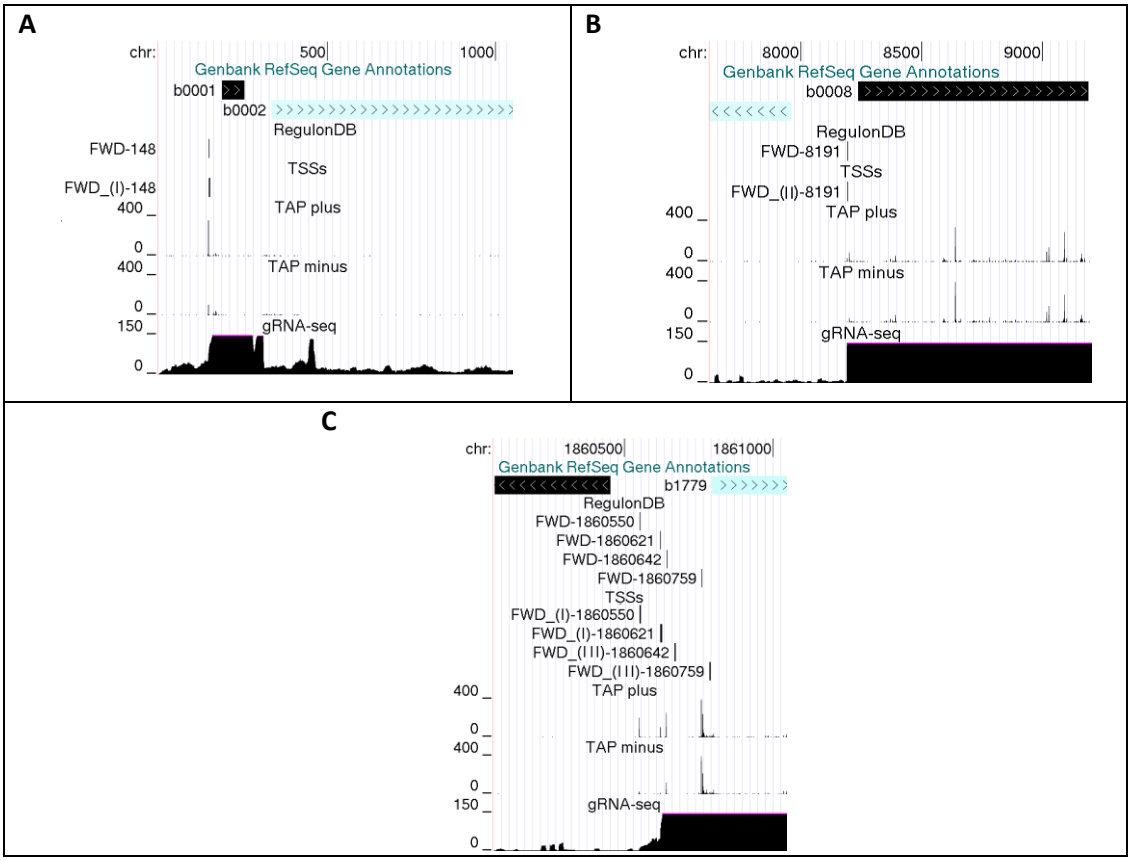

Figure 4.4 - Examples of different classes of transcription start site. All of the TSSs shown in this figure have been previously identified via experimental approaches. Panels A, B and C show the TSSs of *E. coli* genes *thrL* (Gardner, 1982; Lynn et al., 1987), *talB* (Sprenger et al., 1995) and *gapA* (Charpentier and Branlant, 1994) which correspond to a class I, class II and class III TSSs, respectively. The panels are screenshots from the UCSC Microbial Genome Browser (Schneider et al., 2006). Labelling is as in Figure 4.3.

Our dRNA-seq approach was validated by retrieving the sequences upstream of known TSSs for genes associated with the translational machinery. The vast majority of the sequences could be aligned, using MEME (Bailey et al., 2009), with the consensus sequences for the -35 and -10 boxes of vegetative promoters (TTGACA and TATAAT), which are recognised by the housekeeping RNA polymerase (Harley and Reynolds, 1987; Lisser and Margalit, 1993) (Table 4.2). Indeed the consensus sequence for the promoters aligned here was identical to that described previously (Figure 4.5). It also revealed a GC-rich heptanucleotide region immediately downstream of the -10 box which would correspond to the GC rich region which flanks the -10 box (Dickson et al., 1975; Hsu et al., 1984). The analyses described gave confidence that the dRNA-seq approach could identify TSSs and prompted similar analysis of equivalent dRNA-seq data for *P. acnes* (Lin et al., 2013b) and *S. coelicolor* (unpubl. data).

Table 4.2- Table showing the upstream region of the sites of enrichment for tRNA and rRNA TSS.

Suggested operon	TSS position	Enrichment	Strand	Sequence
rrsH-ileV-alaV-rrlH-rrfH-aspU	225017	4.20	+	TCATGGCCCTTACGACCAG GGCTACACACGTGCTACAATGGCGCATACA
aspV	236910	13.95	+	GAAAAACCGTTGACGAAGG TCGAGGCAATCCGTAATATTCGCCTCGTT
Pseudo	344476	5.50	+	ATGTGTTGGTTGACATTCA TATGAAAAAAATCATAATTCCATCA
argU	563921	10.50	+	CAAAAGCCATTGACTCAGA AGGGTTGACCGTATAATTCACGCGATT
infA-serW	925908	9.86	-	TGAATGTTTTCGGCACATT TCTCCCCAGAGTGTTATAATTGCGGTC
serX	1096903	18.50	-	AAAAGTTTGTTGACCTCAG GTCATGATTTCCCTAAATTAGCGCCCGTT
glyW-cysT-leuZ	1990208	23.77	-	AAAATATCGTTGACTCATC GCGCCAGGTAAGTAGAATGCAACGCATC
serU	2041590	4.49	-	CAGGGACTGTTAAAATGCC AAATTTCCTGGCATCATGGCAACCATCT
asnT	2042562	31.50	+	AATTTAGTGTTGACAGACA AGGTACCGCTAAGTAATATTCGCCCCGTT
asnW	2056138	19.53	-	AAACAGGCTTTGACATTGT GGGTGGGCATCGCTAATATTCGCCTCGTT
snU	2057866	12.38	+	AAATCGGGTTTGACAAAAG ATTTTTCGCCGTTAAGATGTGCCTCAAC
asnV	2060273	19.77	+	TTTCACCCTTTGACATCAC CATGCACTGCCATTAATATGCGCCCCGTT
proL	2284227	18.86	+	ATATCTTACTTGCAATCGG TGTGGAAAACGGTAGTATTAGCAGCCACGAG
alaW-alaX	2516279	19.25	-	AATTTGCCGTTGACACATT CGGGCGGAATTCATATGATGCCGCCCGTC
valU-valX-valY-lysV	2518944	11.21	+	GAAAATGCGTTGACTCATT TTGAACTCTCCCTATAATGCGACTCCAC
rrsG-gltW-rrlG-rrfG	2727935	5.92	-	TCATGGCCCTTACGACCAG GGCTACACACGTGCTACAATGGCGCATACA
rrsG-gltW-rrlG-rrfG	2729661	6.63	-	CAGCAAATACTGTCTGGTG AATTGGTTCCGGGTAAAGTGATTCGCCTG
ssrA	2753608	10.22	+	CATTGAGGCTGGTCATGGC GCTCATAAATCTGGTATACTTACCTTTACA
serV-argV-argY-argZ-argQ	2816723	27.73	-	AAAAATTGTTTGACTTATA AGTCTCAGAAAGTAATATGTGCGCCACG
glyU	2997092	15.57	-	GCGTGGCACTTGCTAAGGA GAGCGTAAGGTTTATAATGCCTTACGCA
pheV	3108325	5.56	+	TGCCTGACAATGCGTGCAATATCGGCAAAGTGATGATAGATTGTGCAGTCTG
pheV	3108385	35.08	+	GAAATTTGATTGACGAGAC GAGGCGAATCAGGTTTAATGCGCCCCGTT
MetCAT	3213618	13.50	+	ATGGCTGGATTGCGACACG GAGTTACTTTATAATCCGCTACCAT
leuU	3320658	7.06	-	ACATCGGTTTTGCTGTTTT TTTCCGCAGTTGATACAATGCGATAAAAT
rplM-rpsI	3376832	21.69	-	AAAAAGGGGTCGATCTTTG ACCCCGACTTCTCTATAATCCTGCGACCC
rrsD-ileU-alaU-rrlD-rrfD-thrV-rrfF	3424407	5.27	-	GTGAGCTCGATGAGTAGGG CGGGACACGTGGTATCCTGTCTGAATAT
rrsD-ileU-alaU-rrlD-rrfD-thrV-rrfF	3425540	4.33	-	TCATGGCCCTTACGACCAG GGCTACACACGTGCTACAATGGCGCATACA
rpsL-rpsG-fusA-tufA	3469747	8.90	-	ACCATGACGTTGACTCCTC TGAACTGGCGTTTAAACTGGCTGCTT
proK	3706721	3.44	-	ATTAAGGGATTGACGAGGG CGTATCTGCGCAGTAAGATGCGCCCCGCA
selC	3834221	24.00	+	TGGGGGATGTAGAAACTCA AGGAAGTAGCTATAATGCGCCCCGCC
rrsC-gltU-rrlC-rrfC-aspT-trpT	3941078	9.41	+	TCATGGCCCTTACGACCAG GGCTACACACGTGCTACAATGGCGCATACA
rrsA-ileT-alaT-rrlA-rrfA	4033382	43.00	+	AATAAATGCTTGACTCTGT AGCGGGAAGGCGTATTATGCACACCCCG
rrsA-ileT-alaT-rrlA-rrfA	4034801	10.43	+	TCATGGCCCTTACGACCAG GGCTACACACGTGCTACAATGGCGCATACA
rrsB-gltT-rrlB-rrfB	4165929	9.81	+	TCATGGCCCTTACGACCAG GGCTACACACGTGCTACAATGGCGCATACA
rrsB-gltT-rrlB-rrfB	4166279	8.20	+	AAGCGTACTTTGTAGTGCT CACACAGATTGTCTGATAGAAAGTGAAAAGCA
thrU-tyrU-glyT-thrT	4173404	16.07	+	ATTTTTTAGTTGCATGAAC TCGCATGTCTCCATAGAATGCGCGCTACT
rrsE-gltV-rrlE-rrfE	4207417	11.58	+	TCATGGCCCTTACGACCAG GGCTACACACGTGCTACAATGGCGCATACA

Table 4.2 (continued)

Suggested operon	TSS position	Enrichment	Strand	Sequence
pheU	4360653	19.42	-	GAAATTAGGTTGACGAGAT GTGCAGATTACGGTTTAATGCGCCCCGTT
glyV-glyX-glyY	4390333	11.50	+	TTTGGGGGGTTGCAGAGGG AAAGATTTCTCGTATAATGCGCCTCCCG

Table 4.2- Table showing the upstream region of the sites of enrichment for tRNA and rRNA TSS. tRNA and rRNA genes which are in close proximity and or which gRNA-seq data showed were transcribed as a single unit are presented as "Suggested operons". TSS loci are also presented alongside the strand and enrichment for each of these suggested operons. Sequences upstream of the +1 nucleotide (marked red) were aligned. Nucleotides present around the -10 region and matching the -10 (TATAAT) consensus sequence, as well as those in the -35 region matching the -35 (TTGACA) consensus sequence, are highlighted in grey.

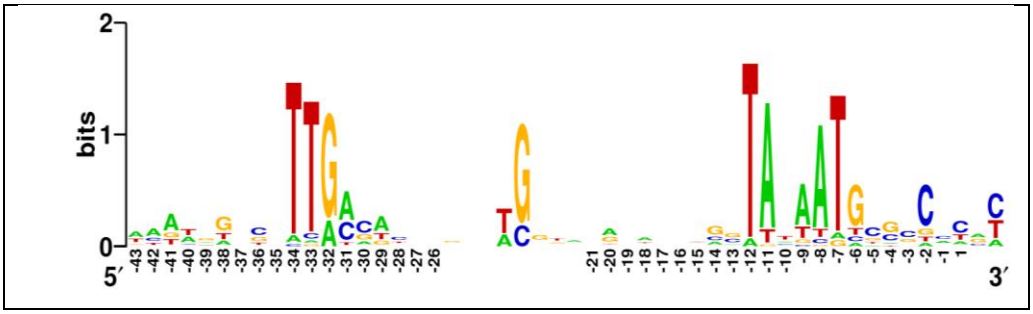

Figure 4.5- Weblogo generated following manual alignment of tRNA TSSs identified by the approach here presented. Conserved sequences in the promoters associated with genes encoding the translational machinery. This figure presents Weblogo represented (Crooks et al., 2004) changing the length of the spacer of individual promoters to maximise the alignment of the -35 box. The combined height of nucleotide symbols shows the level of sequence conservation at a particular position, while the height of individual symbols within a stack of nucleotides indicates the relative frequency at that position. The nucleotide positions are numbered relative to the average position of TSSs and no labelling is present where the spacer was introduced or no consensus was present. Strong conserved sequences can be observed at sites that would correspond to the -10 and-35 boxes (TATAAT and TTGACA respectively).

4.2.2 **Leaderless mRNAs and ribosome processing**

Having access to nucleotide-resolution transcriptional maps annotated to show the positions of sites of transcription initiation and RNA processing and degradation allowed a genome-wide survey of aspects of gene expression and regulation. For example, it was found that in stark contrast to *P. acnes* (Lin et al., 2013b), *E. coli* MG1655 (seq) has only a few mRNAs that lack or have a short 5' leader (<10 nt) and thereby cannot be translated via the canonical Shine-Dalgarno (SD) interaction (Shine and Dalgarno, 1974; Shine and Dalgarno, 1975). It increasingly appears that different species of bacteria differ substantially in the extent to which they use the SD interaction to initiate translation (Nakagawa et al., 2010). While a gene ontology analysis of the 'leaderless' *P. acnes* mRNAs failed to identify enrichment of a particular function(s) (Lin et al., 2013b), 3 of the 5 leaderless *E. coli* mRNAs encoded repressors of prophage (Qin, Rac and e14). This extends the association between leaderless mRNA and repressors of mobile genetic elements in *E. coli*, which has been a major bacterial model for the study of 'leaderless' translation (Moll et al., 2002; Malys and Mccarthy, 2011). The two best-studied leaderless mRNAs in *E. coli* encode the cI repressor of bacteriophage lambda (Walz et al., 1976) and the TetR repressor of transposon Tn1721 (Baumeister et al., 1991). Neither lambda nor Tn1721 are present in the MG1655 (seq) strain of *E. coli* studied here. The two other *E. coli* mRNAs identified as being leaderless (Figure 4.6) encode housekeeping proteins, the RhlB helicase and phosphatidylglycerophosphatase A (PgpA). The former is a component of the RNA

degradosome, which as described earlier is central to both the processing and degradation of RNA (Carpousis et al., 2009), while the latter is involved in the biosynthesis of phospholipids (Lu et al., 2011). The start codon of *rhlB* has been confirmed by N-terminal sequencing of its product (Py et al., 1996). The leaderless mRNAs described above were identified by using the VLOOKUP function of Excel to identify TSSs that were associated with start codons no more than 10 nt downstream (Table 4.3). If the analysis was extended to start codons within a 20 nt leader, another 15 potential leaderless mRNAs would be identified. In all 15 of them, ribosome binding sites were not detected using the RBSfinder software (Suzek et al., 2001). However, leaders of between 10 and 20 nt could enable SD-type interactions

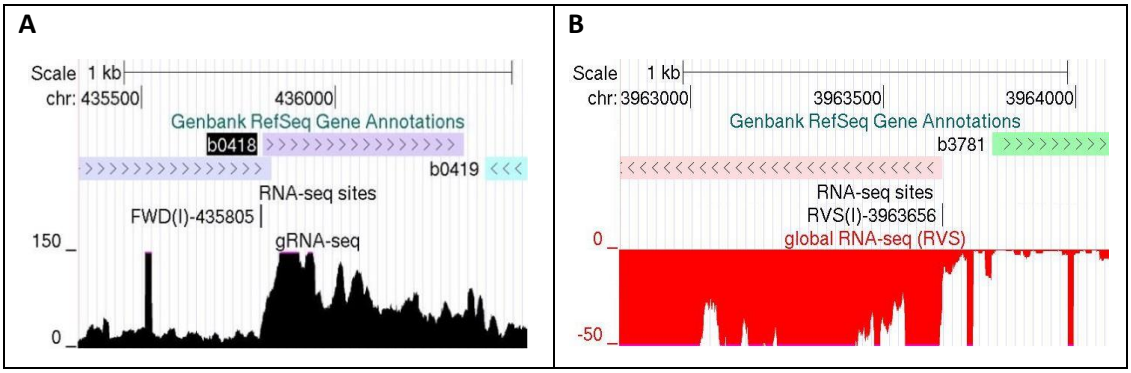

Figure 4.6 – *pgpA* and *rhlB* leaderless mRNAs. Panels A and B correspond to the *pgpA* and *rhlB* transcript respectively. Their TSSs are both class I and fall within 10 nt of the +1 site (in relation to translation) Labelling as in Figure 4.3.

Gene	Start position	Stop position	Closest TSS	Distance	Function
pgpA (+)	435812	436331	FWD(I)-435804	-8	Phosphatidylglycerophosphatase A
ymfK (-)	1202156	1201481	RVS(I)-1202156	0	e14 prophage; repressor protein phage e14
racR (-)	1418265	1417788	RVS(I)-1418265	0	Rac prophage; predicted DNA-binding transcriptional regulator
dicA (+)	1645957	1646365	FWD(I)-1645954	-3	Qin prophage; predicted regulator for DicB
rhlB (-)	3963653	3962387	RVS(I)-3963653	0	ATP-dependent RNA helicase

Table 4.3 - Leaderless mRNAs. This table presents the 5 mRNAs which were identified as being leaderless. From left to right, columns represent the gene name and direction in brackets, start and end codons followed by the closest TSS detected and the relative distance between this TSS and the annotated start codon. The known function for each gene is also presented.

Recently, it has been shown that leaderless *E. coli* mRNAs specifically generated by 5' trimming under conditions of stress are translated by specialised ribosomes from which the last 43 nt of the 3' end of 16S rRNA have been removed endonucleolytically (Vesper et al., 2011). This region contains the anti-SD sequence and the binding site of S1 (Shine and

Dalgarno, 1974; Lauber et al., 2012), a ribosomal protein that augments the SD interaction (Sorensen et al., 1998). This work prompted me to look for evidence of 3' processing of 16S rRNA in our RNA-seq data. Processing precisely 43 nt from the 3' end of *E. coli* 16S rRNA was detected (Figure 4.7). It is possible that specialised ribosomes similar to those generated during stress could facilitate the translation of the handful of leaderless mRNA present in *E. coli* during exponential growth, and that by the virtue of being leaderless the mRNA might be pre-programmed to be translated efficiently during conditions of stress when increasing numbers of specialised ribosomes are produced. For example, the continued repression of phage genes would be required to prevent premature cell death.

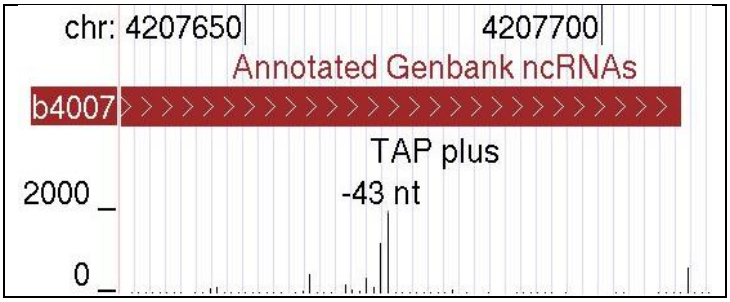

Figure 4.7 - 16S rRNA MazF equivalent cleavage site. This screenshot presents the 3' end of the 16S rRNA from the *rrnE* operon. The site where MazF mediated cleavage normally occurs is indicated by the legend "-43 nt". Labelling as in Figure 4.3. The only dRNA-seq track shown is "TAP plus" as there is no enrichment and "TAP minus" is identical to it.

4.2.3 **The maturation of stable RNAs**

The study of *E. coli* and to a lesser extent *B. subtilis* has revealed that mature ribosomal RNAs are produced via a series of nucleolytic steps involving several ribonucleases and that rRNA can be degraded in response to aberrant assembly of the ribosome or cellular stress (Deutscher, 2009). Remarkably, it was possible to detect for *E. coli* most of the known endonucleolytic processing sites, despite the transitory existence of the corresponding intermediates. All five of the known endonucleolytic steps that produce the mature 16S rRNA via the combined activities of RNases III, E and G and an as yet an unidentified nuclease, possibly RNase YbeY (Jacob et al., 2013), were detected (Young and Steitz, 1978; Li et al., 1999a; Wachi et al., 1999), as well as the RNase III cleavages involved in the maturation of 23S rRNA (Bram et al., 1980), and a tight cluster of cleavages that produce the 5' end of 5S rRNA, at least one of which is produced by RNase E (Misra and Apirion, 1979). Data shown is for the *rrnE* operon, but is representative of all seven *E. coli* operons (Figure 4.8). Several prominent sites internal to the rRNA and on the 3' side of tRNA internal to rRNA operons were also detected. Sites at the precise 3' ends of 23S and 5S RNA

were not detected as these are generated via 3' exonucleolytic trimming by RNase T (Li et al., 1998). Cleavage at a large number of sites internal to the functional regions of the mature rRNAs was also observed for *P. acnes* (Lin et al., 2013b). These nucleolytic events may control the quality of rRNA (and ribosomes) (Jacob et al., 2013), prevent rRNA accumulating in excess over ribosomal proteins (Norris and Koch, 1972; Gausing, 1977) or mediate the rapid turnover of prematurely terminated transcripts, which may occur more frequently in rRNA operons (Condon et al., 1995), or a combination of these. Moreover, the high density of nucleolytic sites in *E. coli* where there is no known 5' exoribonuclease, indicates the use of endoribonucleases on a scale perhaps not appreciated previously.

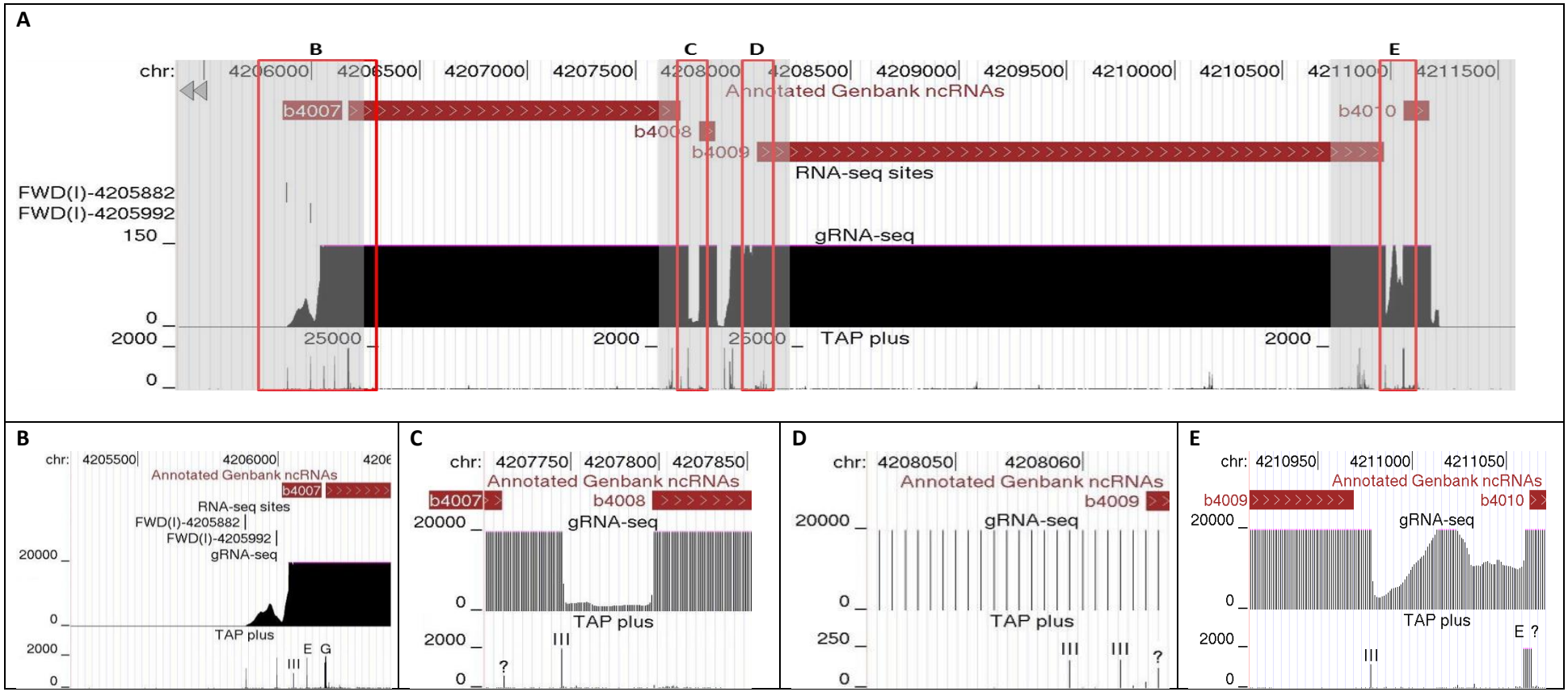

Figure 4.8 - Processing of the *rrnE* operon. This figure presents the rrnE operon (b4007, b4008, b4009 and b4010 which correspond to 16S rRNA, Glu tRNA, 23S rRNA and 5S rRNA respectively). Panel A shows the whole *rrnE* operon; red boxes labelled B, C, D and E are regions that were resolved to a higher degree, thus the decrease in the scale for panel D, and correspond to the panels B, C, D and E, respectively. Cleavage sites for RNaseE, G and III are labelled E, G and III respectively whilst sites for which a ribonuclease is yet to be identified are labelled "?" Labelling as in Figure 4.3 with the exception that the scale in the "TAP plus" lane in Panel A is variable with the 0-2000 scale regions highlighted grey and the 0-25000 scale regions showing no highlighting.

E. coli has been a major model for studying the processing of tRNAs. This study has led to the establishment of the model in which the mature tRNA 5' end is generated by the ubiquitous endonuclease RNase P, and the mature 3' end via endonucleolytic cleavage a few nucleotides downstream the CCA motif followed by 3' exonucleolytic trimming to this motif. The CCA motif is present in all of the 86 tRNA genes in *E. coli.* The maturation of the 3' end can be mediated by RNase BN (known as tRNase Z in other bacteria species), which has dual endo/3' exonucleolytic activity (Dutta and Deutscher, 2010; Dutta et al., 2012), or by the combined action of RNase E and 3' exonucleases, mainly RNases PH and T (Hartmann et al., 2009). Fully consistent with this model, in all cases where 5' and 3' processing sites could be detected for tRNAs, they were at the precise 5' end and within a few nucleotides downstream of the 3' end of *E. coli* tRNAs, respectively (Table 4.4). In contrast, the analysis of *P. acnes* found that most of the CCA-encoding tRNAs are cut *within* this motif between the Cs (Lin et al., 2013b). It is likely that, for *P. acnes*, the homologue of tRNA nucleotidyltransferase (SCO3896), which adds CCA to tRNAs not synthesised with this motif (Cudny and Deutscher, 1986), can recognise partial CCA ends and add only the residues that are missing. There is evidence that at least some tRNA nucleotidyltransferases, including the *E. coli* enzyme (Reuven et al., 1997), have the capability of repairing CCA (Betat et al., 2010). Such an activity in *P. acnes* would mean that cleavages within 3' CCA triplets would not result in terminal inactivation of the tRNA.

tRNA	Strand	5' end	3' end	3' cleavage site
ArgACG	-	2816571	2816494	CCAt↓at
AspGTC	+	236930	237007	CCA↓c↓ta
AspGTC	+	3944894	3944971	CCAccct↓a↓at
CysGCA	-	1990011	1989937	CCA↓c↓ttt↓ct
GlnCTG	-	695839	695764	CCAa↓ttt↓at
GlyGCC	-	1990141	1990065	CCAg↓tt
HisGTG	+	3980532	3980608	CCAtt↓at
MetCAT	-	695963	695886	CCAa↓at
SerGCT	-	2816667	2816574	CCAttt↓gc
ThrGGT	-	3421677	3421601	CCActt↓tt
TyrGTA	-	1286845	1286760	CCAta↓at
TyrGTA	+	4173494	4173579	CCA↓a↓tt

Table 4.4 - Processing of tRNA genes in *E. coli*. tRNA genes for which 3' processing was detected are presented alongside their respective start and end positions and their strand. Cleavage/s detected within the 10 nt downstream of the terminal CCA at the 3' end of each tRNA are indicated with an arrow within the sequence where they occur.

4.2.4 **The degradation and processing of mRNA**

With regard to mRNA, endonucleolytic sites involved in both the degradation and processing of *E. coli* mRNA were detected (Figure 4.9). This included RNase E sites that accelerate the turnover of *eno* (Kime et al., 2008a), *rpsT* (Coburn and Mackie, 1998) and *ompA* (Rasmussen et al., 2005) mRNAs by providing additional ends for 3' exonucleolytic attack. Additionally RNase III sites were detected in the 5' leader of *adhE* and *pnp* mRNA that facilitate translation (Aristarkhov et al., 1996) and an autoregulatory mechanism (Jarrige et al., 2001), respectively as well as RNase P sites that segment polycistronic mRNAs (Lee et al., 2008). In addition, previously uncharacterised cleavage sites within mRNAs that serve as models for understanding mRNA degradation, *e.g.* one internal to the coding region of *rpsT* mRNA and another that removes the 3' stem-loop of *ompA* mRNA, were detected. Thus, our RNA-seq approach not only confirms, but extends knowledge of events controlling the activity and longevity of mRNA in *E. coli*.

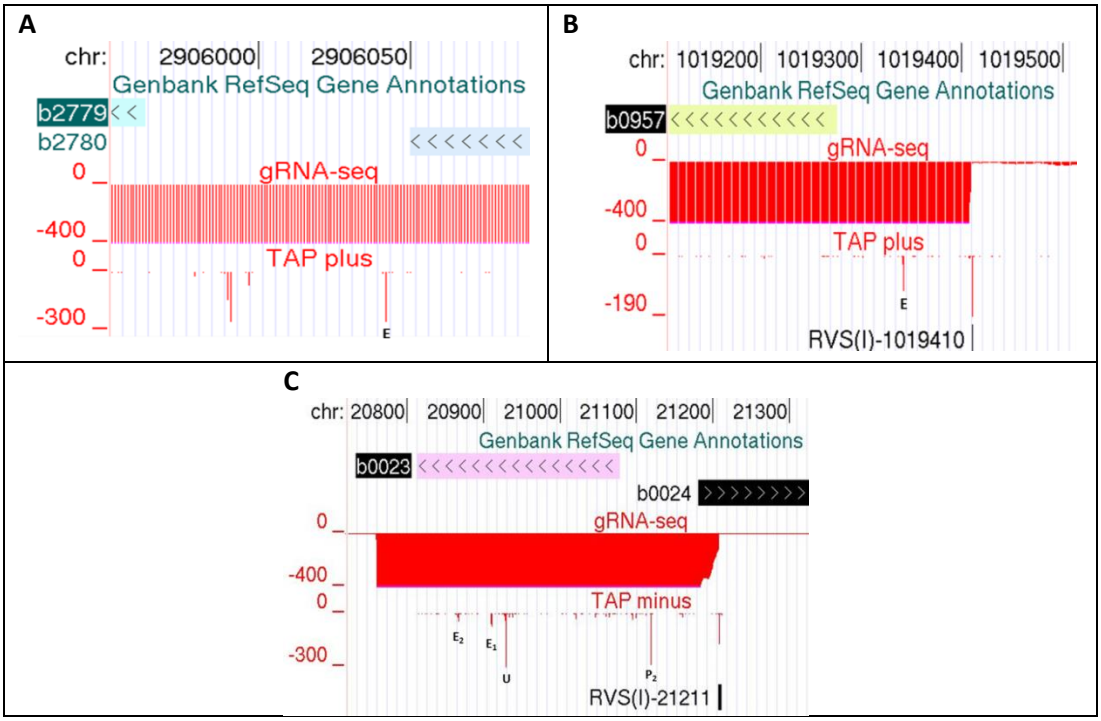

Figure 4.9 - Detection of known cleavage sites. This figure presents the differential and global sequencing profiles for three characterized genes. Panel A presents *eno* where the RNase E site at position -27 is marked with an E. Panel B presents *ompA,* the RNase E cleavage site at -66 nt is labelled E and the promoter RVS(I)-1019410. Panel C presents *rpsT* where the two promoters are labelled RVS(I)-21211 and P_2. A novel, previously unidentified cleavage site is labelled U whilst the RNase E cleavage sites occurring at positions +167 and +210 are labelled E_1 and E_2 respectively. All positions are in relation to the first nucleotide in the start codon. All labelling is as in Figure 4.3. As cleavage site reads are similar in TAP treated and TAP untreated samples only the track presenting the best peaks for both TSSs and cleavage sites are presented.

4.2.5 **Identification of potential sRNAs**

Manual inspection of the global RNA-seq data identified 181 small RNAs (Table S.2) of which 87 appear to have been described previously (Figure 4.10). Within this group, the ubiquitous bacterial sRNAs, 6S RNA, tmRNA, the RNA component of RNase P, and Signal Recognition Particle RNA (Lee et al., 1978; Vioque et al., 1988; Struck et al., 1990; Felden et al., 1997; Schneider et al., 2006; Chan et al., 2012) were all readily identifiable (Figure 4.11). Of the remaining 89, many had 5' ends associated with Class I or III promoters (TAP enrichment) indicating that this group was not dominated by quasi-stable processing or degradation intermediates. However, 69 were associated with REP (repetitive extragenic palindromic) sequences, which are found peppered throughout the genomes of bacteria. REP sequences have been shown to efficiently block the progression of 3' to 5' exonucleolytic attack (Stern et al., 1984; Newbury et al., 1987). Thus, small RNAs overlapping with a REP sequence were also filtered out. It is acknowledged that this may have excluded functional small RNAs that happen to utilise REP sequences to increase their stability. However, collaborators were concerned, perhaps overly, that others would simply interpret the presence of a REP as evidence for quasi-stable processing or degradation intermediates. Moreover, transcripts shorter than 70 nt were not included in our list of small RNAs. This was done to exclude 3' transcriptional terminators, which like REP sequences, form stem-loop structures that are less susceptible to 3' to 5' exonucleolytic attack (Andrade et al., 2009). Another reason for eliminating transcripts shorter than 70 nt was to eliminate artefacts caused by the misalignment of short reads. Where sequences are duplicated in the genome (e.g. ribosomal RNA operons), the alignment algorithms assign reads to them on a random basis. This can produce features which look like 'chimneys' of transcription in genome-wide transcription maps when a read from a highly transcribed region is assigned to a region that is not. After filtering out small RNAs that had been described previous or were associated with REP sequences, 20 small RNA remained. Examples of these novel small RNAs are shown in Figure 4.12.

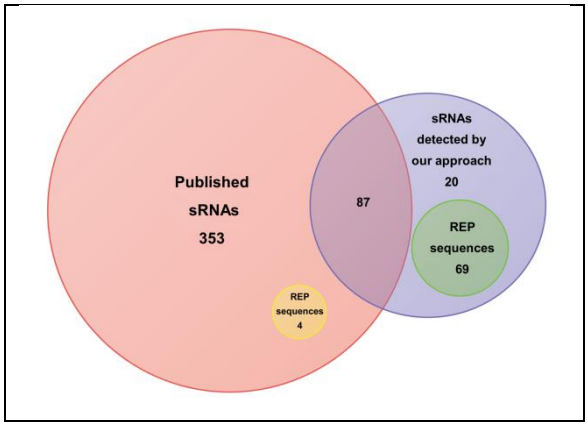

Figure 4.10 - Venn diagram presenting sRNAs identified in our study, previously identified sRNAs and REP sequences. This Venn diagram presents the 181 sRNAs detected by our sequencing approach (labelled "sRNAs detected by our approach"). It also presents the 444 sRNAs that have been previously described in the literature and have experimental evidence associated with them (labelled "Published sRNAs"). There is an overlap of 87 sRNAs between our data and the published data. Regarding overlap with REP sequences (labelled "REP sequences") our data contains 69 sRNAs with REP sequences whilst the published dataset contains only 4.

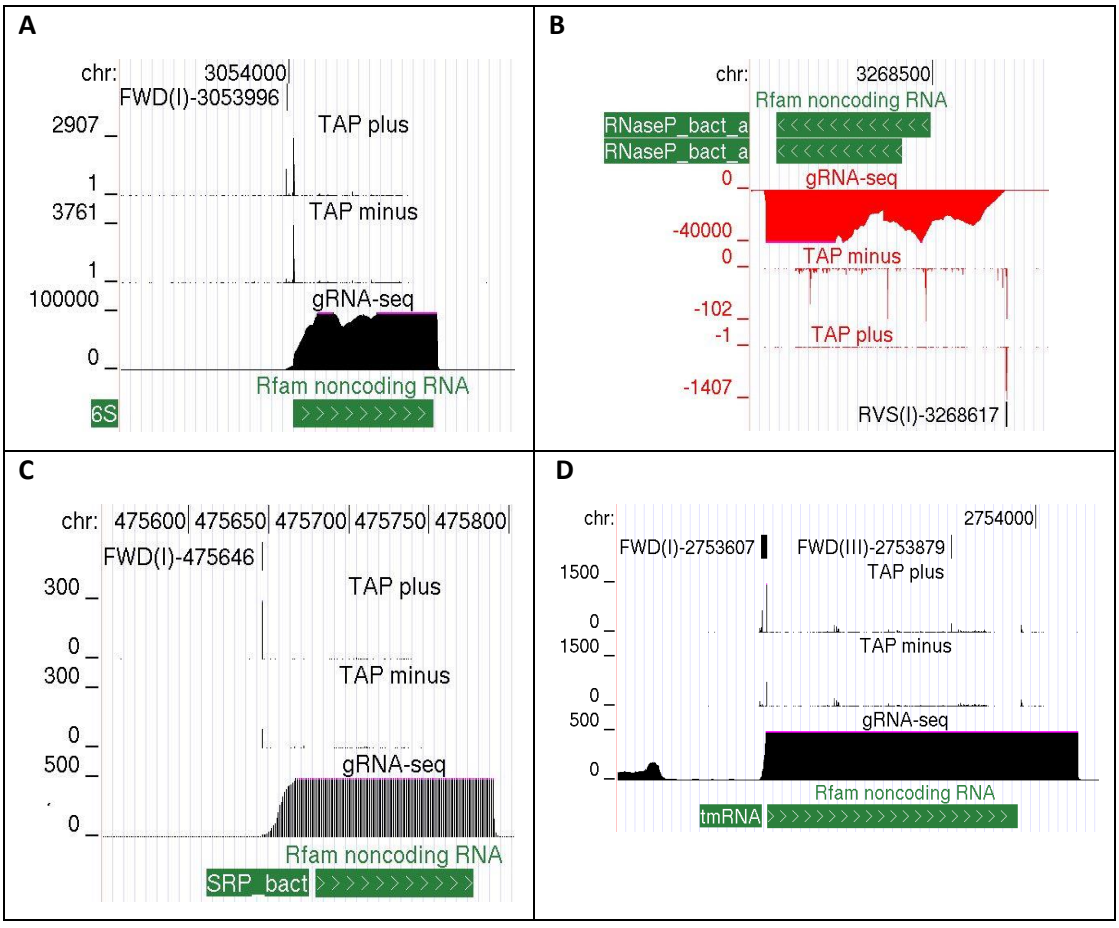

Figure 4.11 - Previously identified sRNAs supported by our sequencing approach. This figure presents the ubiquitous bacterial sRNAs, 6S RNA (Panel A), the RNA component of RNase P (M1RNA) (Panel B), Signal Recognition Particle RNA (Panel C) and tmRNA (Panel D). Labelling is as in Figure 4.3.

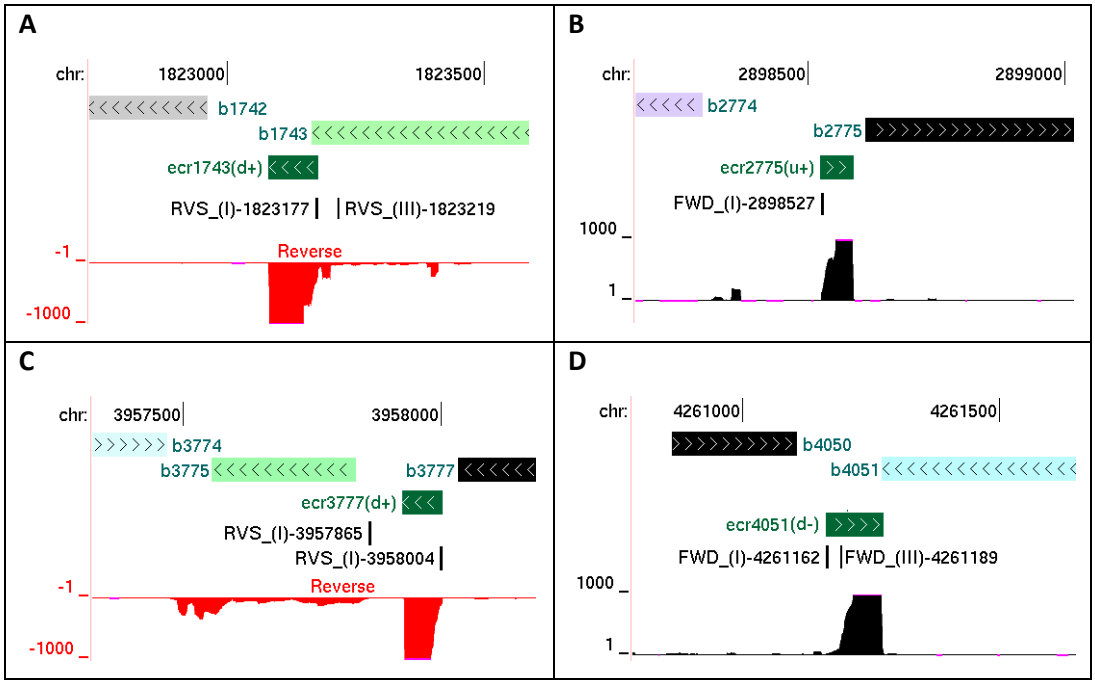

Figure 4.12 - Novel candidate sRNAs identified by our sequencing approach. This figure presents four novel sRNA candidates (panels A to D). Labelling in parentheses indicates whether the sRNA is upstream (u) or downstream (d) of the nearest protein-coding gene and whether it is on the same (+) or opposite (-) strand. The prefix 'ecr' is used for discrete RNAs of unknown function. The remaining of the labeling is as in Figure 4.3.

4.2.6 **Mapping of sites of cleavage by T170V *in vitro***

After the generation of reliable maps with regard to transcription and the positions of sites involved in RNA processing and degradation, the results of incubating enriched mRNA with the T170V mutant (Figure 3.9) were analysed next. Libraries were prepared from aliquots of RNA isolated from the *ΔrppH* strain that had been incubated with and without T170V for 20 min. This was done as described for dRNA-seq with the exclusion of the TAP treatment. For each of the libraries, the genome positions of the 5'-monophosphorylated ends were mapped and an estimate of the relative abundance of the corresponding fragments was obtained by counting the numbers of reads starting at each of the positions. Positions not associated with reads before and after incubation with enzyme were not included in the scatterplot. Where reads were obtained under only one condition (*i.e.* before or after incubation with T170V), the read for the other was given a nominal value of 1 (the lowest limit of detection). In this way, ~567 thousand ends were mapped.

The reads obtained before and after incubation with T170V were then compared using M (ratio)-A (intensity) scatterplots, where M = Log_2 (reads after/reads before incubation with T170V), and A = (log_2 reads before + log_2 reads after incubation)/2 (Figure 4.13). This

revealed a cone-shaped population of points that were distributed around an average M value of -0.43. This population corresponds to 5'-monophosphorylated ends that were present in the starting material and were not cleaved significantly. Above this population, there was a large 'cloud' of points of which 15 had M values greater than 10, 185 had M values greater than 8, 2667 had M values greater than 5 and 13551 had M values greater than 3. As the 5' ends associated with transcripts that are rapidly degraded following *in vitro* incubation with T170V show a decrease in reads, a noise envelope could not be determined. Thus a cut-off where M values greater than 5, which correspond to a fold increase of 32 (2 to the power 5), were considered as being significantly enriched following incubation with T170V *in vitro* was used. Thus, despite being deficient in 5'-monophosphate sensing, T170V appears to be able to cleave at a large number of sites in the *E. coli* transcriptome. The proportion of the cleavages mediated by T170V that are likely to be of functional significance was then estimated, by analysing the extent of the overlap between sites cleaved *in vitro* and those identified *in vivo*. This was done using the VLOOKUP function of Excel. For *in vitro* sites with M values of 3, 5 and 8, the percentage overlap with *in vivo* was 61.4, 72.1 and 75.7 % respectively. Thus, while RNA binding proteins block access *in vivo* to sites that were available to RNase E *in vitro*, it appears, nevertheless, that direct entry by RNase E may mediate a substantial number of cleavages detected *in vivo*. Moreover, this is likely to be an underestimate as detection of cleavages *in vitro* requires that the substrate is present at sufficient levels *in vivo*, against a background of active processing and degradation, and because some of the downstream species of RNase E cleavage might be degraded so rapidly that they cannot be detected *in vivo*. Examples of some top candidates where *in vivo* cleavage sites are reconstituted following *in vitro* incubation with T170V are presented in Figure 4.14.

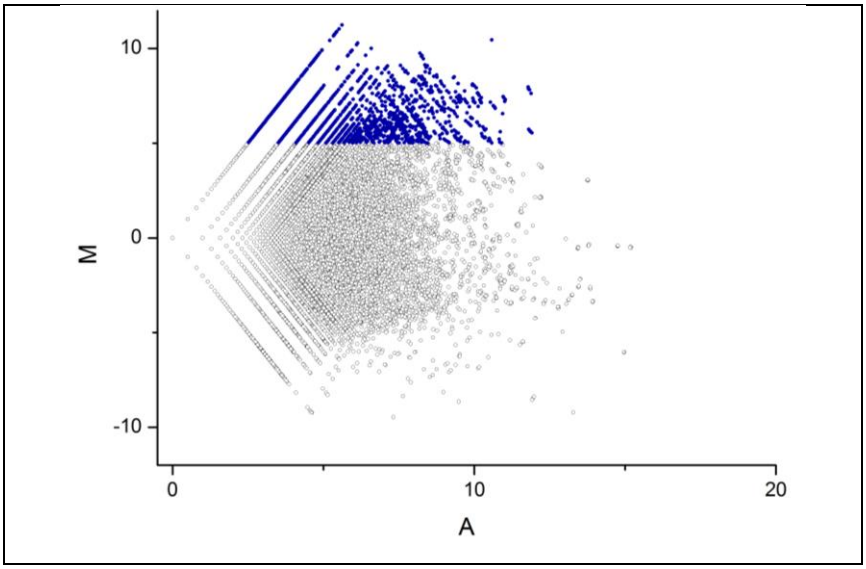

Figure 4.13 - Scatterplot analysis of RNA-seq data following total mRNA incubation with T170V. The values corresponding to the reads obtained before and after the incubation of enriched RNA with T170V for 20 min. The points coloured blue have M values ≥5.

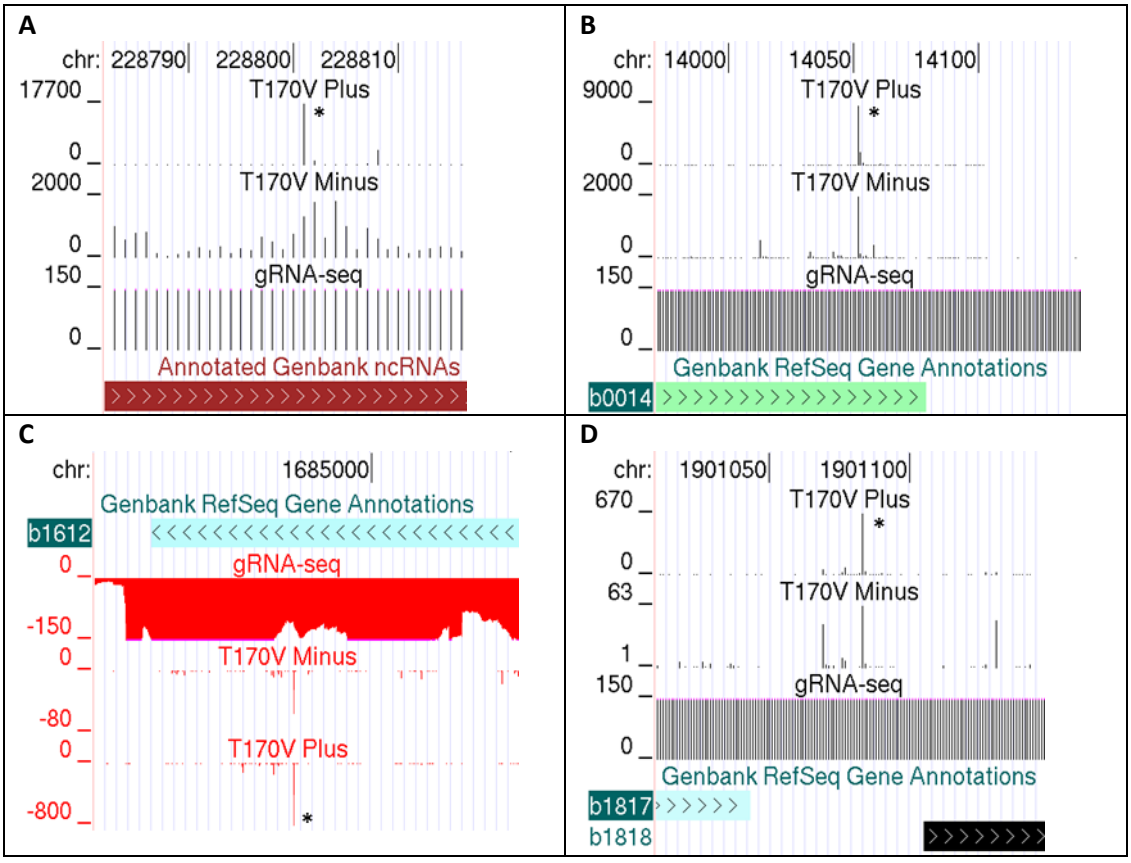

Figure 4.14 – *in vivo* cleavage sites reconstituted *in vitro* following incubation with T170V. This figure presents four top examples of cleavage sites (labeled *) which are present *in vivo* (T170V minus) and are reconstituted when total mRNA in incubated with T170V *in vitro* (T170V plus). The examples presented are 5S rRNA, *dnaK*, *fumA* and *manX* (Panels A, B, C and D respectively). Scales are different in the T170V minus and T170V plus tracks. Labelling is as in Figure 4.3

4.3 Discussion

This chapter describes the successful development of a pipeline that utilises RNA-seq to map the location of sites involved in RNA processing and decay (Figure 4.1) against the transcriptional landscape present under the same conditions and to investigate the contribution of direct entry cleavage by RNase E using the T170V mutant (Figure 4.13). The results of the latter suggested that one of the hallmark properties of RNase E, *i.e.* its ability to interact with 5'-monophosphorylated ends, is not required for efficient cleavage at a plethora of sites within the *E. coli* transcriptome (Figure 4.13). Consistent with this notion, a selection of 5'-triphosphorylated transcripts was found to be cleaved by the T170V mutant *in vitro* (see Chapter 5).

Previous RNA-seq analyses of sites involved in bacterial gene regulation have had a different focus, concentrating instead on the identification of transcription start sites. As described in the Introduction (Section 1.4), TSSs in previous studies were identified by removing the 5'-monophosphorylated products of RNA processing and degradation and then treating the remaining RNA with TAP to allow the cloning of fragments corresponding to the 5'-triphosphorylated ends of nascent transcripts (Figure 1.5). The approach described here, which took advantage of the ability of Illumina Solexa sequencing to generate ~10 million reads per run, also identified TSSs, but by virtue of the increase in the number of the corresponding reads that followed treatment with TAP. Sites involved in RNA processing and degradation, which show no such increase, provided a clear baseline for identifying TSSs (Figure 4.2) as well as sites cut by T170V *in vitro* (Figure 4.13). In a recent study of the *E. coli* transcriptome that did not include a step to differentiate TSSs, sites of processing as determined here, some of which are well characterised and documented (*e.g.* RNase P maturation of the 5' end of tRNA), were identified erroneously as TSSs (Cho et al., 2009). Moreover, the inclusion of a differential step is critical to identify transcription start sites that are positioned downstream of others that produce significant transcription. By widening the pipeline to include replicates and a statistical analysis that took into account the inherent noise in RNA-seq data, it was possible in a recent study of the transcriptional landscape of *Propionibacterium acnes*, a major contributor to wide-spread human disease, to identify not only nested promoters upstream of coding regions, but sites that have been attributed to pervasive transcription (Cho et al., 2009; Georg et al., 2009; Guell et al., 2009; Jacquier, 2009; Jager et al., 2009; Mendoza-Vargas et al., 2009b; Rasmussen et al., 2009; Toledo-Arana et al., 2009; Albrecht et al., 2010; Beaume et al.,

2010; Dornenburg et al., 2010; Filiatrault et al., 2010; Marguerat and Bahler, 2010; Martin et al., 2010; Sharma et al., 2010; Wurtzel et al., 2010; Lasa et al., 2011; Mitschke et al., 2011). The latter may be largely the consequence of the ability of RNA polymerases to initiate transcription from sub-optimal sequences (Lin et al., 2013b).

As a result of the encouraging results described here, the use of TAP or another RNA pyrophosphatase as a means of distinguishing nascent 5' ends is advisable. However, treatment with TEX may offer better discrimination of TSSs that correspond to transcripts that are efficiently 'decapped' by an RNA pyrophosphatase *in vivo* and as a consequence cannot be enriched significantly by TAP *in vitro*. In this scenario, TEX treatment would remove the majority of the species allowing those with nascent 5' triphosphorylated ends to be detected. In this study, 311 leading edges of transcription for *E. coli* were detected, and although not associated with obvious enrichment (Class II sites), the majority had 5' ends which were detectable by dRNA seq. Most of these TSSs were associated with extremely low A values (Figure 4.2). Furthermore, analysis of the corresponding 5' ends did not identify terminal stem-loops, which are known to block RNA pyrophosphatases (Celesnik et al., 2007). Considered together, these results suggest that 5' ends associated with leading edges of transcription were not identified by dRNA-seq as they could not compete with others during the PCR step. About 25% of the promoters recorded in RegulonDB (Salgado et al., 2006) that have been verified by transcript-specific mapping (e.g. nuclease protection or primer extension assays) and were associated in this study with obvious transcription extending downstream (data not shown) were assigned to Class II.

In addition, background transcription could explain at least a proportion of the TSSs in Class III, which are associated with TAP enrichment, but not an obvious step increase in transcription. However, verification of background transcription initiation will require a number of biological replicates and statistical analysis, as applied recently to *P. acnes* (Lin, 2013) as the sequencing runs for the analyses here presented were done once and not in duplicate. TSSs associated with alternative promoters nested downstream of ones that produce a substantial increase in transcription were also assigned to Class III. This represents about 15-20% of the promoters recorded in RegulonDB (Salgado et al., 2006) that have been verified by transcript-specific mapping (e.g. nuclease protection or primer extension assays) and were associated in this study with obvious transcription extending downstream (data not shown).

The mapping of TSSs could be further improved by analysing strains which are deficient in the RNA pyrophosphatase; this might not always be desirable given the resulting changes in gene expression and presumably cell physiology accompanying this deletion. The addition of a phosphorylation step would allow the identification of the cleavage sites of RNases that produce downstream products with a 5'-hydroxyl group. This is likely to be particularly relevant to studies of conditions suboptimal for growth under which such RNases, e.g. MazF (Vesper et al., 2011), are activated. It is advisable that RNA is fragmented after the addition of the 5' sequencing adaptor as this step made it possible to detect the 5' ends of large and small transcripts, and the use of the FRT methods of global RNA-seq, which does not require an amplification step prior to sequencing (Mamanova et al., 2010). The latter may be of more importance in the study of GC-rich organisms such as the streptomycetes as their transcriptomes tend to form more stable secondary structures, which can introduce strong bias in assays involving PCR (Mcdowell et al., 1998).

In relation to the 20 novel sRNAs identified, it is not believed that this group represents metastable decay intermediates, as most are associated with TSSs in Class I. Furthermore, as the expression of several sRNAs has been shown to be regulated by growth (Vogel et al., 2003) and the RNA-seq was performed on RNA extracted from cells grown exponentially, it is expected that not all sRNAs in the literature will be detected. As a result of this, the overlap between our data and the library of sRNAs previously identified under a vast array of conditions is not as prominent as expected.

New features within the transcriptome of *E. coli*, despite it being one of the most studied organisms, such as the 20 small RNAs that had been un-described previously (Table S.2 and Figure 4.10) and ribosomal RNA processing (Figure 4.7) that might produce a subpopulation of ribosomes during exponential growth that ensure the translation of the mRNAs, identified here, that are leaderless in their nascent form (Table 4.3) are discussed at the end of the Chapter 5.

Chapter 5

5 Confirmation of features within the transcriptional landscape of *E. coli.*

5.1 Introduction

Bacteria not only have to be able to respond rapidly to changes in nutrient availability, they have to be able to survive periods of nutrient deprivation and other conditions that are not conducive to growth. For example, *Mycobacterium tuberculosis*, which causes tuberculosis (TB), one of the deadliest diseases ever to afflict mankind (Cruz-Knight and Blake-Gumbs, 2013), can produce latent TB infections in which the bacterial cell is thought to exist in a non-replicating dormant state, from which it may later reactivate and cause disease (Sutherland et al., 1976; Manabe and Bishai, 2000; Lillebaek et al., 2002). Although much surrounding latent infection by TB remains elusive, the molecular mechanisms by which bacteria are able to initiate and maintain a non-replicating dormant state, which is called persistence, is beginning to emerge (Zahrt and Deretic, 2001; Betts et al., 2002; Singh et al., 2006). Major advances in our understanding of bacterial persistence has emerged from the study of *E. coli.* These advances include the description of persister cells, even during exponential growth, as a result of stochastic fluctuations in gene expression that lead to phenotypic heterogeneity among cells within a clonal population (Elowitz et al., 2002; Davidson and Surette, 2008; Nikolic et al., 2013). Furthermore, the key role that toxin-antitoxin (TA) systems play in mediating the persistent state across prokaryotes is also being elucidated (Wagner and Unoson, 2012; Schuster and Bertram, 2013; Butt et al., 2014).

The hallmarks of TA systems are a stable toxin that can inhibit an important cellular process, and an antitoxin that must be constitutively present in the same cell (Buts et al., 2005; Gerdes et al., 2005). Although initially discovered as a mechanism associated with certain plasmids, that ensures only cells that continue to carry the plasmids survive (Bernard and Couturier, 1991; Thisted et al., 1994a; Thisted et al., 1994b), TA systems are now known to be encoded widely in the chromosomes of bacteria, where they are reported primarily to down-regulate cellular processes in response to stress rather than induce cell death (Yamaguchi and Inouye, 2009). However, this may be an overly simplistic view. Recently it has been shown that MazF, an endoribonuclease that serves as the toxin

component of a stress-induced TA system in *E. coli*, while destroying the majority of cellular transcripts, ensures the continued translation of a subset by cleaving at a site immediately upstream or close to their AUG start codon, thereby generating leaderless mRNAs (Vesper et al., 2011). Moreover, these leaderless mRNAs are translated by 'specialised' ribosomes that are also generated by MazF, via removal of 43 nucleotides from the 3' terminus of 16S rRNA, which is within the decoding centre of 30S ribosomal subunits (Vesper et al., 2011). The mRNAs that continue to be translated by the modified translation machinery are not just associated with cell survival but also cell death. It is thought that the induction of MazF leads to the death of most of the population and the survival of a small sub-population that can then expand when conditions become conducive to growth (Amitai et al., 2009).

The ability of *E. coli* to adapt its physiology is now increasingly associated with the production of small RNAs. For example, the long-term survival of *E. coli* during stationary phase is enhanced by 6S RNA, which binds specifically to RNA polymerase containing sigma70 (Wassarman and Storz, 2000). This interaction represses transcription of the many housekeeping genes, thereby reducing the energy burden on the cell during stationary phase. Furthermore, sRNAs have been shown to block or enhance translation and mRNA degradation (see Section 1.3.3.1)

In the previous chapter, RNA-seq analysis of the transcriptome of *E. coli* cells that were growing exponentially in rich media identified five mRNA transcripts that, in their nascent form, appear to lack a 5' leader sequence long enough to mediate a Shine-Dalgarno (SD) interaction, which is key to canonical translation in this model organism. Moreover, the processing site at the 3' end of 16S ribosomal RNA, the same as that cleaved by MazF, was detected. This raised the possibility that in rapidly growing cultures a subset of specialised ribosomes, similar to those generated during stress, could facilitate translation of leaderless mRNA transcripts. The RNA-seq analysis also revealed over a hundred small RNAs that appear not to have been described previously. This suggested that regulation by small RNAs may have a more extensive role than is currently appreciated. In this chapter, the existence of leaderless mRNAs under non-stress conditions, MazF-like processing of 16S rRNA, which further work by others implicates RNase E, and novel sRNAs are investigated further using techniques complementary to RNA-seq.

5.2 Results

5.2.1 **Leaderless mRNA**

RNA ligase-mediated reverse transcription PCR (RLM-RT-PCR, (Kime et al., 2008a)) before and after treatment with TAP was used to confirm the position of the 5'-triphosphorylated ends of leaderless mRNAs identified in the previous chapter (Table 4.3). The mRNAs of four genes were selected for this analysis: *rhlB*, *pgpA*, *ftsT* (also known as *dicA*) and *ymfK*. The first two encode RhlB, which is a core component of the degradosome (Miczak et al., 1996; Py et al., 1996), and a non-essential phosphatidylglycerolphosphatase (Icho and Raetz, 1983; Funk et al., 1992), respectively. The second two encode repressors of genes within the Qin and e14 prophages (Keseler et al., 2013), respectively. RLM-RT-PCR involves the ligation of an RNA adaptor to the 5' ends of transcripts, which can only occur if the 5' end is monophosphorylated (Kime et al., 2008a). This was followed by reverse transcription of the RNA using random hexamers, and then the amplification of specific 5'-end fragment(s) using a primer that is specific for the transcript of interest and another specific for the RNA adaptor. The amplicons were then sequenced to identify the positions corresponding to the 5' end of the transcripts. Two different primers complementary to the RNA adaptor were used (RLM1 and RLM2, see Table 2.2). It was determined empirically if one or the other could facilitate better amplification when combined with gene-specific primers. The *pgpA* transcript was analysed using RLM1 and was predicted to produce an amplicon of 125 bp should it have a 5' end that matches the site identified by RNA-seq (Figure 4.6). The *rhlB*, *ftsT* and *ymfK* transcripts were amplified using RLM2 and were expected similarly to produce amplicons of 156, 300 and 335 bp, respectively. The PCR products of *ftsT* and *ymfK* were analysed by electrophoresis using agarose gels whilst those of *pgpA* and *rhlB* were analysed using polyacrylamide gels (Figure 5.1). To differentiate 5' ends that might normally be 'decapped' efficiently *in vivo*, RNA from an *ΔrppH* mutant as well as the congenic wild-type strain (BW25113) were analysed. Assays minus the ligation step were included as negative controls.

All of the reactions produced amplicons of the expected size. However, in the case of *rhlB*, a second amplicon of ~230 bp was identified (Figure 5.1, panel D), which initially suggested the production of a leadered as well as leaderless transcript. For all of the transcripts, the abundance of the amplicons increased substantially following treatment of RNA from the *ΔrppH* mutant with TAP. This is consistent with the corresponding 5' ends being the products of transcription initiation. Similar results were obtained using RNA from the wild-

type strain, with the exception that the abundance of the amplicons produced for the leaderless *ymfK* and *rhlB* transcript did not increase following TAP treatment. This suggested that the *ymfK* and *rhlB* transcripts might be more efficiently decapped by RppH than the other transcripts under these conditions. Had a quantitative PCR been used to analyse the ligation product (Vanguilder et al., 2008), it might have been possible to discriminate the effects of TAP treatment. This was deemed unnecessary as the results of the RLM-RT-PCR analysis of RNA from the *ΔrppH* mutant agreed broadly with those of RNA-seq.

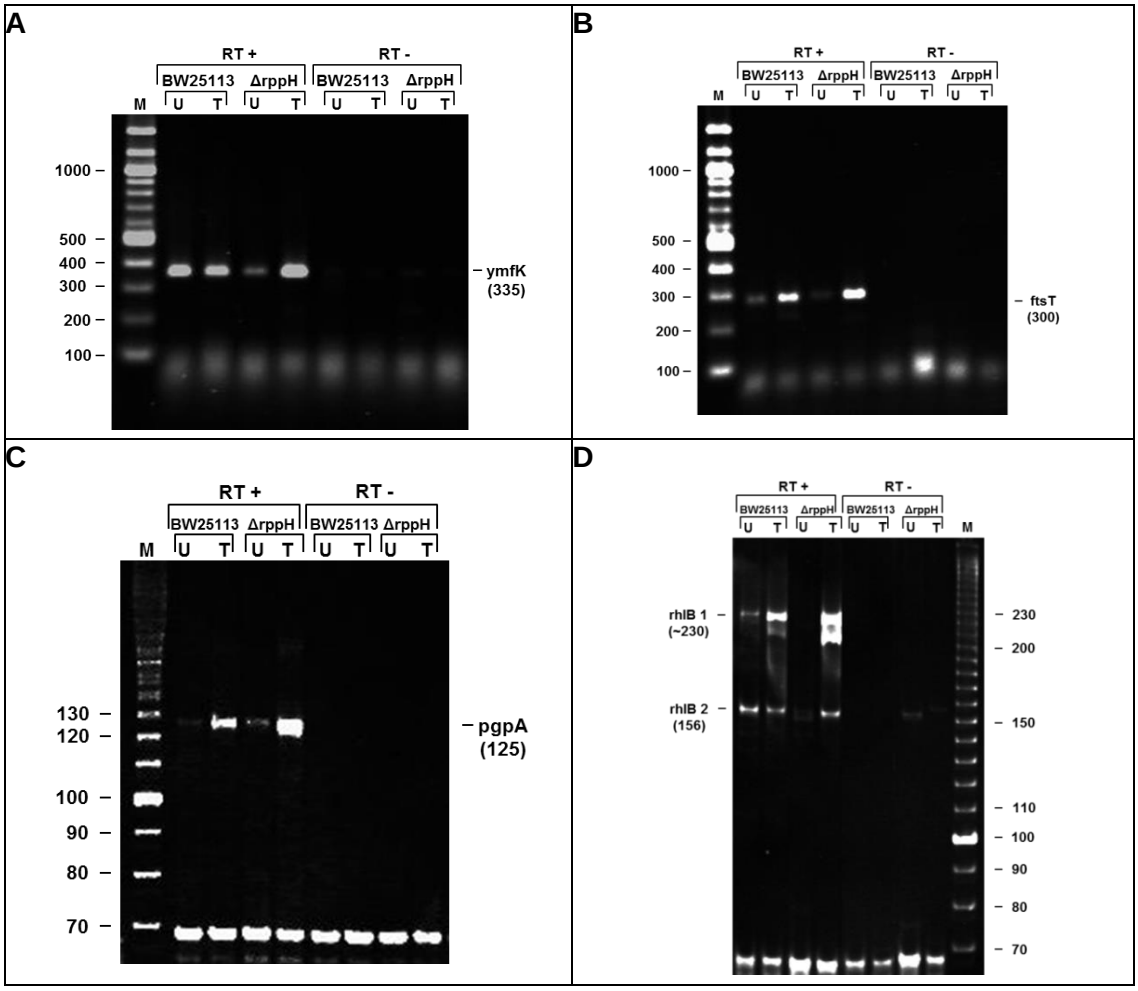

Figure 5.1 - Gel electrophoresis analysis of the products of RLM-RT-PCR Panels A, B, C and D correspond to the *ymfK*, *ftsT*, *pgpA* and *rhlB* transcripts, respectively. RNA extracted from BW25113 and ΔrppH (labelled accordingly) were either treated with TAP (labelled T) or left untreated (labelled U) prior to RLM-RT-PCR. Reverse transcription reaction samples were then PCR amplified and are labelled RT+ whilst their negative controls, which were not reverse transcribed, are labelled RT-. The expected products are labelled *ymfK*, *ftsT*, *pgpA*, *rhlB* 2 respectively; their sizes are indicated in brackets. For *rhlB* an unexpected product is labelled *rhlB* 1. The source of markers (labelled M) were the GeneRuler 100 bp Plus DNA Ladder (Fermentas) for panels A and B and the 10 bp DNA Ladder (Invitrogen) for panels C and D. The samples were electrophoresed on a 2% agarose gel (panels A and B) and 10% acrylamide (panels C and D).

The finding that *rhlB* produced a second amplicon of ~230 bp (Figure 5.1, panel D) was followed up by repeating the RLM-RT-PCR analysis, purifying the PCR products (without separating on the basis of size) and cutting with the restriction enzyme *MseI*. This enzyme cuts 87 bp upstream from the 3' side of the binding site of the *rhlB*-specific primer (Figure 5.2). Following digestion with *MseI*, two products consistent with the migration of this 87 bp 3' side fragment (P_2) and its 69 bp 5' side counterpart (P_3) were detected; together they add up to the 156 bp fragment corresponding to the leaderless transcript (Figure 4.6). A faint product of ~230 bp was also detected (rhlB1), consistent with transcription also starting ~70 bp upstream of the start codon. However, the analysis was complicated by the presence of a species of ~70 bp of unknown identity, but probably a primer dimer from the RLM-RT-PCR reaction. Therefore, the *rhlB* amplicons of 156 and ~230 bp were gel purified, cloned and sequenced (Table S.3). The analysis of two independent clones of the smaller transcript confirmed that the 5' end of this amplicon corresponded exactly to the +1 position of *rhlB* (relative to the start codon), while the analysis of three independent clones of the larger amplicon positioned the second TSS at +70. The TSSs identified in this analysis are indicated in relation to gRNA-seq and d-RNA seq data (Figure 5.3).

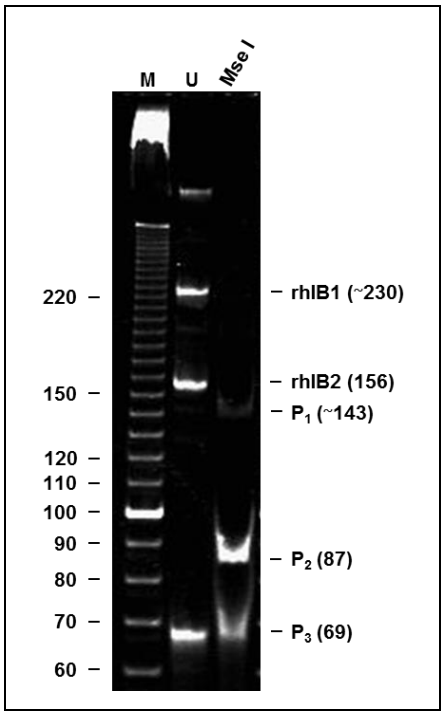

Figure 5.2 – Restriction enzyme digest of the rhlB RLM-RT-PCR products. RLM-RT-PCR products for *rhlB* before and after cleavage with restriction enzyme *MseI* (labelled U and *MseI* respectively). RLM-RT-PCR products prior to incubation with *MseI* are labelled rhlB 1 and rhlB 2. Following restriction digest, the products are labelled P_1, P_2 and P_3; their sizes are indicated in brackets.

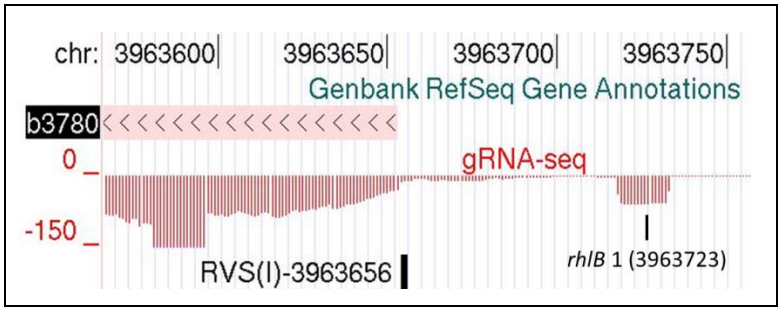

Figure 5.3 - TSSs identified for *rhlB* following RLM-RT-PCR and cloning. RLM-RT-PCR results together with sequencing and restriction mapping indicate that *rhlB* possesses two TSSs, one of which is leaderless and another one that has a leader sequence of 70 nt. The mapping of these TSSs onto the genome browser view for *rhlB* are presented here. Labelling is as in Figure 4.3.

5.2.2 **Specialized ribosome-like processing of 16S rRNA**

As described in the Introduction to this chapter, the generation of specialized ribosomes in which the 43 nt at the 3' end of the 16S rRNA containing the anti-Shine-Dalgarno sequence is removed via endoribonucleolytic cleavage by MazF has been recently described (Vesper et al., 2011). Cleavage at this site was also identified by RNA sequencing (Section 4.2.2), which is specific for 5'-monophosphorylated ends, not the 5'-hydroxylated ends generated by MazF and other mRNases associated with TA systems (Vesper et al., 2011; Cook et al., 2013). To allow the basis of cleavage at the +43 site (relative to the 3' end of mature 16S rRNA) to be studied specifically, RLM-RT-PCR was employed (Figure 5.4). However, the protocol was modified to allow good coverage of short RNA fragments by reverse transcription. The template produced by random priming could not be amplified using a gene-specific primer that binds to a site within the 43-nt segment but it could when reverse transcription was primed from the very 3' end by adding a poly(A) tail and then using $d(T)_{10}(V)$ as a primer. The expected size of the amplicon corresponding to cleavage at the +43 site was 87 bp. Longer and shorter amplicons were also detected: this was not unexpected as cleavage sites upstream and downstream of the +43 position were detected by RNA-seq. The analysis included RNA isolated from a congenic *ΔmazF* strain during stationary phase as well as exponential growth. RNA was also treated with TAP and polynucleotide kinase, the latter to allow detection of 5'-hydroxylated ends.

For the sample isolated from the wild-type strain during exponential growth, an amplicon of 87 bp was detected, albeit of low abundance, without treatment with TAP or PNK. This is consistent with the RNA-seq data. In the PNK-treated sample, the abundance of this amplicon was substantially higher indicating that much of the cleavage at this site

generates downstream product with a 5'-hydroxyl group. In other words, the RNA-seq approach only detected a minority of the cleavage at this site. The abundance of the 87-bp amplicon was also higher in the TAP-treated sample, although not nearly to the same extent as the PNK-treated sample. This suggests that the +43 site might also be associated with some transcription initiation. Analysis of the RNA from the *ΔmazF* strain during exponential growth produced a similar pattern to that of the wild-type strain, but with one clear exception: the abundance of the amplicon was *not* higher in the PNK-treated sample. This suggests that MazF normally cleaves at the +43 site in cultures in which the bulk of the cells are growing exponentially. For RNA from the *ΔmazF* strain during stationary phase, the detection of the 87-bp amplicon was dependent on PNK treatment. This suggests that an RNase other than MazF can cleave at the +43 site during stationary phase. The identity of the 87 bp amplicon was confirmed by cutting the products of the RLM-RT-PCR (untreated $\Delta mazF_{ex}$) with *BstEII*, which was predicted to produce fragments of 62 and 25 bp (Figure 5.5). Fragments with mobility consistent with the sizes of these products were detected. The significantly higher abundance of the 25 bp fragment reflects that it is also common to amplicons longer than the 87 bp (Figure 5.5).

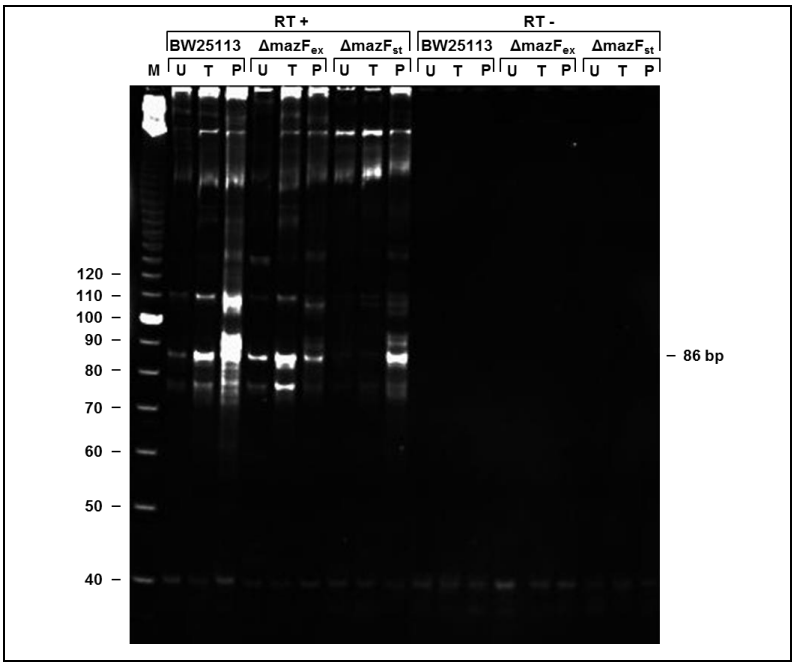

Figure 5.4 – RLM-RT-PCR analysis of cleavage at the -43 site of 16S rRNA. This figure shows the products generated following RLM-RT-PCR to detect cleavage at the -43 site of 16S rRNA. RNA extracted from BW25113, *ΔmazF* at exponential growth (*Δmaz*F$_{ex}$) and *ΔmazF* at stationary growth ($\Delta mazF_{st}$) were untreated (labelled U) or treated with TAP (labelled T) or PNK (labelled P). The expected 86 bp product is labelled on the right of the panel. Reverse transcription samples were PCR amplified and are labelled RT+ whilst their negative controls are labelled RT-. The samples were electrophoresed on a 10% polyacrylamide gel.

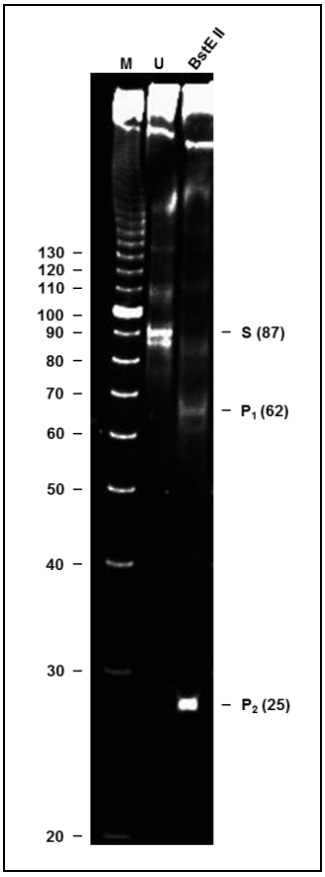

Figure 5.5 - Restriction assay of the 16S 3' end cleavage product. The 87 bp RLM-RT-PCR product generated by the cleavage at the 3' end of 16S rRNA before and after cleavage with restriction enzyme *BstEII* (labelled U and *BstEII* respectively). The RLM-RT-PCR product prior to incubation with *BstEII* is labelled S. Following restriction digest, the products are labelled P_1 and P_2. Sizes are indicated in brackets. The gel is 10% polyacrylamide.

5.2.3 **Confirmation of sRNAs by northern blotting**

To verify the presence of small RNAs not described previously, at least to our knowledge, six of these sRNAs were selected randomly and probed using northern blotting under stringent conditions. RNA used for northern blot was separate from that used for RNA seq analysis. Moreover, as the expression of several sRNAs has been shown to be regulated by growth (Vogel et al., 2003), we isolated RNA from cells growing exponentially in M9 glucose as well as LB. As a control, we included the analysis of AgrB sRNA, which is an RNA regulator of the SOS-related DinQ protein (Weel-Sneve et al., 2013). Positive results were obtained for five of sRNAs and the AgrB control (Figure 5.6). The actual abundance of all of the detected sRNAs, apart from the AgrB control, was dependent on the growth media. The small RNA that could not be detected by northern blotting is identified as ecr0770(d-). Overall, the vast majority of sRNAs tested by northern blotting was detected. The northern blotting experiments were done in conjunction with collaborators, Olatz Ruiz Larrabeiti and

Vladimir Kaberdin (University of the Basque Country, Faculty of Science and Technology, Spain).

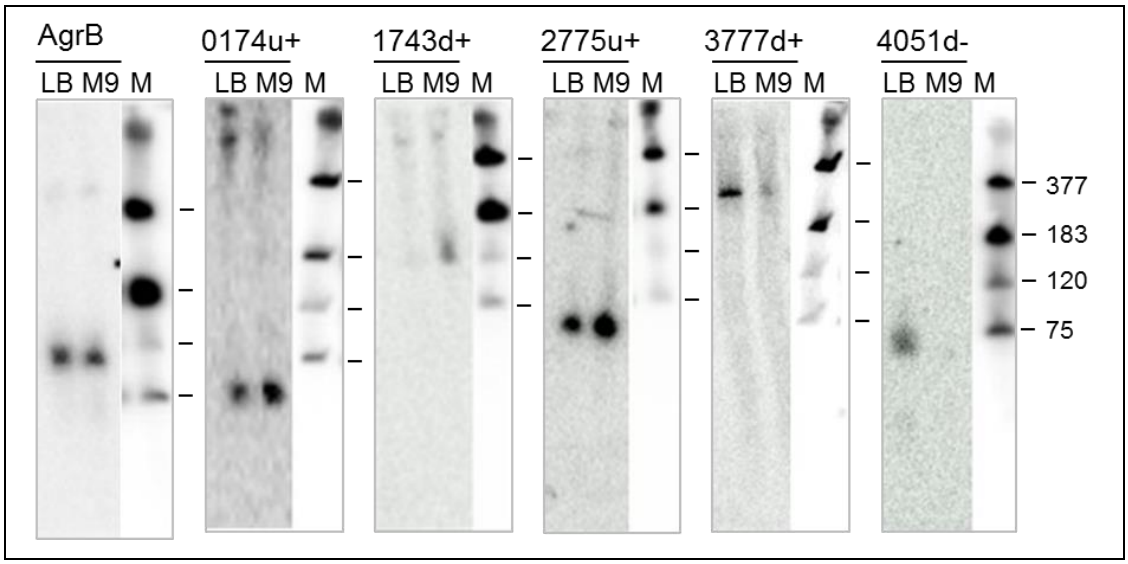

Figure 5.6 – sRNA detection with radiolabelled probes. Northern blot results obtained by our collaborators. 30 µg of total RNA was extracted from MG1655, grown in LB and M9 minimal media, and was blotted onto 6 nitrocellulose strips, each containing 5 µg total RNA. Each strip was then hybridized with 5 different radio-labelled oligonucleotide probes. The first strip was hybridized with an oligoprobe for AgrB and acted as a positive control. Following the control, from left to right, probes to detect sRNAs ecr0174(u+), ecr3777(d+), ecr4051(u+) and ecr1743(d+) were used and are labeled accordingly (for probe sequences see Table S.2). Markers are presented on the right and labelled M.

5.2.4 **Confirmation of novel cleavage sites present *in vivo* reconstituted following incubation of total mRNA with T170V *in vitro.***

Using the RNA-seq approach described in the previous chapter, it was shown that the T170V mutant of RNase E is able to cleave at many sites within the *E. coli* transcriptome (Section 4.2.6). To investigate further whether these sites are cleaved efficiently by RNase E in the absence of binding to a 5'-monophosphorylated end, transcripts for a selection of candidates were synthesized *in vitro* and incubated with T170V. The selection was based on manual inspection of the RNA-seq data using a genome browser. Three mRNAs were selected *dnaK, fumA* and *manX*, as they were of average abundance (400-1500 reads, gRNA-seq data) and T170V cleavage occurred at sites mapped as being involved in processing or degradation *in vivo*. 5S RNA was included as a curiosity, as it is widely known as a product, not a substrate, of RNase E cleavage (Deutscher, 2009). The tRNA precursors, *argX-hisR-leuT-proM*, *glyW-cysT-leuZ* and *pheU*, were also included as a recent study by others had indicated that their processing *in vivo* was not dependent on the 5' sensor of RNase E (Garrey and Mackie, 2011).

Using conditions described previously (Section 2.5.7), cleavage of 5'-triphosphorylated 5S RNA and *dnaK* mRNA by T170V and NTH-RNase E was readily detected (Figure 5.7, panels A & B). Based on the RNA-seq data, product sizes should be 120/55 and 214/238 nt, respectively. 5S RNA produced a cleavage profile that is consistent with a single cleavage site with product sizes being close to those expected. *dnaK* mRNA also shows significant cleavage but the number of products generated shows that an additional site which is not available *in vivo* is made available *in vitro.* The efficiency of these cleavages, as determined from the half-lives of the substrates, was similar to that reported previously for *cspA* mRNA (see Figure 3.8). These results confirmed that at least some, if not many of the sites identified by RNA-seq sequencing are cleaved efficiently by RNase E in the absence of binding to a 5'-monophosphorylated end, *i.e.* direct entry may be a major route by which processing and degradation occurs. Cleavage of *manX* and *fumA* (Figure 5.7, panels C and D) was also detected; however, it was much less efficient. Possibly, the corresponding sites were not within the conformational context *in vitro* that allowed cleavage *in vivo*. In support of this notion, efficient cleavage of *cspA* mRNA can be reduced substantially by certain truncations from the 5' end, but then restored by further truncation (Kime et al., 2010). A limitation of the RNA-seq approach is that while sites of cleavage can be identified readily, the 5' and 3' boundaries of the corresponding substrate(s) is not always obvious.

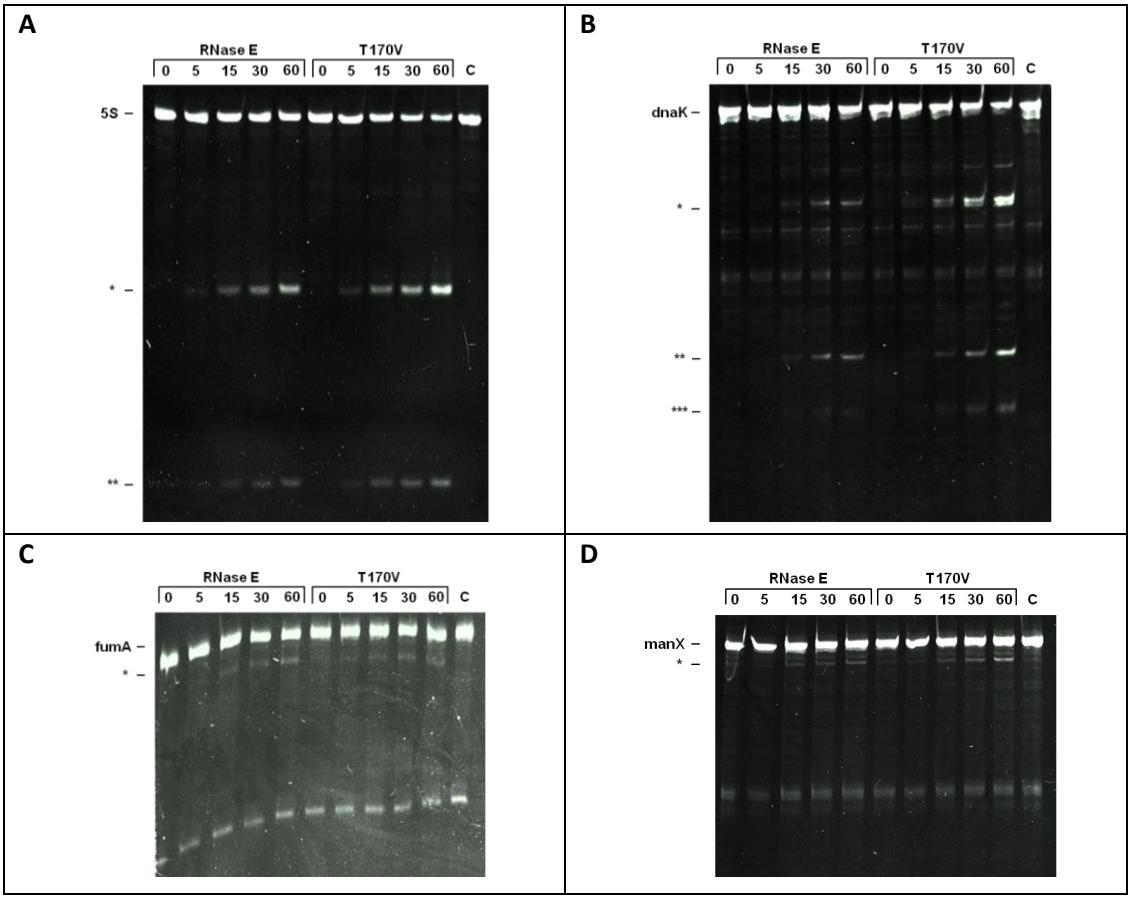

Figure 5.7 – Assay of RNase E direct entry candidates identified via *in vitro* reconstitution. The four RNase E direct entry mRNA candidates were incubated with NTH-RNase E and T170V. The reactions contained 0.64 µM of 5S RNA (A), *dnaK* mRNA (B) *fumA* mRNA (C) or *manX* mRNA (D) and 25 nM of wild-type NTH-RNase E or the T170V mutant. Samples were taken after 0, 5, 15, 30 and 60 min. A control of each mRNA, incubated alongside the cleavage assays without enzyme for 60 min, was also present (labelled C in all cases). Substrates are labelled 5S, *dnaK*, *fumA* and *manX*; cleavage products for each substrate are labelled *, ** or *** depending on the number of cleavages occurring within each transcript. Substrates for 5S RNA, *dnaK* mRNA, *fumA* mRNA and *manX* mRNA were generated by PCR and *in vitro* transcription with primer combinations 5SF/5SR, dnaKF/dnaKR, fumAF/fumAR and manXF/manXR, respectively (Table 2.2). All samples were analysed under denaturing conditions using a 7 M urea, 7% (w/v) polyacrylamide (19:1) gel.

Incubation of the *argX-hisR-leuT-proM* precursor with NTH-RNase E and T170V produced a complex pattern of cleavage products that were similar, although not identical (Figure 5.8, panel A). Reaction conditions were consistent with those previously used (Section 2.5.7). The result described here suggested that the majority of the cleavages within the *argX-hisR-leuT-proM* precursor were independent of 5'-monophosphate sensing. Consistent with this notion, subsequent experiments showed that the pattern and rate of cleavage was not altered dramatically when the 5' end was treated to produce a monophosphate group (Kime et al., 2014).

Similar results to those described above were obtained for the *glyW-cysT-leuZ* precursor (Figure 5.8, panel B). Although initial cleavage does not appear to be dependent on 5' sensing, an intermediate (labelled 2 in panel B) accumulated when incubated with T170V, but not NTH-RNase E. This suggests that, normally, this intermediate would be subsequently cleaved via a 5'-monophosphate-dependent step. Similar intermediates accumulated with the *argX-hisR-leuT-proM* precursor (labelled 2 and 3 in panel A). The *pheU* transcript was cleaved by T170V (Figure 5.8, panel C), but not as efficiently as the other precursors. This indicates that not all initial cleavages are necessarily mediated efficiently by direct entry.

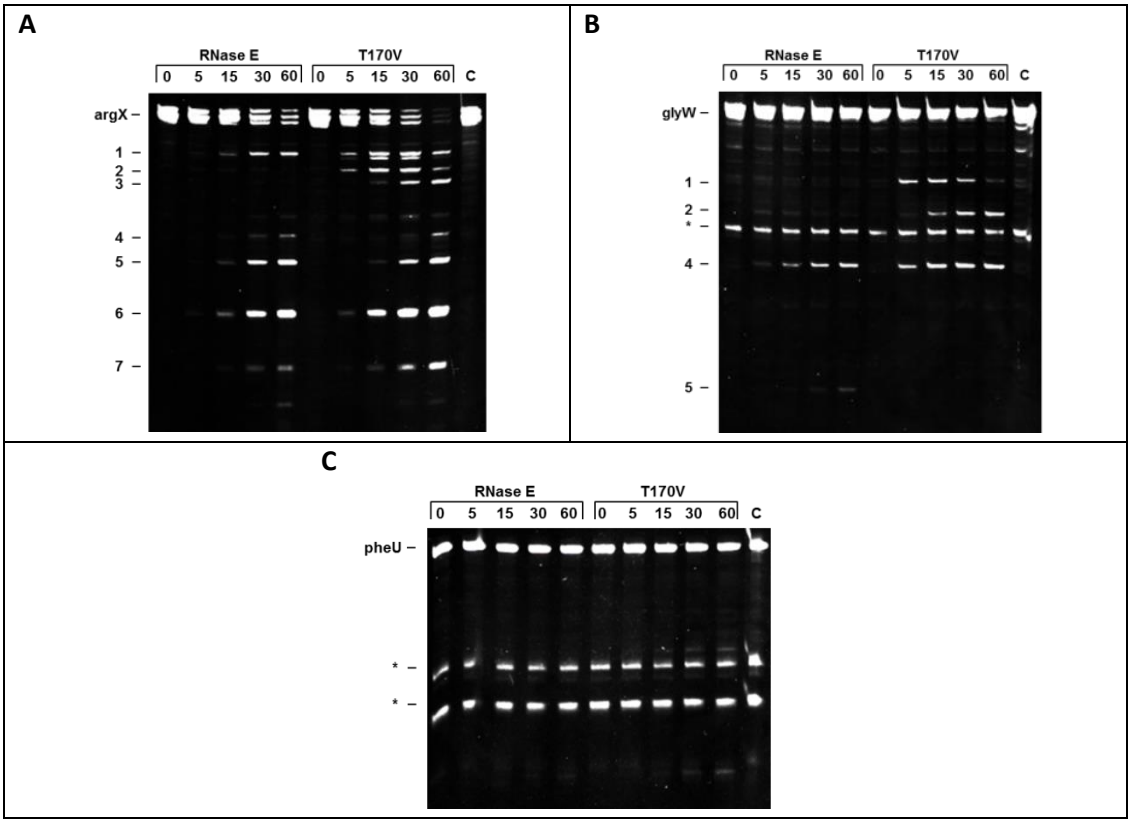

Figure 5.8 – *In vitro* RNase E cleavage assays of tRNA precursors. tRNA operons *argX-hisR-leuT-proM* and *glyW-cysT-leuZ* alongside the *pheU* transcript were incubated with NTH-RNase E and T170V. The reactions contained 0.64 µM substrate *argX-hisR-leuT-proM* RNA, *glyW-cysT-leuZ* RNA and *pheU* RNA (Panels A, B and C respectively) and 25 nM of wild-type NTH-RNase E or the T170V mutant. Samples were taken after 0, 5, 15, 30 and 60 min. A control of each mRNA, incubated alongside the cleavage assays without enzyme for 60 min, was also present (labelled C in all cases). Substrates are labelled according to the first gene of the operon i.e. *argX, glyW* and *pheU*; cleavage products for each substrate are labelled with numbers depending on the number of cleavages occurring within each transcript. Shorter transcripts which were a result of the *in vitro* transcription and not of the cleavage assay are indicated with an asterisk. Substrates for the *argX, glyW* and *pheU* precursors were generated via PCR and *in vitro* transcription with primer combinations argXF/argXR, glyWF/glyWR and pheUF/pheUR, respectively (Table 2.2). All samples were analysed under denaturing conditions using a 7 M urea, 7% (w/v) polyacrylamide (19:1) gel.

After undertaking the comparisons described above, it was confirmed that the cleavage products generated by the N-terminal half of RNase E are the same as those produced by the RNA degradosome under conditions in which PNPase was not active (Louise Kime, unpubl.result). Others have also found that N-terminal half of RNase E is sufficient to direct all of the cleavages produced by the degradosome (Mackie, 2013a). The analysis of 5' sensing described here was based on the NTH-RNase E rather than the degradosome, as our laboratory and its collaborators have so far been unable to purify degradosome preparations that incorporate RNase E with mutations in its 5' sensor.

5.3 Discussion

In this chapter, the existence of leaderless mRNAs as well as the processing of 16S rRNA implicated in generating ribosomes capable of translating such mRNAs as well as the presence of small RNAs that were not previously described was demonstrated using RLM-RT-PCR (Figure 5.1 and Figure 5.4) and northern blotting (Figure 5.6). Results concluded that the RNA-seq approach described in the previous chapter provides reliable maps of the transcriptional landscape that can be further annotated. With regard to the latter, preliminary analysis of *in vitro* cleavage sites mapped for T170V (Figure 5.7) confirmed that many are likely to be sites of cleavage that are mediated by direct entry rather than residual 5'-sensing activity.

In addition to confirming the existence of leaderless mRNAs and 16S rRNA processing that is implicated in generating specialised ribosomes capable of translating such mRNAs, the RLM-RT-PCR analysis suggests that MazF also has a role in producing specialised ribosomes during exponential growth (Figure 5.4). Whether this is true for all cells within the culture or only a sub-population, and whether the level of processing determined is biologically relevant remains to be determined. The sensitivity of RNA-seq is such that potentially rare events, including pervasive transcription, can be detected. An assessment of biological relevance will require the integration of genetic approaches. For example, by engineering 5'-leadered segments into the mRNAs of certain repressors of prophage (Qin, Rac and e14) and assaying for the release of phage into the culture, it would be possible to determine whether normally their leaderless status provides a simple mechanism by which their translation, and thus repression of the lytic cycle, is maintained during conditions of stress when increasing numbers of specialised ribosomes are produced. A detailed assessment of a *ΔrhlB* might shed light on why the mRNA of a degradosome component is leaderless. One possibility is that it is required for the remodelling of ribosomes following processing by

MazF and perhaps other ribonucleases. If this notion is correct, it might be expected that extracts from *ΔrhlB* would be less efficient at translating leaderless mRNAs *in vitro*. A role of RNA helicases in ribosome biogenesis is known (Jain, 2008).

It would also be interesting to determine whether the RppH 5' pyrophosphohydrolase has a role in the translation of leaderless mRNAs. Recent work has shown that the 5'-phosphorylation status of the mRNA of the cI repressor of bacteriophage lambda might influence the efficiency of translation *in vivo* as well as the binding of ribosomes *in vitro* with a 5'-monophosphate group stimulating binding of the ribosome and thus translation (Giliberti et al., 2012). As of yet, no one else seems to have realised that 5' pyrophosphate removal might influence translation as well as cleavage by RNase E. If this is the case, it might be expected that more phages will be released into the cultures of *ΔrppH* mutants, as a result of the diminished repression of lytic genes. It would also be interesting to investigate if any proportion of the leaderless mRNAs generated by MazF during stationary phase have 5'-monophosphorylated ends. It might be that the 5'-hydroxylated, leaderless mRNAs generated by MazF are translated sufficiently to allow the expression of cell survival/death genes (Giliberti et al., 2012). On the other hand, it is not inconceivable they might be 5' monophosphorylated by a kinase to enhance their translation. After all, the polynucleotide kinase used in recombinant DNA technology, is encoded by a bacteriophage that infects *E. coli* (Cameron et al., 1978; Uzan, 2009).

Genetic approaches are now also required to assess the function of the growing number of sRNAs currently being identified in bacteria (Figure 4.10). However, while it is relatively straightforward to knock out the genes of sRNAs that do not overlap with other genes, it is difficult, and more often than not impossible, to predict how they might function.

Chapter 6

6 Further work and concluding remarks

By using two sequencing technologies (dRNA-seq and gRNA-seq), the generation of nucleotide resolution maps of both the primary and secondary transcriptomes of *E. coli* was achieved. From these maps, sites of transcriptional initiation, stable RNAs, small non-coding RNAs, leaderless mRNAs, specialized ribosome generation and sites of processing and degradation were identified. From this repertoire of features, we detected a reasonable overlap with previously identified sites in addition to a large number of novel ones. From the data here presented a number of previously undiscovered leaderless mRNAs and the identification of sites involved in the formation of specialized ribosomes under conditions which both match and diverge from those known to produce such specialized machinery has been further established (Chapter 4). These features and the sequencing approach used were finally further validated experimentally (Chapter 5).

There is great interest in the deciphering of the biochemical requirements for recognition and action by the repertoire of RNases and other members of the RNA degradation machinery present in *E. coli* and how they act synergistically together. We decided to characterise RNase E because, in addition to having many of its substrates being well characterised (examples presented in Chapter 3, Chapter 4 and Chapter 5), it is a central component of the degradosome (Miczak et al., 1996; Taghbalout and Rothfield, 2007) and is essential for cell viability (Apirion and Lassar, 1978; Ono and Kuwano, 1979); all features which made it a suitable candidate for characterisation. We initially wanted to characterise the contribution of the 5' monophosphate sensing pocket towards stimulation of cleavage during RNase E mediated processing and decay. Total RNA extracted from the wild-type strain BW25113 was incubated *in vitro* either in the absence or presence of the 5' monophosphate sensing deficient mutant NTH-RNase E T170V (Jourdan and Mcdowall, 2008), which was followed by dRNA-seq analysis. By comparing these two samples, it was possible to detect cleavage via direct entry and the evidence indicates that it is widespread within the transcriptome (Chapter 4). Furthermore, cleavage sites that were detected *in vivo* and enriched following *in vitro* T170V incubation were then reconstituted *in vitro* (Chapter 4). These results, in conjunction with those obtained by further analysis of tRNA precursor cleavage (Chapter 5) strongly indicate that direct entry is a major pathway for

RNA degradation and processing in *E. coli*, contrary to what is currently understood (Mackie, 2013a).

Indeed, following results presented in Chapter 4 and Chapter 5, a colleague in the laboratory, Louise Kime, took the lead in exploring substrate requirements for tRNA processing (Kime et al., 2014). She was drawn to study these substrates not only because their processing represents one of the main activities of RNase E (Li and Deutscher, 2002; Ow and Kushner, 2002) in *E. coli* and other bacteria (Li et al., 2005), but because the localised folding that produces tRNAs limits the formation of alternative secondary structures within the precursor (and derivatives) that can complicate the analysis of RNA: protein interactions. She focussed on the processing of the polycistronic *argX-hisR-leuT-proM* precursor, as it has been the subject of *in vivo* studies by others (Li and Deutscher, 2002; Ow and Kushner, 2002), including a recent study that concluded its processing was not dependent on the 5' sensor of RNase E (Garrey and Mackie, 2011). The study confirmed that direct entry is central to the processing of tRNA in *E. coli* and provided the first biochemical evidence for natural transcripts that direct entry is indeed mediated by specific unpaired RNA regions that are adjacent to, but not contiguous with, segments cleaved by RNase E. In addition, the study showed that direct entry at a site on the 5' side of the precursor triggers a series of 5'-monophosphate-dependent cleavages. Consistent with a major role for direct entry in tRNA processing, it was also shown by Justin Clarke in the group, contrary to a report by others (Garrey et al., 2009), that a 5'-monophosphate is not required to 'activate' the catalytic step, *i.e.* efficient cleavage does not have an absolute requirement for access to a 5'-monophosphorylated end (Kime et al., 2014). The dissection of the requirements for the cleavage of the *argX-hisR-leuT-proM* precursor made extensive use of the same T170V 5' sensor mutant of NTH-RNase E described here.

These results, together with the detection of cleavage sites *in vivo* and their reconstitution and enrichment following *in vitro* incubation of total mRNA with T170V (Chapter 4 and Chapter 5), prompted a more extensive study of RNA cleavage by T170V, which was also led by Louise Kime and Justin Clarke. To enrich further for sites of direct entry, the total RNA used as substrate for *in vitro* cleavage assays was treated enzymatically to remove 5'-monophosphorylated ends. Furthermore, a group of processing and degradation sites that are highly dependent on RNase E *in vivo* were identified by making and comparing libraries of enriched mRNA isolated from an *rne*^{ts} strain and its congenic wild-type partner at a non-permissive temperature (Clarke et al., "in preparation"). The pipeline for the follow-on

study was the one developed here. This study also confirmed that RNase E cleaves sites by direct entry within mRNA, tRNA, rRNA and sRNA transcripts. Thus, contrary to earlier expectations (Schoenberg, 2007), the recognition of substrates by direct entry (Joyce and Dreyfus, 1998; Leroy et al., 2002; Baker and Mackie, 2003; Bardey et al., 2005; Hammarlof and Hughes, 2008) pervades in RNA metabolism in *Escherichia coli* and likely in many other bacterial species that contain homologues of RNase E (Danchin, 2009; Kaberdin et al., 2011). It will be interesting to investigate whether other RNases that are unrelated to RNase E, but have been found to prefer the 5'-monophosphorylated forms of some substrates, e.g. *Bacillus subtilis* RNase Y (Shahbabian et al., 2009), can also cleave a subset of transcripts by direct entry. As a result of the work initiated here, it has been concluded that the ability of RNase E to interact with 5'-monophosphorylated ends is not required for efficient cleavage at a plethora of sites within all classes of RNA within the *E. coli* transcriptome (Clarke et al., "in preparation").

Moving forward in the determination of the contribution of other enzymes towards RNA degradation and processing, using the methodology described herein, the research group has now produced a collection of dRNA-seq maps, which include additional *in vitro* enzyme incubations and the deletion or inactivation of other components of the RNA degradation machinery in *E. coli*, such as, RNase G, RNase III and PNPase. The future aim of having now obtained these maps will be to provide a genome wide characterisation of sites that are exclusive to each of these enzymes, enabling a better understanding of the biochemical requirements for recognition and cleavage. This will help elucidate the interactions that occur between the various components of the RNA degradation and processing machinery within the cell.

An additional factor which has been shown to affect RNA processing is the repertoire of sRNAs within *E. coli*. Work presented here detected a vast quantity of novel sRNAs (Chapter 4). This relates to RNA processing as previously identified sRNAs have been shown to play an important role in the regulation of gene expression and mRNA decay (see Section 1.3). As explained in Section 1.3, sRNA/RNase E interactions are usually mediated by Hfq. Nevertheless, it is uncertain whether RNA destabilisation is a result of the blockage of translation and lack of ribosome coverage or the directed action of RNase E on a transcript via Hfq. Thus performing dRNA-seq using Hfq deficient strains together with the *ΔrppH* strain could shed further light on which sRNAs are recruited by Hfq and which of these stimulate 5' monophosphate dependent cleavage. Additionally, proteins such as RapZ,

which also recruits RNase E and interacts with adapter sRNAs and Hfq in order to direct cleavage (Gopel et al., 2013), can also be assessed via dRNA-seq. Nevertheless, the possibility remains that RppH is not the only 5' pyrophosphohydrolase present in *E. coli* and another pyrophosphohydrolase could direct an alternative mechanism for decapping of 5' triphosphorylated RNA.

An additional serendipitous finding was the detection of MazF-like processing at the 3' end of 16S rRNA, which removes the last 43 nt of this RNA along with the anti-SD sequence (Vesper et al., 2011) in order to generate specialized ribosomes which, in turn, transcribe leaderless mRNAs (Vesper et al., 2011). This was unexpected as the MazF toxin generates a 5'-hydroxylated downstream product following cleavage and this ribonuclease is expressed only during stress conditions (Nirenberg, 2004; Vesper et al., 2011). These findings were unexpected for two reasons. Firstly, we did not knowingly subject the cells to any kind of stress. Secondly, if the fragment was generated following MazF cleavage, it would not be detected by our approach as we designed the approach to only detect 5' monophosphate or 5' triphosphate ends depending on how total RNA was treated following extraction. Further analysis by our laboratory has indicated that this MazF like processing of 16S rRNA could be RNase E dependent. Of further interest was the finding that RhlB has a leaderless transcript (Chapter 4); it is possible that the up-regulation of RhlB during stress conditions could result in the modification of the architecture of other leaderless and leadered mRNAs in order to facilitate ribosome access or their decay, respectively. With this in mind and the possibility of a yet unidentified 5' kinase phosphorylating leaderless mRNAs to facilitate their translation by specialised ribosomes a modified mechanism to that previously reported (Vesper et al., 2011) is proposed (Figure 6.1); nevertheless, as is the case with other models (Lim et al., 2005; Woodall, 2006), it is possible that other factors have a direct or indirect input to this model.

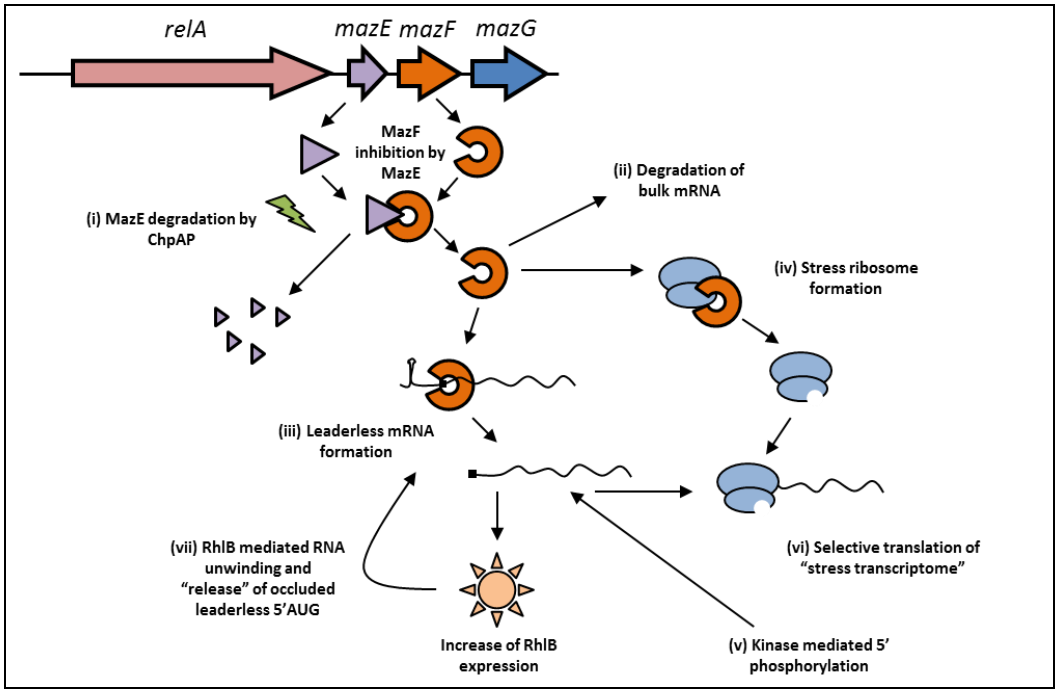

Figure 6.1 - MazF leaderless mRNA stress transcriptome model. Diagram showing the proposed mechanism in which MazF acts, in concert with a group of enzymes, to shift the transcriptome of a cell from an exponential growth conformation to a stress response one. Following stress induction, the mazEF module is expressed. This results in the degradation of anti-toxin MazE by ChpAP (i). MazF is thus released and its activity leads to the degradation of the majority of transcripts (ii), furthermore, it removes the 5' UTR of specific transcripts and renders them leaderless (iii); moreover it also removes the 3' terminal 43 nts of 16S rRNA which contain the anti-Shine-Dalgarno sequence (iv). These specialized ribosomes mediate the selective translation of the leaderless mRNA pool (vi). In addition to this, it is proposed that a 5' kinase phosphorylates the 5' hydroxylated leaderless mRNAs generated by mazF so that they can be readily translated (v) and that the leaderless *rhlB* transcript is responsible for the up regulation of RhlB which aids translation of leaderless mRNAs with have a 5' structure (vii). Adapted from (Vesper et al., 2011).

The work described here presents a novel approach for the sequencing and mapping of the primary and secondary transcriptomes of bacterial genomes. Additionally, this work presented a method that allows the number of individual transcript numbers per mRNA per cell to be roughly calculated by calibrating the number of reads for each mRNA against that of RNAs whose abundance is well characterized (Section 4.3). However, this method is semi quantitative and calibrates the transcriptome of a whole population of cells. Even during exponential growth in nutrient rich media, gene expression has stochastic fluctuations (Elowitz et al., 2002; Davidson and Surette, 2008; Nikolic et al., 2013). This means that it is impossible to determine if observations, such as 16S rRNA processing by MazF in exponentially growing cultures is a common feature of all cells or a result of a sub-population expressing stress genes. Recently, the whole transcriptome of a single bacterial cell was sequenced (Kang et al., 2011), a feat that was previously restricted to eukaryotic

cells. This advance will further aid the investigation of stochastic gene expression and help determine whether features observed in a population are an artefact caused by a sub-population or indeed a common feature across the whole population.

An additional application for these technologies that is worth mentioning would be their use to better understand the interactions that take place between the host and the pathogen during infection. This has been recently suggested to be possible by simultaneously sequencing and studying the transcriptomes of both organisms; thus being able to assess the reaction of the host during infection and that of both host and pathogen to treatments at the transcript level (Westermann et al., 2012). This has now evolved beyond concept as the authors of the cited article have communicated that this approach has now been performed experimentally. Alongside this, a study dually assessing the transcriptome of the host and the pathogenic murine cytomegalovirus has also been recently published (Juranic Lisnic et al., 2013).

In addition to *E. coli, t*he combination of improved global and differential RNA-seq presented in this thesis was extended to the analysis of *Propionibacterium acnes*, a major contributor to wide-spread human disease, and *Streptomyces coelicolor*, a model system for understanding the control of the production of secondary metabolites, such as antibiotics, produced by the streptomycetes.

Supplement

Table S.1 - Transcriptional start sites identified for *E. coli*. This table presents all the values that fell above the upper envelope boundary (Figure 4.2). The first column represents the TSS id which was constructed based on the direction of the TSS, forward (FWD) or reverse (RVS), followed by the TSS class and the first nucleotide of the TSS. The second and third columns represent the left and right limits of the TSS respectively (nucleotide positions within 8 nt of each other were classed as belonging to the same TSS) (see Section 4.2.1).The last column refers to whether or not the TSSs identified showed any overlap with experimentally verified TSSs in RegulonDB. See Excel file on attached CD.

Table S.2 – sRNAs detected. This table presents the sRNAs identified by our approach. The columns, left to right, show the sRNA id, flanking genes, flanking genes orientation, sRNA left and right positions, sRNA length, elements overlapping the sRNA, databases associated with the sRNA and whether the sRNA has been identified by others or is novel. See Excel file on attached CD.

Product size	Sequence	Distance between TSS and +1 (nt)
Small	5'- CATGAGGATTACCCATGTCGAAGACAACAAGTTCAACTCTTTATGTATTATGAGC -3'	**0**
Small	5'-CATGAGGATTACCCATGTCGAAGACAACAAAGAAGTTCAACTCTTTATGTATTATGAGC-3'	**0**
Large	5'-CATGAGGATTACCCATGTCGAAGACAAAGAAGTTCAACTCTTTATGTATTACTAAAGGTTGACTTTATTTCACCGGATACGCTTTCGTAAAGCAATAGTAAGCTGATATTCTACCACACTATGAGC-3'	**70**
Large	5'-CATGAGGATTACCCATGTCGAAGACAACAAAGAAGTTCAACTCTTTATGTATTACTAAAGGTTGACTTTATTTCACCGGATACGCTTTCGTAAAGCAATAGTAAGCTGATATTCTACCACACTATG -3'	**70**
Large	5'-CATGAGGATTACCCatgtCGAAGACAAAGAAGTTCAACTCTTTATGTATTACTAAAGGTTGACTTTATTTCACCGGATACGCTTTCGTAAAGCAATAGTAAGCTGATATTCTACCACACTATGAGC-3'	**70**

Table S.3 - *rhlB* mRNA TSS sequencing results. Sequences of the small and large PCR products of *rhlB.* (Figure 5.1). The sequence for the adaptor ligated during the RLM-RT stage is highlighted blue and the +1 site is presented in red. The PCR product size and the distance between the +1 and the TSSs are also indicated.

References

Ahlquist, P. 2002. RNA-dependent RNA polymerases, viruses, and RNA silencing. Science, 296**,** 1270-3.

Al Rashdi, A. S., Jadhav, B. L., Deshpande, T. & Deshpande, U. 2014. Whole-Genome Sequencing and Annotation of a Clinical Isolate of Mycobacterium tuberculosis from Mumbai, India. Genome Announc, 2.

Albrecht, M., Sharma, C. M., Reinhardt, R., Vogel, J. & Rudel, T. 2010. Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. Nucleic Acids Res., 38**,** 868-77.

Amitai, S., Kolodkin-Gal, I., Hananya-Meltabashi, M., Sacher, A. & Engelberg-Kulka, H. 2009. Escherichia coli MazF leads to the simultaneous selective synthesis of both "death proteins" and "survival proteins". PLoS Genet, 5**,** e1000390.

Andrade, J. M., Pobre, V., Silva, I. J., Domingues, S. & Arraiano, C. M. 2009. The role of 3'-5' exoribonucleases in RNA degradation. In: CONDON, C. (ed.) Molecular Biology of RNA Processing and Decay in Prokaryotes. Academic Press.

Anupama, K., Leela, J. K. & Gowrishankar, J. 2011. Two pathways for RNase E action in Escherichia coli in vivo and bypass of its essentiality in mutants defective for Rho-dependent transcription termination. Mol. Microbiol., 82**,** 1330-48.

Apirion, D. & Lassar, A. B. 1978. Conditional lethal mutant of *Escherichia coli* which affects processing of ribosomal RNA. J. Biol. Chem., 253**,** 1738-1742.

Appel, R. D., Sanchez, J. C., Bairoch, A., Golaz, O., Ravier, F., Pasquali, C., Hughes, G. J. & Hochstrasser, D. F. 1996. The SWISS-2DPAGE database of two-dimensional polyacrylamide gel electrophoresis, its status in 1995. Nucleic Acids Res., 24**,** 180-1.

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

Aristarkhov, A., Mikulskis, A., Belasco, J. G. & Lin, E. C. C. 1996. Translation of the *adhE* transcript to produce ethanol dehydrogenase requires RNase III cleavage in *Escherichia coli*. J. Bacteriol., 178**,** 4327-4332.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. & Mori, H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol., 2.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W. & Noble, W. S. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res., 37**,** W202-8.

Baker, K. E. & Mackie, G. A. 2003. Ectopic RNase E sites promote bypass of 5 '-end-dependent mRNA decay in *Escherichia coli*. Mol. Microbiol., 47**,** 75-88.

Bandyra, K. J. & Luisi, B. F. 2013. Licensing and due process in the turnover of bacterial RNA. RNA Biol., 10**,** 627-35.

Bardey, V., Vallet, C., Robas, N., Charpentier, B., Thouvenot, B., Mougin, A., Hajnsdorf, E., Regnier, P., Springer, M. & Branlant, C. 2005. Characterization of the molecular mechanisms involved in the differential production of erythrose-4-phosphate dehydrogenase, 3-phosphoglycerate kinase and class II fructose-1,6-bisphosphate aldolase in *Escherichia coli*. Mol. Microbiol., 57**,** 1265-1287.

Baumeister, R., Flache, P., Melefors, O., Vongabain, A. & Hillen, W. 1991. Lack of a 5' noncoding region in Tn1721-encoded *tetR* mRNA is associated with a low efficiency of translation and a short half-life in *Escherichia coli*. Nucleic Acids Res., 19**,** 4595-4600.

Beaume, M., Hernandez, D., Farinelli, L., Deluen, C., Linder, P., Gaspin, C., Romby, P., Schrenzel, J. & Francois, P. 2010. Cartography of methicillin-resistant *S. aureus* transcripts: detection, orientation and temporal expression during growth phase and stress conditions. Plos One, 5**,** e10725.

Benahmed, F. H., Gopinath, G. R., Wang, H., Jean-Gilles Beaubrun, J., Grim, C., Cheng, C. M., McClelland, M., Ayers, S., Abbott, J., Desai, P., Frye, J. G., Weinstock, G., Hammack, T. S., Hanes, D. E., Rasmussen, M. A. & Davidson, M. K. 2014. Whole-Genome Sequencing of Salmonella enterica subsp. enterica Serovar Cubana Strains Isolated from Agricultural Sources. Genome Announc, 2.

Bernard, P. & Couturier, M. 1991. The 41 carboxy-terminal residues of the miniF plasmid CcdA protein are sufficient to antagonize the killer activity of the CcdB protein. Mol. Gen. Genet., 226**,** 297-304.

Betat, H., Rammelt, C. & Morl, M. 2010. tRNA nucleotidyltransferases: ancient catalysts with an unusual mechanism of polymerization. Cell. Mol. Life Sci., 67**,** 1447-63.

Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. & Duncan, K. 2002. Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. Mol. Microbiol., 43**,** 717-731.

Bier, F. F. & Kleinjung, F. 2001. Feature-size limitations of microarray technology--a critical review. Fresenius. J. Anal. Chem., 371**,** 151-6.

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. 1997. The complete genome sequence of Escherichia coli K-12. Science, 277**,** 1453-62.

Blum, E., Carpousis, A. J. & Higgins, C. F. 1999. Polyadenylation Promotes Degradation of 3â€²-Structured RNA by theEscherichia coli mRNA Degradosome in Vitro. Journal of Biological Chemistry, 274**,** 4009-4016.

Bouvet, P. & Belasco, J. G. 1992. Control of RNase E-mediated RNA degradation by 5'-terminal base-pairing in *Escherichia coli*. Nature, 360**,** 488-491.

Bram, R. J., Young, R. A. & Steitz, J. A. 1980. The ribonuclease III site flanking 23S sequences in the 30S ribosomal precursor RNA of *E. coli*. Cell, 19**,** 393-401.

Breter, H. J. & Rhoads, R. E. 1979. Analysis of $NaIO_4$-oxidized/$NaBH_4$-reduced mRNA cap analogs by high-performance liquid anion-exchange chromatography and tobacco acid pyrophosphatase. H.-S. Z. Physiol. Chem., 360**,** 240-240.

Brooks, B. E. & Buchanan, S. K. 2008. Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. Biochimica et Biophysica Acta (BBA) - Biomembranes, 1778**,** 1930-1945.

Browning, D. F. & Busby, S. J. W. 2004. The regulation of bacterial transcription initiation. Nat Rev Micro, 2**,** 57-65.

Buck, M., Gallegos, M.-T., Studholme, D. J., Guo, Y. & Gralla, J. D. 2000. The Bacterial Enhancer-Dependent sigma 54 (sigma N) Transcription Factor. J. Bacteriol., 182**,** 4129-4136.

Bugrysheva, J. V. & Scott, J. R. 2010. The ribonucleases J1 and J2 are essential for growth and have independent roles in mRNA decay in Streptococcus pyogenes. Mol. Microbiol., 75**,** 731-43.

Butcher, J. & Stintzi, A. 2013. The transcriptional landscape of Campylobacter jejuni under iron replete and iron limited growth conditions. PLoS One, 8**,** e79475.

Buts, L., Lah, J., Dao-Thi, M. H., Wyns, L. & Loris, R. 2005. Toxin-antitoxin modules as bacterial metabolic stress managers. Trends Biochem. Sci., 30**,** 672-9.

Butt, A., Higman, V. A., Williams, C., Crump, M. P., Hemsley, C. M., Harmer, N. & Titball, R. W. 2014. The HicA toxin from Burkholderia pseudomallei has a role in persister cell formation. Biochem. J., 459**,** 333-44.

Callaghan, A. J., Aurikko, J. P., Iiag, L. L., Grossmann, J. G., Chandran, V., Kuhnel, K., Poljak, L., Carpousis, A. J., Robinson, C. V., Symmons, M. F. & Luisi, B. F. 2004. Studies of the RNA degradosome-organizing domain of the Escherichia coli ribonuclease RNase E. J. Mol. Biol., 340**,** 965-979.

Callaghan, A. J., Grossmann, J. G., Redko, Y. U., Ilag, L. L., Moncrieffe, M. C., Symmons, M. F., Robinson, C. V., McDowall, K. J. & Luisi, B. F. 2003. Quaternary structure and catalytic activity of the *Escherichia coli* ribonuclease E amino-terminal catalytic domain. Biochemistry, 42**,** 13848-13855.

Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G. & Luisi, B. F. 2005. Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. Nature, 437**,** 1187-1191.

Callahan, C., Neri-Cortes, D. & Deutscher, M. P. 2000. Purification and characterization of the tRNA-processing enzyme RNase BN. J. Biol. Chem., 275**,** 1030-4.

Cameron, V., Soltis, D. & Uhlenbeck, O. C. 1978. Polynucleotide kinase from a T4 mutant which lacks the 3' phosphatase activity. Nucleic Acids Res., 5**,** 825-33.

Cannistraro, V. J. & Kennell, D. 1989. Purification and characterization of ribonuclease M and mRNA degradation in Escherichia coli. Eur. J. Biochem., 181**,** 363-70.

Cannistraro, V. J. & Kennell, D. 1991. RNase I*, a form of RNase I, and mRNA degradation in Escherichia coli. J. Bacteriol., 173**,** 4653-9.

Carpousis, A. J. 2002. The *Escherichia coli* RNA degradosome: Structure, function and relationship to other ribonucleolytic multienyzme complexes. Biochem. Soc. Trans., 30**,** 150-155.

Carpousis, A. J., Leroy, A., Vanzo, N. & Khemici, V. 2001. *Escherichia coli* RNA degradosome. Methods Enzymol., 342**,** 333-345.

Carpousis, A. J., Luisi, B. F., McDowall, K. J. & Ciaran, C. 2009. Endonucleolytic Initiation of mRNA Decay in Escherichia coli. Progress in Molecular Biology and Translational Science. Academic Press.

Caruthers, J. M., Feng, Y. N., McKay, D. B. & Cohen, S. N. 2006. Retention of core catalytic functions by a conserved minimal ribonuclease E peptide that lacks the domain required for tetramer formation. J. Biol. Chem., 281**,** 27046-27051.

Celesnik, H., Deana, A. & Belasco, J. G. 2007. Initiation of RNA decay in *Escherichia coli* by 5 ' pyrophosphate removal. Mol. Cell, 27**,** 79-90.

Chan, P. P., Holmes, A. D., Smith, A. M., Tran, D. & Lowe, T. M. 2012. The UCSC Archaeal Genome Browser: 2012 update. Nucleic Acids Res., 40**,** D646-52.

Chandran, V. & Luisi, B. F. 2006. Recognition of enolase in the *Escherichia coli* RNA degradosome. J. Mol. Biol., 358**,** 8-15.

Charpentier, B. & Branlant, C. 1994. The Escherichia coli gapA gene is transcribed by the vegetative RNA polymerase holoenzyme E sigma 70 and by the heat shock RNA polymerase E sigma 32. J. Bacteriol., 176**,** 830-9.

Cheng, Z. F., Zuo, Y., Li, Z., Rudd, K. E. & Deutscher, M. P. 1998. The vacB gene required for virulence in Shigella flexneri and Escherichia coli encodes the exoribonuclease RNase R. J. Biol. Chem., 273**,** 14077-80.

Cho, B.-K., Zengler, K., Qiu, Y., Park, Y. S., Knight, E. M., Barrett, C. L., Gao, Y. & Palsson, B. O. 2009. The transcription unit architecture of the Escherichia coli genome. Nat. Biotech., 27**,** 1043-1049.

Christensen, S. K., Pedersen, K., Hansen, F. G. & Gerdes, K. 2003. Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. J. Mol. Biol., 332**,** 809-19.

Clarke, J. E., Kime, L., Romero A., D. & McDowall, K. J. "in preparation". Direct entry by RNase E pervades in mRNA degradation in *Escherichia coli*. "in preparation".

Coburn, G. A. & Mackie, G. A. 1998. Reconstitution of the degradation of the mRNA for ribosomal protein S20 with purified enzymes. J. Mol. Biol., 279**,** 1061-1074.

Condon, C. 2007. Maturation and degradation of RNA in bacteria. Curr. Opin. Microbiol., 10**,** 271-278.

Condon, C., Squires, C. & Squires, C. L. 1995. Control of rRNA transcription in *Escherichia coli*. Microbiol Rev, 59**,** 623-45.

Conrad, C., Schmitt, J. G., Evguenieva-Hackenberg, E. & Klug, G. 2002. One functional subunit is sufficient for catalytic activity and substrate specificity of Escherichia coli endoribonuclease III artificial heterodimers. FEBS Lett., 518**,** 93-6.

Cook, G. M., Robson, J. R., Frampton, R. A., McKenzie, J., Przybilski, R., Fineran, P. C. & Arcus, V. L. 2013. Ribonucleases in bacterial toxin-antitoxin systems. Biochim. Biophys. Acta, 1829**,** 523-31.

Cortes, T., Schubert, O. T., Rose, G., Arnvig, K. B., Comas, I., Aebersold, R. & Young, D. B. 2013. Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. Cell Rep, 5**,** 1121-31.

Crick, F. 1970. Central dogma of molecular biology. Nature, 227**,** 561-3.

Crick, F. H. 1958. On protein synthesis. Symp Soc Exp Biol, 12**,** 138-63.

Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. 2004. WebLogo: A sequence logo generator. Genome Res., 14**,** 1188-1190.

Cruz-Knight, W. & Blake-Gumbs, L. 2013. Tuberculosis: An Overview. Primary Care: Clinics in Office Practice, 40**,** 743-756.

Cudny, H. & Deutscher, M. P. 1986. High-level overexpression, rapid purification, and properties of *Escherichia coli* tRNA nucleotidyltransferase. J. Biol. Chem., 261**,** 6450-3.

Cudny, H., Zaniewski, R. & Deutscher, M. P. 1981. Escherichia coli RNase D. Purification and structural characterization of a putative processing nuclease. J. Biol. Chem., 256**,** 5627-32.

Curnow, A. W., Kung, F. L., Koch, K. A. & Garcia, G. A. 1993. Transfer RNA-Guanine transglycosylase from *Escherichia Coli*- Gross transfer-RNA structural requirements for recognition. Biochemistry, 32**,** 5239-5246.

Danchin, A. 2009. A Phylogenetic View of Bacterial Ribonucleases. Molecular Biology of RNA Processing and Decay in Prokaryotes. Prog Mol Biol Transl Sci.

Darfeuille, F., Unoson, C., Vogel, J. & Wagner, E. G. H. 2007. An Antisense RNA Inhibits Translation by Competing with Standby Ribosomes. Mol. Cell, 26**,** 381-392.

Davidson, C. J. & Surette, M. G. 2008. Individuality in Bacteria. Annu. Rev. Genet., 42**,** 253-268.

De Smet, R. & Marchal, K. 2010. Advantages and limitations of current network inference methods. Nature Rev. Microbiol., 8**,** 717-729.

Deana, A., Celesnik, H. & Belasco, J. G. 2008. The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. Nature, 451**,** 355-8.

Deutscher, M. P. 2006. Degradation of RNA in bacteria: Comparison of mRNA and stable RNA. Nucleic Acids Res., 34**,** 659-666.

Deutscher, M. P. 2009. Maturation and degradation of ribosomal RNA in bacteria. In: CONDON, C. (ed.) Molecular Biology of RNA Processing and Decay in Prokaryotes. Academic press.

Deutscher, M. P. & Marlor, C. W. 1985. Purification and characterization of Escherichia coli RNase T. J. Biol. Chem., 260**,** 7067-71.

Deutscher, M. P., Marshall, G. T. & Cudny, H. 1988. RNase PH: an Escherichia coli phosphate-dependent nuclease distinct from polynucleotide phosphorylase. Proc Natl Acad Sci U S A, 85**,** 4710-4.

Dickson, R. C., Abelson, J., Barnes, W. M. & Reznikoff, W. S. 1975. Genetic regulation: the Lac control region. Science, 187**,** 27-35.

Dornenburg, J. E., DeVita, A. M., Palumbo, M. J. & Wade, J. T. 2010. Widespread antisense transcription in *Escherichia coli*. Mbio, 1**,** 1-4.

Dreyfus, M. 2009. Killer and Protective Ribosomes. Molecular Biology of Rna Processing and Decay in Prokaryotes. San Diego: Elsevier Academic Press Inc.

Dutta, T. & Deutscher, M. P. 2010. Mode of action of RNase BN/RNase Z on tRNA precursors: RNase BN does not remove the CCA sequence from tRNA. J. Biol. Chem., 285**,** 22874-81.

Dutta, T., Malhotra, A. & Deutscher, M. P. 2012. Exoribonuclease and endoribonuclease activities of RNase BN/RNase Z both function in vivo. J. Biol. Chem., 287**,** 35747-55.

Edgar, R., Domrachev, M. & Lash, A. E. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res., 30**,** 207-10.

Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. 2002. Stochastic gene expression in a single cell. Science, 297**,** 1183-6.

Emory, S. A., Bouvet, P. & Belasco, J. G. 1992. A 5'-terminal stem-loop structure can stabilize messenger RNA in *Escherichia coli*. Genes Dev., 6**,** 135-148.

Escherich, T. 1885. Die Darmbakterien des Neugeborenen und Säuglinge. Fortschr. Med**,** 515-522.

Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Brechemmier-Baey, D. & Putzer, H. 2005. Ribonucleases J1 and J2: two novel endoribonucleases in *B. subtilis* with functional homology to *E. coli* RNase E. Nucleic Acids Res., 33**,** 2141-2152.

Ezraty, B., Dahlgren, B. & Deutscher, M. P. 2005. The RNase Z homologue encoded by Escherichia coli elaC gene is RNase BN. J. Biol. Chem., 280**,** 16542-5.

Felden, B., Himeno, H., Muto, A., McCutcheon, J. P., Atkins, J. F. & Gesteland, R. F. 1997. Probing the structure of the Escherichia coli 10Sa RNA (tmRNA). RNA, 3**,** 89-103.

Filiatrault, M. J., Stodghill, P. V., Bronstein, P. A., Moll, S., Lindeberg, M., Grills, G., Schweitzer, P., Wang, W., Schroth, G. P., Luo, S. J., Khrebtukova, I., Yang, Y., Thannhauser, T., Butcher, B. G., Cartinhour, S. & Schneider, D. J. 2010. Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs, and antisense activity. J. Bacteriol., 192**,** 2359-2372.

Funk, C. R., Zimniak, L. & Dowhan, W. 1992. The pgpA and pgpB genes of Escherichia coli are not essential: evidence for a third phosphatidylglycerophosphate phosphatase. J. Bacteriol., 174**,** 205-213.

Gao, J., Lee, K., Zhao, M., Qiu, J., Zhan, X., Saxena, A., Moore, C. J., Cohen, S. N. & Georgiou, G. 2006. Differential modulation of E. coli mRNA abundance by inhibitory proteins that alter the composition of the degradosome. Mol. Microbiol., 61**,** 394-406.

Gardner, J. F. 1982. Initiation, pausing, and termination of transcription in the threonine operon regulatory region of Escherichia coli. J. Biol. Chem., 257**,** 3896-904.

Garrey, S. M., Blech, M., Riffell, J. L., Hankins, J. S., Stickney, L. M., Diver, M., Hsu, Y. H. R., Kunanithy, V. & Mackie, G. A. 2009. Substrate binding and active site residues in RNases E and G: role of the 5' sensor. J. Biol. Chem., 284**,** 31843-31850.

Garrey, S. M. & Mackie, G. A. 2011. Roles of the 5'-phosphate sensor domain in RNase E. Mol. Microbiol., 80**,** 1613-24.

Gausing, K. 1977. Regulation of ribosome production in *Escherichia coli*: synthesis and stability of ribosomal RNA and of ribosomal protein messenger RNA at different growth rates. J. Mol. Biol., 115**,** 335-54.

Georg, J., Voss, B., Scholz, I., Mitschke, J., Wilde, A. & Hess, W. R. 2009. Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. Mol. Syst. Biol., 5**,** 305.

Gerdes, K., Christensen, S. K. & Lobner-Olesen, A. 2005. Prokaryotic toxin-antitoxin stress response loci. Nat Rev Microbiol, 3**,** 371-82.

Ghora, B. K. & Apirion, D. 1978. Structural analysis and *in vitro* processing to p5 ribosomal RNA of a 9S RNA molecule isolated from an *rne* mutant of *Escherichia coli*. Cell, 15**,** 1055-1066.

Giliberti, J., O'Donnell, S., Etten, W. J. & Janssen, G. R. 2012. A 5'-terminal phosphate is required for stable ternary complex formation and translation of leaderless mRNA in Escherichia coli. RNA, 18**,** 508-18.

Goecks, J., Nekrutenko, A. & Taylor, J. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol., 11**,** R86.

Goldsmith, Z. G. & Dhanasekaran, N. 2004. The microrevolution: applications and impacts of microarray technology on molecular biology and medicine (review). Int J Mol Med, 13**,** 483-95.

Golomb, M. & Chamberlin, M. 1974. Characterization of T7-specific Ribonucleic Acid Polymerase: IV. RESOLUTION OF THE MAJOR IN VITRO TRANSCRIPTS BY GEL ELECTROPHORESIS. J. Biol. Chem., 249**,** 2858-2863.

Gopel, Y., Papenfort, K., Reichenbach, B., Vogel, J. & Gorke, B. 2013. Targeted decay of a regulatory small RNA by an adaptor protein for RNase E and counteraction by an anti-adaptor RNA. Genes Dev., 27**,** 552-64.

Gorna, M. W., Pietras, Z., Tsai, Y. C., Callaghan, A. J., Hernandez, H., Robinson, C. V. & Luisi, B. F. 2010. The regulatory protein RraA modulates RNA-binding and helicase activities of the E. coli RNA degradosome. RNA, 16**,** 553-562.

Gruber, T. M. & Gross, C. A. 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu. Rev. Microbiol., 57**,** 441-466.

Guell, M., van Noort, V., Yus, E., Chen, W. H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kuhner, S., Rode, M., Suyama, M., Schmidt, S., Gavin, A. C., Bork, P. & Serrano, L. 2009. Transcriptome complexity in a genome-reduced bacterium. Science, 326**,** 1268-1271.

Gupta, R. S., Kasai, T. & Schlessinger, D. 1977. Purification and some novel properties of Escherichia coli RNase II. J. Biol. Chem., 252**,** 8945-9.

Gutgsell, N. S. & Jain, C. 2010. Coordinated Regulation of 23S rRNA Maturation in Escherichia coli. J. Bacteriol., 192**,** 1405-1409.

Hammarlof, D. L. & Hughes, D. 2008. Mutants of the RNA-processing enzyme RNase E reverse the extreme slow-growth phenotype caused by a mutant translation factor EF-Tu. Mol. Microbiol., 70**,** 1194-1209.

Hankins, J. S., Denroche, H. & Mackie, G. A. 2010. Interactions of the RNA-Binding Protein Hfq with cspA mRNA, Encoding the Major Cold Shock Protein. J. Bacteriol., 192**,** 2482-2490.

Hankins, J. S., Zappavigna, C., Prud'homme-Genereux, A. & Mackie, G. A. 2007. Role of RNA structure and susceptibility to RNase E in regulation of a cold shock mRNA, *cspA* mRNA. J. Bacteriol., 189**,** 4353-4358.

Harley, C. B. & Reynolds, R. P. 1987. Analysis of *Escherichia coli* promoter sequences. Nucleic Acids Res., 15**,** 2343-2361.

Hartmann, R. K., Gossringer, M., Spath, B., Fischer, S. & Marchfelder, A. 2009. The Making of tRNAs and More - RNase P and tRNase Z. In: CONDON, C. (ed.) Molecular Biology of RNA Processing and Decay in Prokaryotes. Academic Press.

Hasman, H., Saputra, D., Sicheritz-Ponten, T., Lund, O., Svendsen, C. A., Frimodt-Moller, N. & Aarestrup, F. M. 2014. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. J. Clin. Microbiol., 52**,** 139-46.

Helmann, J. D. 2002. The extracytoplasmic function (ECF) sigma factors. Adv Microb Physiol, 46**,** 47-110.

Ho, T. D. & Ellermeier, C. D. 2012. Extra cytoplasmic function σ factor activation. Curr. Opin. Microbiol., 15**,** 182-188.

Hong, B., Phornphisutthimas, S., Tilley, E., Baumberg, S. & McDowall, K. J. 2007. Streptomycin production by Streptomyces griseus can be modulated by a mechanism not associated with change in the adpA component of the A-factor cascade. Biotechnol. Lett., 29**,** 57-64.

Hook-Barnard, I., Johnson, X. B. & Hinton, D. M. 2006. Escherichia coli RNA Polymerase Recognition of a {sigma}70-Dependent Promoter Requiring a -35 DNA Element and an Extended -10 TGn Motif. J. Bacteriol., 188**,** 8352-8359.

Hsu, L. M., Klee, H. J., Zagorski, J. & Fournier, M. J. 1984. Structure of an Escherichia coli tRNA operon containing linked genes for arginine, histidine, leucine, and proline tRNAs. J. Bacteriol., 158**,** 934-42.

Hughes, K. T. & Mathee, K. 1998. The anit-sigma factors. Annu. Rev. Microbiol., 52**,** 231-286.

Icho, T. & Raetz, C. R. 1983. Multiple genes for membrane-bound phosphatases in Escherichia coli and their action on phospholipid precursors. J. Bacteriol., 153**,** 722-730.

Iost, I. & Dreyfus, M. 1995. The stability of *Escherichia coli lacZ* messenger RNA depends upon the simultaneity of its synthesis and translation. EMBO J., 14**,** 3252-3261.

Ishihama, A. 1990. Molecular assembly and functional modulation of Escherichia coli RNA polymerase. Advances in Biophysics, 26**,** 19-31.

Jacob, A. I., Kohrer, C., Davies, B. W., Rajbhandary, U. L. & Walker, G. C. 2013. Conserved Bacterial RNase YbeY Plays Key Roles in 70S Ribosome Quality Control and 16S rRNA Maturation. Mol. Cell, 49**,** 427-38.

Jacquier, A. 2009. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat. Rev. Genet., 10**,** 833-844.

Jager, D., Sharma, C. M., Thomsen, J., Ehlers, C., Vogel, J. & Schmitz, R. A. 2009. Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. Proc. Natl. Acad. Sci. USA, 106**,** 21878-21882.

Jain, C. 2008. The E. coli RhlE RNA helicase regulates the function of related RNA helicases during ribosome assembly. RNA, 14**,** 381-9.

Jain, C. & Belasco, J. G. 1995. RNase E autoregulates its synthesis by controlling the degradation rate of its own messenger RNA in *Escherichia coli*: Unusual sensitivity of the *rne* transcript to RNase E activity. Genes Dev., 9**,** 84-96.

Jarrige, A. C., Mathy, N. & Portier, C. 2001. PNPase autocontrols its expression by degrading a double-stranded structure in the *pnp* mRNA leader. EMBO J., 20**,** 6845-6855.

Jiang, X. Q., Diwa, A. & Belasco, J. G. 2000. Regions of RNase E important for 5'-end-dependent RNA cleavage and autoregulated synthesis. J. Bacteriol., 182**,** 2468-2475.

Jourdan, S. S., Kime, L. & McDowall, K. J. 2009. The sequence of sites recognised by a member of the RNase E/G family can control the maximal rate of cleavage, while a 5'-monophosphorylated end appears to function cooperatively in mediating RNA binding. Biochem. Biophys. Res. Commun., 391**,** 879-883.

Jourdan, S. S. & McDowall, K. J. 2008. Sensing of 5 ' monophosphate by *Escherichia coli* RNase G can significantly enhance association with RNA and stimulate the decay of functional mRNA transcripts *in vivo*. Mol. Microbiol., 67**,** 102-115.

Joyce, S. A. & Dreyfus, M. 1998. In the absence of translation, RNase E can bypass 5 ' mRNA stabilizers in *Escherichia coli*. J. Mol. Biol., 282**,** 241-254.

Juranic Lisnic, V., Babic Cac, M., Lisnic, B., Trsan, T., Mefferd, A., Das Mukhopadhyay, C., Cook, C. H., Jonjic, S. & Trgovcich, J. 2013. Dual Analysis of the Murine Cytomegalovirus and Host Cell Transcriptomes Reveal New Aspects of the Virus-Host Cell Interface. PLoS Pathog, 9**,** e1003611.

Kaberdin, V. R., Singh, D. & Lin-Chao, S. 2011. Composition and conservation of the mRNA-degrading machinery in bacteria. J Biomed Sci, 18**,** 23.

Kaberdin, V. R., Walsh, A. P., Jakobsen, T., McDowall, K. J. & von Gabain, A. 2000. Enhanced cleavage of RNA mediated by an interaction between substrates and the arginine-rich domain of *E. coli* ribonuclease E. J. Mol. Biol., 301**,** 257-264.

Kang, Y., Norris, M. H., Zarzycki-Siek, J., Nierman, W. C., Donachie, S. P. & Hoang, T. T. 2011. Transcript amplification from single bacterium for transcriptome analysis. Genome Res., 21**,** 925-935.

Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A. M., Kothari, A., Krummenacker, M., Latendresse, M., Muñiz-Rascado, L., Ong, Q., Paley, S., Schröder, I., Shearer, A. G., Subhraveti, P., Travers, M., Weerasinghe, D., Weiss, V., Collado-Vides, J., Gunsalus, R. P., Paulsen, I. & Karp, P. D. 2013. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res., 41**,** D605-D612.

Khemici, V., Poljak, L., Luisi, B. F. & Carpousis, A. J. 2008. The RNase E of Escherichia coli is a membrane-binding protein. Mol. Microbiol., 70**,** 799-813.

Kime, L., Clarke, J. E., Romero A., D. & McDowall, K. J. 2014. Adjacent single-stranded regions mediate processing of tRNA precursors by RNase E direct entry. Nucleic Acids Res.**,** in press.

Kime, L., Jourdan, S. S. & McDowall, K. J. 2008a. Identifying and characterizing substrates of the RNase E/G family of enzymes. In: MAQUAT, L. E. & ARRAIANO, C. M. (eds.) RNA Turnover in Bacteria, Archaea and Organelles.

Kime, L., Jourdan, S. S. & McDowall, K. J. 2008b. Identifying and characterizing substrates of the RNase E/G family of enzymes. In: MAQUAT, L. E. & ARRAIANO, C. M. (eds.) Methods in Enzymology: RNA Turnover in Bacteria, Archaea and Organelles. Academic Press.

Kime, L., Jourdan, S. S., Stead, J. A., Hidalgo-Sastre, A. & McDowall, K. J. 2010. Rapid cleavage of RNA by RNase E in the absence of 5' monophosphate stimulation. Mol. Microbiol., 76**,** 590-604.

Kobayashi, M., Nagata, K. & Ishihama, A. 1990. Promoter selectivity of *Escherichia coli* RNA-polymerase 8. Promoter selectivity of *Escherichia coli* RNA-polymerase - Effect of base substitutions in the promoter -35 region on promoter strength. Nucleic Acids Res., 18**,** 7367-7372.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227**,** 680-5.

Langmead, B. & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods, 9**,** 357-U54.

Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., de los Mozos, I. R., Vergara-Irigaray, M., Segura, V., Fagegaltier, D., Penades, J. R., Valle, J., Solano, C. & Gingeras, T. R. 2011. Genome-wide antisense transcription drives mRNA processing in bacteria. Proc. Natl. Acad. Sci. USA, 108**,** 20172-20177.

Lauber, M. A., Rappsilber, J. & Reilly, J. P. 2012. Dynamics of ribosomal protein S1 on a bacterial ribosome with cross-linking and mass spectrometry. Mol. Cell. Proteomics, 11**,** 1965-1976.

Lease, R. A., Cusick, M. E. & Belfort, M. 1998. Riboregulation in Escherichia coli: DsrA RNA acts by RNA:RNA interactions at multiple loci. Proceedings of the National Academy of Sciences, 95**,** 12456-12461.

Lee, J., Kim, Y., Hong, S. K. & Lee, Y. 2008. RNase P-dependent cleavage of polycistronic mRNAs within their downstream coding regions in Escherichia coli. Bull. Korean Chem. Soc., 29**,** 1137-1140.

Lee, K., Bernstein, J. A. & Cohen, S. N. 2002. RNase G complementation of *rne* null mutation identifies functional interrelationships with RNase E in *Escherichia coli*. Mol. Microbiol., 43**,** 1445-1456.

Lee, S. Y., Bailey, S. C. & Apirion, D. 1978. Small stable RNAs from *Escherichia coli*: evidence for existence of new molecules and for a new ribonucleoprotein particle containing 6S RNA. J. Bacteriol., 133**,** 1015-1023.

Leroy, A., Vanzo, N. F., Sousa, S., Dreyfus, M. & Carpousis, A. J. 2002. Function in *Escherichia coli* of the non-catalytic part of RNase E: role in the degradation of ribosome-free mRNA. Mol. Microbiol., 45**,** 1231-1243.

Li, Z., Gong, X., Joshi, V. H. & Li, M. 2005. Co-evolution of tRNA 3' trailer sequences with 3' processing enzymes in bacteria. RNA, 11**,** 567-77.

Li, Z., Pandit, S. & Deutscher, M. P. 1999a. The CafA protein (RNase G) together with RNase E is required for the maturation of 16S rRNA in *Escherichia coli*. FASEB J., 13**,** A1410-A1410.

Li, Z. W. & Deutscher, M. P. 2002. RNase E plays an essential role in the maturation of *Escherichia coli* tRNA precursors. RNA, 8**,** 97-109.

Li, Z. W., Pandit, S. & Deutscher, M. P. 1998. 3 ' exoribonucleolytic trimming is a common feature of the maturation of small, stable RNAs in *Escherichia coli*. Proc. Natl. Acad. Sci. USA, 95**,** 2856-2861.

Li, Z. W., Pandit, S. & Deutscher, M. P. 1999b. RNase G (CafA protein) and RNase E are both required for the 5 ' maturation of 16S ribosomal RNA. EMBO J., 18**,** 2878-2885.

Lillebaek, T., Dirksen, A., Baess, I., Strunge, B., Thomsen, V. Ø. & Andersen, Å. B. 2002. Molecular Evidence of Endogenous Reactivation of Mycobacterium tuberculosis after 33 Years of Latent Infection. J. Infect. Dis., 185**,** 401-404.

Lim, M. S., Hellard, M. E. & Aitken, C. K. 2005. The case of the disappearing teaspoons: longitudinal cohort study of the displacement of teaspoons in an Australian research institute. BMJ, 331**,** 1498-500.

Lin-Chao, S. & Cohen, S. N. 1991. The rate of processing and degradation of antisense RNAI regulates the replication of ColE1-type plasmids *in vivo*. Cell, 65**,** 1233-1242.

Lin-Chao, S., Wei, C. L. & Lin, Y. T. 1999. RNase E is required for the maturation of *ssrA* RNA and normal *ssrA* RNA peptide-tagging activity. Proc. Natl. Acad. Sci. USA, 96**,** 12406-12411.

Lin, Y.-f. 2013. Genome-wide analysis of *Propionibacterium acnes* gene regulation. PhD, University of Leeds.

Lin, Y. F., A, D. R., Guan, S., Mamanova, L. & McDowall, K. J. 2013a. A combination of improved differential and global RNA-seq reveals pervasive transcription initiation and events in all stages of the life-cycle of functional RNAs in *Propionibacterium acnes*, a major contributor to wide-spread human disease. BMC Genomics, 14**,** 620.

Lin, Y. F., Romero. A., D., Guan, S., Mamanova, L. & McDowall, K. J. 2013b. A combination of improved differential and global RNA-seq reveals pervasive transcription initiation and events in all stages of the life-cycle of functional RNAs in Propionibacterium acnes, a major contributor to wide-spread human disease. BMC Genomics, 14**,** 620.

Liou, G. G., Jane, W. N., Cohen, S. N., Lin, N. S. & Lin-Chao, S. 2001. RNA degradosomes exist *in vivo* in *Escherichia coli* as multicomponent complexes associated with the cytoplasmic membrane via the N-terminal region of ribonuclease E. Proc. Natl. Acad. Sci. USA, 98**,** 63-68.

Lisser, S. & Margalit, H. 1993. Compilation of *Escherichia coli* mRNA promoter sequences. Nucleic Acids Res., 21**,** 1507-1516.

Littauer, U. Z. & Soreq, H. 1982. The regulatory function of poly(A) and adjacent 3' sequences in translated RNA. Prog. Nucleic Acid Res. Mol. Biol., 27**,** 53-83.

Lu, Y. H., Guan, Z., Zhao, J. & Raetz, C. R. 2011. Three phosphatidylglycerol-phosphate phosphatases in the inner membrane of Escherichia coli. J. Biol. Chem., 286**,** 5506-18.

Lundberg, U. & Altman, S. 1995. Processing of the precursor to the catalytic RNA subunit of RNase P from *Escherichia coli*. RNA, 1**,** 327-334.

Lynn, S. P., Burton, W. S., Donohue, T. J., Gould, R. M., Gumport, R. I. & Gardner, J. F. 1987. Specificity of the attenuation response of the threonine operon of Escherichia coli is determined by the threonine and isoleucine codons in the leader transcript. J. Mol. Biol., 194**,** 59-69.

Mackie, G. A. 1998. Ribonuclease E is a 5'-end-dependent endonuclease. Nature, 395**,** 720-723.

Mackie, G. A. 2000. Stabilization of circular *rpsT* mRNA demonstrates the 5 '-end dependence of RNase E action *in vivo*. J. Biol. Chem., 275**,** 25069-25072.

Mackie, G. A. 2013a. Determinants in the *rpsT* mRNAs recognized by the 5'-sensor domain of RNase E. Mol. Microbiol., 89**,** 388-402.

Mackie, G. A. 2013b. RNase E: at the interface of bacterial RNA processing and decay. Nature Rev. Microbiol., 11**,** 45-57.

Madan Babu, M. & Teichmann, S. A. 2003. Evolution of transcription factors and the gene regulatory network in Escherichia coli. Nucleic Acids Res., 31**,** 1234-1244.

Madhugiri, R. & Evguenieva-Hackenberg, E. 2009. RNase J is involved in the 5'-end maturation of 16S rRNA and 23S rRNA in Sinorhizobium meliloti. FEBS Lett., 583**,** 2339-42.

Maki, K., Uno, K., Morita, T. & Aiba, H. 2008. RNA, but not protein partners, is directly responsible for translational silencing by a bacterial Hfq-binding small RNA. Proceedings of the National Academy of Sciences, 105**,** 10332-10337.

Malys, N. & McCarthy, J. E. G. 2011. Translation initiation: variations in the mechanism can be anticipated. Cell. Mol. Life Sci., 68**,** 991-1003.

Mamanova, L., Andrews, R. M., James, K. D., Sheridan, E. M., Ellis, P. D., Langford, C. F., Ost, T. W., Collins, J. E. & Turner, D. J. 2010. FRT-seq: amplification-free, strand-specific transcriptome sequencing. Nat. Methods, 7**,** 130-2.

Mamanova, L. & Turner, D. J. 2011. Low-bias, strand-specific transcriptome Illumina sequencing by on-flowcell reverse transcription (FRT-seq). Nat Protoc, 6**,** 1736-47.

Manabe, Y. C. & Bishai, W. R. 2000. Latent Mycobacterium tuberculosis-persistence, patience, and winning by waiting. Nat Med, 6**,** 1327-1329.

Marcaida, M. J., DePristo, M. A., Chandran, V., Carpousis, A. J. & Luisi, B. F. 2006. The RNA degradosome: Life in the fast lane of adaptive molecular evolution. Trends Biochem. Sci., 31**,** 359-365.

Marguerat, S. & Bahler, J. 2010. RNA-seq: from technology to biology. Cell. Mol. Life Sci., 67**,** 569-579.

Martin, J., Zhu, W., Passalacqua, K. D., Bergman, N. & Borodovsky, M. 2010. *Bacillus anthracis* genome organization in light of whole transcriptome sequencing. BMC Bioinformatics, 11 Suppl 3**,** S10.

Massé, E., Escorcia, F. E. & Gottesman, S. 2003. Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. Genes Dev., 17**,** 2374-2383.

McDowall, K. J. & Cohen, S. N. 1996. The N-terminal domain of the rne gene product has RNase E activity and is non-overlapping with the arginine-rich RNA-binding site. J. Biol. Chem., 255**,** 349-55.

McDowall, K. J., Hernandez, R. G., Lin-Chao, S. & Cohen, S. N. 1993. The *ams-1* and *rne-3071* temperature-sensitive mutations in the *ams* gene are in close proximity to

each other and cause substitutions within a domain that resembles a product of the *Escherichia coli mre* locus. J. Bacteriol., 175**,** 4245-4249.

McDowall, K. J., Kaberdin, V. R., Wu, S. W., Cohen, S. N. & Lin-Chao, S. 1995. Site-specific RNase E cleavage of oligonucleotides and inhibition by stem-loops. Nature, 374**,** 287-90.

McDowall, K. J., Lin-Chao, S. & Cohen, S. N. 1994. A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. J. Biol. Chem., 269**,** 10790-6.

McDowell, D. G., Burns, N. A. & Parkes, H. C. 1998. Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR. Nucleic Acids Res., 26**,** 3340-3347.

Meador, J., 3rd & Kennell, D. 1990. Cloning and sequencing the gene encoding Escherichia coli ribonuclease I: exact physical mapping using the genome library. Gene, 95**,** 1-7.

Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juarez, K., Contreras-Moreira, B., Huerta, A. M., Collado-Vides, J. & Morett, E. 2009a. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. PLoS One, 4.

Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juarez, K., Contreras-Moreira, B., Huerta, A. M., Collado-Vides, J. & Morett, E. 2009b. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. Plos One, 4**,** e7526.

Mian, I. S. 1997. Comparative sequence analysis of ribonucleases HII, III, II PH and D. Nucleic Acids Res., 25**,** 3187-95.

Miczak, A., Kaberdin, V. R., Wei, C. L. & Lin-Chao, S. 1996. Proteins associated with RNase E in a multicomponent ribonucleolytic complex. Proc. Natl. Acad. Sci. USA, 93**,** 3865-9.

Miczak, A., Srivastava, R. A. K. & Apirion, D. 1991. Location of the RNA processing enzymes RNase III, RNase E and RNase P in the *Escherichia coli* cell. Mol. Microbiol., 5**,** 1801-1810.

Misra, T. K. & Apirion, D. 1979. RNase E, an RNA processing enzyme from *Escherichia coli*. J. Biol. Chem., 254**,** 1154-1159.

Mitschke, J., Georg, J., Scholz, I., Sharma, C. M., Dienst, D., Bantscheff, J., Voss, B., Steglich, C., Wilde, A., Vogel, J. & Hess, W. R. 2011. An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp PCC6803. Proc. Natl. Acad. Sci. USA, 108**,** 2124-2129.

Mohanty, B. K. & Kushner, S. R. 2000. Polynucleotide phosphorylase functions both as a 3 '-> 5 ' exonuclease and a poly(A) polymerase in *Escherichia coli*. Proc. Natl. Acad. Sci. USA, 97**,** 11966-11971.

Moll, I., Grill, S., Gualerzi, C. O. & Blasi, U. 2002. Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. Mol. Microbiol., 43**,** 239-246.

Morita, T., Maki, K. & Aiba, H. 2005. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. Genes Dev., 19**,** 2176-2186.

Morita, T., Mochizuki, Y. & Aiba, H. 2006. Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. Proc. Natl. Acad. Sci. USA, 103**,** 4858-4863.

Mudd, E. A. & Higgins, C. F. 1993. *Escherichia coli* endoribonuclease RNase E: Autoregulation of expression and site-specific cleavage of messenger RNA. Mol. Microbiol., 9**,** 557-568.

Munoz-Gomez, A. J., Santos-Sierra, S., Berzal-Herranz, A., Lemonnier, M. & Diaz-Orejas, R. 2004. Insights into the specificity of RNA cleavage by the Escherichia coli MazF toxin. FEBS Lett., 567**,** 316-20.

Nakagawa, S., Niimura, Y., Miura, K. & Gojobori, T. 2010. Dynamic evolution of translation initiation mechanisms in prokaryotes. Proc Natl Acad Sci U S A, 107**,** 6382-7.

Newbury, S. F., Smith, N. H., Robinson, E. C., Hiles, I. D. & Higgins, C. F. 1987. Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell, 48**,** 297-310.

Nikolic, N., Barner, T. & Ackermann, M. 2013. Analysis of fluorescent reporters indicates heterogeneity in glucose uptake and utilization in clonal bacterial populations. BMC Microbiology, 13**,** 258.

Nirenberg, M. 2004. Historical review: Deciphering the genetic code--a personal account. Trends Biochem. Sci., 29**,** 46-54.

Niyogi, S. K. & Datta, A. K. 1975. A novel oligoribonuclease of Escherichia coli. I. Isolation and properties. J. Biol. Chem., 250**,** 7307-12.

Norris, T. E. & Koch, A. L. 1972. Effect of growth rate on the relative rates of synthesis of messenger, ribosomal and transfer RNA in *Escherichia coli*. J. Mol. Biol., 64**,** 633-49.

Nurmohamed, S., Vincent, H. A., Titman, C. M., Chandran, V., Pears, M. R., Du, D., Griffin, J. L., Callaghan, A. J. & Luisi, B. F. 2011. Polynucleotide phosphorylase activity may be modulated by metabolites in Escherichia coli. J. Biol. Chem., 286**,** 14315-23.

Ono, M. & Kuwano, M. 1979. Conditional lethal mutation in an *Escherichia coli* strain with a longer chemical lifetime of messenger RNA. J. Mol. Biol., 129**,** 343-357.

Otsuka, Y., Ueno, H. & Yonesaki, T. 2003. Escherichia coli endoribonucleases involved in cleavage of bacteriophage T4 mRNAs. J. Bacteriol., 185**,** 983-90.

Otsuka, Y. & Yonesaki, T. 2005. A novel endoribonuclease, RNase LS, in Escherichia coli. Genetics, 169**,** 13-20.

Ow, M. C. & Kushner, S. R. 2002. Initiation of tRNA maturation by RNase E is essential for cell viability in *E. coli*. Genes Dev., 16**,** 1102-1115.

Pace, N. R. & Brown, J. W. 1995. Evolutionary perspective on the structure and function of ribonuclease P, a ribozyme. J. Bacteriol., 177**,** 1919-28.

Pfeifer-Sancar, K., Mentz, A., Ruckert, C. & Kalinowski, J. 2013. Comprehensive analysis of the Corynebacterium glutamicum transcriptome using an improved RNAseq technique. BMC Genomics, 14**,** 888.

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J. C. D. & Vogel, J. 2009. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nat. Struct. Mol. Biol., 16**,** 840-846.

Py, B., Higgins, C. F., Krisch, H. M. & Carpousis, A. J. 1996. A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. Nature, 381**,** 169-172.

Quinlan, A. R. & Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26**,** 841-842.

Rajagopala, S. V., Sikorski, P., Kumar, A., Mosca, R., Vlasblom, J., Arnold, R., Franca-Koh, J., Pakala, S. B., Phanse, S., Ceol, A., Hauser, R., Siszler, G., Wuchty, S., Emili, A., Babu, M., Aloy, P., Pieper, R. & Uetz, P. 2014. The binary protein-protein interaction landscape of Escherichia coli. Nat. Biotech., 32**,** 285-290.

Rasmussen, A. A., Eriksen, M., Gilany, K., Udesen, C., Franch, T., Petersen, C. & Valentin-Hansen, P. 2005. Regulation of ompA mRNA stability: the role of a small regulatory RNA in growth phase-dependent control. Mol. Microbiol., 58**,** 1421-1429.

Rasmussen, S., Nielsen, H. B. & Jarmer, H. 2009. The transcriptionally active regions in the genome of *Bacillus subtilis*. Mol. Microbiol., 73**,** 1043-1057.

Redko, Y., Tock, M. R., Adams, C. J., Kaberdin, V. R., Grasby, J. A. & McDowall, K. J. 2003. Determination of the catalytic parameters of the N-terminal half of *Escherichia coli*

ribonuclease E and the identification of critical functional groups in RNA substrates. J. Biol. Chem., 278**,** 44001-8.

Reuven, N. B., Zhou, Z. H. & Deutscher, M. P. 1997. Functional overlap of tRNA nucleotidyltransferase, poly(A) polymerase I, and polynucleotide phosphorylase. J. Biol. Chem., 272**,** 33255-33259.

Richards, J., Liu, Q., Pellegrini, O., Celesnik, H., Yao, S., Bechhofer, D. H., Condon, C. & Belasco, J. G. 2011. An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in Bacillus subtilis. Mol. Cell, 43**,** 940-9.

Robertson, H. D., Webster, R. E. & Zinder, N. D. 1967. A nuclease specific for double-stranded RNA. Virology, 32**,** 718-9.

Rumbo-Feal, S., Gomez, M. J., Gayoso, C., Alvarez-Fraga, L., Cabral, M. P., Aransay, A. M., Rodriguez-Ezpeleta, N., Fullaondo, A., Valle, J., Tomas, M., Bou, G. & Poza, M. 2013. Whole transcriptome analysis of Acinetobacter baumannii assessed by RNA-sequencing reveals different mRNA expression profiles in biofilm compared to planktonic cells. PLoS One, 8**,** e72968.

Salgado, H., Gama-Castro, S., Peralta-Gil, M., Diaz-Peredo, E., Sanchez-Solano, F., Santos-Zavaleta, A., Martinez-Flores, I., Jimenez-Jacinto, V., Bonavides-Martinez, C., Segura-Salazar, J., Martinez-Antonio, A. & Collado-Vides, J. 2006. RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. Nucleic Acids Res., 34**,** D394-D397.

Sambrook, J. 2001. Molecular cloning : a laboratory manual. In: RUSSELL, D. W. (ed.) 3rd ed. ed. Cold Spring Harbor, N.Y. :: Cold Spring Harbor Laboratory Press.

Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 270**,** 467-70.

Schneider, K. L., Pollard, K. S., Baertsch, R., Pohl, A. & Lowe, T. M. 2006. The UCSC archaeal genome browser. Nucleic Acids Res., 34**,** D407-D410.

Schoenberg, D. R. 2007. The end defines the means in bacterial mRNA decay. Nat. Chem. Biol., 3**,** 535-536.

Schuster, C. F. & Bertram, R. 2013. Toxin–antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate. FEMS Microbiol. Lett., 340**,** 73-85.

Selinger, D. W., Saxena, R. M., Cheung, K. J., Church, G. M. & Rosenow, C. 2003. Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. Genome Res., 13**,** 216-223.

Shahbabian, K., Jamalli, A., Zig, L. & Putzer, H. 2009. RNase Y, a novel endoribonuclease, initiates riboswitch turnover in *Bacillus subtilis*. EMBO J., 28**,** 3523-33.

Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermuller, J., Reinhardt, R., Stadler, P. F. & Vogel, J. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature, 464**,** 250-255.

Sharma, U. K. & Chatterji, D. 2010. Transcriptional switching in Escherichia coli during stress and starvation by modulation of σ70 activity. FEMS Microbiol. Rev., 34**,** 646-657.

Shelness, G. S. & Williams, D. L. 1985. Secondary structure analysis of apolipoprotein II mRNA using enzymatic probes and reverse transcriptase. Evaluation of primer extension for high resolution structure mapping of mRNA. J. Biol. Chem., 260**,** 8637-46.

Shine, J. & Dalgarno, L. 1974. 3'-terminal sequence of *Escherichia coli* 16S rRNA: possible role in initiation and termination of protein synthesis. Proceedings of the Australian Biochemical Society, 7**,** 72-72.

Shine, J. & Dalgarno, L. 1975. Terminal sequence analysis of bacterial rRNA: correlation between 3'-terminal polypyrimidine sequence of 16S RNA and translational specificity of ribosome. Eur. J. Biochem., 57**,** 221-230.

Singh, A., Mai, D., Kumar, A. & Steyn, A. J. 2006. Dissecting virulence pathways of *Mycobacterium tuberculosis* through protein-protein association. Proc. Natl. Acad. Sci. USA, 103**,** 11346-51.

Sorensen, M. A., Fricke, J. & Pedersen, S. 1998. Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in Escherichia coli in vivo. J. Mol. Biol., 280**,** 561-569.

Soreq, H. & Littauer, U. Z. 1977. Purification and characterization of polynucleotide phosphorylase from Escherichia coli. Probe for the analysis of 3' sequences of RNA. J. Biol. Chem., 252**,** 6885-8.

Sousa, S., Marchand, I. & Dreyfus, M. 2001. Autoregulation allows Escherichia coli RNase E to adjust continuously its synthesis to that of its substrates. Mol. Microbiol., 42**,** 867-878.

Sprenger, G. A., Schorken, U., Sprenger, G. & Sahm, H. 1995. Transaldolase B of Escherichia coli K-12: cloning of its gene, talB, and characterization of the enzyme from recombinant strains. J. Bacteriol., 177**,** 5930-6.

Stern, M. J., Ames, G. F., Smith, N. H., Robinson, E. C. & Higgins, C. F. 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell, 37**,** 1015-26.

Storms, V., Claeys, M., Sanchez, A., De Moor, B., Verstuyf, A. & Marchal, K. 2010. The effect of orthology and coregulation on detecting regulatory motifs. PLoS One, 5**,** e8938.

Struck, J. C., Lempicki, R. A., Toschka, H. Y., Erdmann, V. A. & Fournier, M. J. 1990. Escherichia coli 4.5S RNA gene function can be complemented by heterologous bacterial RNA genes. J. Bacteriol., 172**,** 1284-1288.

Subbarayan, P. R. & Deutscher, M. P. 2001. Escherichia coli RNase M is a multiply altered form of RNase I. RNA, 7**,** 1702-7.

Sutherland, I., Svandova, E. & Radhakrishna, S. 1976. Alternative models for the development of tuberculosis disease following infection with tubercle bacilli. Bull Int Union Tuberc, 51**,** 171-9.

Suzek, B. E., Ermolaeva, M. D., Schreiber, M. & Salzberg, S. L. 2001. A probabilistic, method for identifying start codons in bacterial genomes. Bioinformatics, 17**,** 1123-1130.

Taghbalout, A. & Rothfield, L. 2007. RNaseE and the other constituents of the RNA degradosome are components of the bacterial cytoskeleton. Proc. Natl. Acad. Sci. USA, 104**,** 1667-1672.

Taghbalout, A. & Rothfield, L. 2008. RNaseE and RNA helicase B play central roles in the cytoskeletal organization of the RNA degradosome. J. Biol. Chem., 283**,** 13850-5.

Taraseviciene, L., Bjork, G. R. & Uhlin, B. E. 1995. Evidence for an RNA binding region in the *Escherichia coli* processing endoribonuclease RNase E. J. Biol. Chem., 270**,** 26391-26398.

The_ENCODE_Project_Consortium 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science, 306**,** 636-640.

Thisted, T., Nielsen, A. K. & Gerdes, K. 1994a. Mechanism of post-segregational killing: translation of Hok, SrnB and Pnd mRNAs of plasmids R1, F and R483 is activated by 3'-end processing. EMBO J, 13**,** 1950-9.

Thisted, T., Sorensen, N. S., Wagner, E. G. & Gerdes, K. 1994b. Mechanism of post-segregational killing: Sok antisense RNA interacts with Hok mRNA via its 5'-end single-stranded leader and competes with the 3'-end of Hok mRNA for binding to the mok translational initiation region. EMBO J., 13**,** 1960-8.

Tock, M. R., Walsh, A. P., Carroll, G. & McDowall, K. J. 2000. The CafA protein required for the 5'-maturation of 16 S rRNA is a 5'- end-dependent ribonuclease that has context-dependent broad sequence specificity. J. Biol. Chem., 275**,** 8726-32.

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., Barthelemy, M., Vergassola, M., Nahori, M. A., Soubigou, G., Regnault, B., Coppee, J. Y., Lecuit, M., Johansson, J. & Cossart, P. 2009. The *Listeria* transcriptional landscape from saprophytism to virulence. Nature, 459**,** 950-956.

Tomcsanyi, T. & Apirion, D. 1985. Processing enzyme ribonuclease E specifically cleaves RNAI: An inhibitor of primer formation in plasmid DNA synthesis. J. Mol. Biol., 185**,** 713-720.

Uguru, G. C., Stephens, K. E., Stead, J. A., Towle, J. E., Baumberg, S. & McDowall, K. J. 2005. Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in Streptomyces coelicolor. Mol. Microbiol., 58**,** 131-50.

Uzan, M. 2009. RNA Processing and Decay in Bacteriophage T4. Prog Mol Transl Sci., 85**,** 43-89.

van Wezel, G. P. & McDowall, K. J. 2011. The regulation of the secondary metabolism of Streptomyces: new links and experimental advances. Nat. Prod. Rep., 28**,** 1311-33.

VanGuilder, H. D., Vrana, K. E. & Freeman, W. M. 2008. Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques, 44**,** 619-26.

Vanzo, N. F., Li, Y. S., Py, B., Blum, E., Higgins, C. F., Raynal, L. C., Krisch, H. M. & Carpousis, A. J. 1998. Ribonuclease E organizes the protein interactions in the *Escherichia col*i RNA degradosome. Genes Dev., 12**,** 2770-2781.

Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A. C., Engelberg-Kulka, H. & Moll, I. 2011. Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. Cell, 147**,** 147-157.

Vioque, A., Arnez, J. & Altman, S. 1988. Protein-RNA interactions in the RNase P holoenzyme from Escherichia coli. J. Mol. Biol., 202**,** 835-848.

Vogel, J., Bartels, V., Tang, T. H., Churakov, G., Slagter-Jager, J. G., Huttenhofer, A. & Wagner, E. G. 2003. RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res., 31**,** 6435-43.

Vogel, J. & Luisi, B. F. 2011. Hfq and its constellation of RNA. Nat Rev Micro, 9**,** 578-589.

von Gabain, A., Belasco, J. G., Schottel, J. L., Chang, A. C. & Cohen, S. N. 1983. Decay of mRNA in Escherichia coli: investigation of the fate of specific segments of transcripts. Proc. Natl. Acad. Sci. USA, 80**,** 653-657.

Wachi, M. 2001. *Escherichia coli* RNase G. Nippon Nogeikagaku Kaishi-Journal of the Japan Society for Bioscience Biotechnology and Agrochemistry, 75**,** 121-127.

Wachi, M., Umitsuki, G., Shimizu, M., Takada, A. & Nagai, K. 1999. *Escherichia coli cafA* gene encodes a novel RNase, designated as RNase G, involved in processing of the 5' end of 16S rRNA. Biochem. Biophys. Res. Commun., 259**,** 483-488.

Wagner, E. G. & Unoson, C. 2012. The toxin-antitoxin system tisB-istR1: Expression, regulation, and biological role in persister phenotypes. RNA Biol., 9**,** 1513-9.

Wang, S., Dong, X., Zhu, Y., Wang, C., Sun, G., Luo, T., Tian, W., Zheng, H. & Gao, Q. 2013. Revealing of Mycobacterium marinum transcriptome by RNA-seq. PLoS One, 8**,** e75828.

Wassarman, K. M. & Storz, G. 2000. 6S RNA regulates E. coli RNA polymerase activity. Cell, 101**,** 613-23.

Weaver, R. F. & Weissmann, C. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15 S beta-globin mRNA precursor and mature 10 s beta-globin mRNA have identical map coordinates. Nucleic Acids Res., 7**,** 1175-93.

Weel-Sneve, R., Kristiansen, K. I., Odsbu, I., Dalhus, B., Booth, J., Rognes, T., Skarstad, K. & Bjoras, M. 2013. Single transmembrane peptide DinQ modulates membrane-dependent activities. PLoS Gen., 9.

Westermann, A. J., Gorski, S. A. & Vogel, J. 2012. Dual RNA-seq of pathogen and host. Nat Rev Micro, 10**,** 618-630.

Woodall, A. A. 2006. Disappearing teaspoons: teabags and forks are confounding factors. BMJ, 332**,** 121.

Wösten, M. M. S. M. 1998. Eubacterial sigma-factors. FEMS Microbiol. Rev., 22**,** 127-150.

Wurtzel, O., Sapra, R., Chen, F., Zhu, Y. W., Simmons, B. A. & Sorek, R. 2010. A single-base resolution map of an archaeal transcriptome. Genome Res., 20**,** 133-141.

Yamaguchi, Y. & Inouye, M. 2009. mRNA interferases, sequence-specific endoribonucleases from the toxin-antitoxin systems. Prog Mol Biol Transl Sci, 85**,** 467-500.

Young, R. A. & Steitz, J. A. 1978. Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. Proc Natl Acad Sci U S A, 75**,** 3593-7.

Zahrt, T. C. & Deretic, V. 2001. Mycobacterium tuberculosis signal transduction system required for persistent infections. Proceedings of the National Academy of Sciences, 98**,** 12706-12711.

Zhang, X., Zhu, L. & Deutscher, M. P. 1998. Oligoribonuclease is encoded by a highly conserved gene in the 3'-5' exonuclease superfamily. J. Bacteriol., 180**,** 2779-81.

Zhang, Y., Zhang, J., Hoeflich, K. P., Ikura, M., Qing, G. & Inouye, M. 2003. MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. Mol. Cell, 12**,** 913-23.

Zhou, Z. & Deutscher, M. P. 1997. An essential function for the phosphate-dependent exoribonucleases RNase PH and polynucleotide phosphorylase. J. Bacteriol., 179**,** 4391-5.