CHARACTERIZING TRANSCRIPTIONAL AND POST-TRANSCRIPTIONAL REGULATORS IN *STREPTOMYCES* BACTERIA

CHARACTERIZING TRANSCRIPTIONAL AND POST-TRANSCRIPTIONAL REGULATORS IN *STREPTOMYCES* BACTERIA

By

RACHEL ANN YOUNG, B.Sc.

A Thesis
Submitted to the School of Graduate Studies
In Partial Fulfillment of the Requirements of the Degree of
Master of Science

McMaster University

Master of Science (2015) McMaster University
(Biology) Hamilton, Ontario

TITLE: Characterizing transcriptional and post-transcriptional regulators in *Streptomyces* bacteria

AUTHOR: Rachel Ann Young, B.Sc. (McMaster University)

SUPERVISOR: Dr. Marie A. Elliot

NUMBER OF PAGES: xi, 94

Abstract

The final stage of transcription, termed termination, is required for proper gene expression in bacteria. However, the mechanism of transcription termination is poorly understood in the multicellular *Streptomyces* bacteria. Motivated by the lack of well characterized terminators, we systematically searched for intrinsic termination signals in three divergent *Streptomyces* species. Using a novel computational approach we identified hundreds of biologically relevant terminators in each species. Many of these terminators were found downstream of annotated genes and aligned nicely with intrinsic terminators identified by other prediction algorithms. Further *in silico* analyses indicated that the streptomycetes prefer to terminate transcription at non-canonical structures characterized by a long hairpin lacking a trailing U-rich tract. We prioritized three structurally diverse intrinsic terminators for *in vivo* analysis. Due to difficulties optimizing a fluorescent reporter assay, we have not yet been unable to measure termination efficiency *in vivo*. Current work is focused on developing a new reporter system that can be used to characterize putative *Streptomyces* terminators.

In this work, we have also investigated small RNA (sRNA)-mediated regulation in *Streptomyces*. These bacteria encode hundreds of sRNAs, although very few have been characterized to date. As a result, we designed an affinity purification system that can be used to identify sRNA interaction partners (*e.g.* proteins and RNAs). Initially, we tried to implement a system using a highly structured streptavidin aptamer; however, sRNAs tagged with this aptamer were not stably expressed *in vivo* and did not appear to re-fold effectively *in vitro*. More recently, we have tagged several sRNAs with a different aptamer which binds the MS2 phage coat protein. Future experiments will focus on testing the functionality of these new RNA fusions *in vivo*.

Acknowledgements

I want to first thank Dr. Marie Elliot for her support, guidance and patience during the last three years. This process would certainly not have been possible without her unwavering confidence in my abilities and my work. Her enthusiasm for science is contagious and has always inspired me to push the boundaries of my own research.

I would also like to thank the members of my supervisory committee, Dr. Brian Coombes and Dr. Brian Golding, for their insightful and helpful suggestions on my research. They have taught me so much in such a short amount of time.

Thanks also go to the past and present members of the Elliot lab: Dr. Julia Swiercz, Dr. David Capstick, Dr. Emma Sherwood, Matthew Moody, Andrew Duong, Renée St-Onge, Danielle Sexton, Stephanie Jones, Chris Firby and Savannah Colameco. I am grateful to have worked with such an intelligent and kind group of people. Their willingness to help and share expertise was always greatly appreciated. Without their friendship, this experience would not be nearly as memorable and I will miss spending time with each one of them.

Finally, I would like to thank my family for being the most selfless and caring people I know. Their unconditional love and support made this process easier and for that I am eternally thankful. Special thanks also go to Solomon Barkley, who has been by my side through this entire process. He was a source of calm and positivity and always ensured I never lost perspective. I am grateful to have had a constant reminder of life outside of the lab.

Table of Contents

List of Figures

List of Tables

List of abbreviations

^{32}P	phosphorus-32 isotope
A	adenine
a.u.	arbitrary units
amp	ampicillin
ATP	adenosine triphosphate
BED	Browser Extensible Data
bp	base pair
C	Cytosine; Celsius
DMSO	dimethyl sulfoxide
DNA	Difco nutrient agar; deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleotide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein
F	forward
fwd	forward
G	Guanine
g	gravity
h	hour
IGV	Integrated Genomics Viewer
K	lysine
kDa	kilodalton
kV	kilovolts
L	Litre
LB	Luria Bertani
max	maximum
Mb	megabase pairs
mCherry	monomeric Cherry protein
MCS	multiple cloning site
mg	milligram
min	minutes; minimum
mL	milliliter
mm	millimeter
mM	millimolar
mRNA	messanger RNA
mRNA	messanger RNA
MS	mannitol-soy flour
msfGFP	monomeric, superfolder green fluorescent protein
MYM	maltose-yeast extract-malt extract
NADH	nicotinamide adenine dinucleotide
ng	nanogram
nm	nanometer
nt	nucleotide
OD	optical density

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pmol	picomole
Q	glutamine
R	reverse; arginine
RBP	RNA binding protein
RBS	ribosome binding site
rev	reverse
RNA	ribonucleic acid
RNAP	RNA polymerase
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	reverse transcription
rut	Rho utilization site
SDS	sodium dodecyl sulfate
sec	seconds
seq	sequencing
SOB	super optimal broth
sRNA	small RNA
T	thymine
term	terminator
TRAP	*trp* RNA-binding attenuation protein
tRNA	transfer RNA
tRSA	tRNA-scaffolded streptavidin aptamer
TSB	tryptone soya broth
U	uracil; units
UTR	untranslated region
V	valine
YEME	yeast extract-malt extract
YT	yeast extract-tryptone
µg	microgram
µL	microlitres
µM	micromolar

Chapter 1: Introduction

1.1 The actinomycetes

For decades following their initial discovery in 1877 (Harz, 1877), the actinomycetes were assumed to be fungi due to their filamentous morphology and spore-bearing life cycle (Waksman, 1961). However, their sensitivity to antibiotics and bacterial-like cell wall have since provided evidence that the actinomycetes are instead prokaryotic organisms (Waksman, 1961; Hopwood, 1999). Today, the term actinomycete is used to describe the Gram-positive, high-GC bacteria that belong to the order *Actinomycetales* (Goodfellow and Williams, 1983; McNeil and Brown, 1994). Members of this taxonomic order are typically saprophytes which inhabit both terrestrial (soil) and aquatic (freshwater and marine) environments (Goodfellow and Williams, 1983); although, a few actinomycetes form pathogenic relationships with plants (*e.g.* the potato pathogen, *Streptomyces scabies*) and animals (*e.g.* the human pathogens, *Mycobacterium tuberculosis, M. lyprae* and *Corynebacterium diphtheriae*) (Stackebrandt and Woese, 1981; Goodfellow and Williams, 1983). The actinomycetes also display diversity in their metabolic properties (Ventura *et al.*, 2007; Nett *et al.*, 2009), and produce numerous bioactive metabolites (Bérdy, 2005; Wietz *et al.*, 2013).

1.2 The streptomycetes

Streptomyces are soil-dwelling actinomycetes best known for their ability to produce a wide range of medically relevant secondary metabolites (Nett *et al.*, 2009). These bacteria produce over two-thirds of clinical antibiotics (Bentley *et al.*, 2002), in addition to many other antifungal, anticancer, antiparasitic and immunosuppressive compounds (Nett *et al.*, 2009). While collectively these characterized compounds represent over 7,500 bioactive metabolites (Bérdy, 2005), the streptomycetes have the biosynthetic capacity to produce many more potentially useful compounds (Bentley *et al.*, 2002; Ikeda *et al.*, 2003; Ohnishi *et al.*, 2008).

These bacteria are also characterized by a multicellular life cycle that is distinct from typical binary cell division (Angert, 2005) (Figure 1.1). This complex process begins with a dormant spore, which germinates and extends germ tubes after encountering favourable growth conditions (Flärdh and Buttner, 2009). The germ tubes then grow by tip extension and branching, ultimately forming a network of substrate mycelium (Elliot *et al.*, 2008). When nutrients become limiting or in response to other unknown signals, aerial hyphae begin to extend from the vegetative substrate mycelium into the air (Flärdh and Buttner, 2009). This morphological transformation is also associated with the beginning of secondary metabolite production. Aerial hyphae then grow by tip extension (Flärdh, 2003) until a synchronous septation event divides the hyphae into multiple unigenomic pre-spore compartments. Maturation of these pre-spores ultimately results in the release of individual, metabolically inactive spores (Swiercz and Elliot, 2012).

Streptomyces coelicolor has served as the model streptomycete since the 1950s (Sermonti and Spada-Sermonti, 1955; Hopwood, 1957). More recent sequencing of the *S. coelicolor* genome revealed a large (8.7 Mb), linear chromosome which is predicted to encode 7,825 genes (Bentley *et al.*, 2002). The average GC content of the chromosome is also extremely high at 72.1% (Bentley *et al.*, 2002). This species was originally used extensively for genetic studies because of its two pigmented and easily detectable antibiotics (Hopwood, 1999): actinorhodin (blue) and undecylprodigeosin (red). However, the fact that *S. coelicolor* only grows vegetatively in liquid culture (it will only differentiate into spores during growth on solid media) has made it a challenging system for studying cellular differentiation (Bibb *et al.*, 2012). As a result, *Streptomyces venezuelae* has been adopted as a new model species. This streptomycete is favourable due to its rapid growth rate (Flärdh and Buttner, 2009) and ability to fully (and synchronously) differentiate in liquid culture (Glazebrook *et al.*, 1990; Bibb *et al.*, 2012). Notably, when grown in liquid media, *S. venezuelae* follows an analogous differentiation cycle to solid-grown cultures, only instead of raising aerial hyphae, the entire culture transitions from vegetatively growing cells to mycelial fragments which then transform into spores (Jones *et al.*, 2014). Like *S. coelicolor, S. venezuelae* is known to produce at least two antibiotics: jadomycin and chloramphenicol (Ayer *et al.*, 1991), although these lack the brilliant pigmentation of *S. coelicolor* antibiotics.

The genomes of many divergent *Streptomyces* species have now been sequenced, including *S. avermitilis* [produces the antiparastic agent, avermectin (Ikeda *et al.*, 2003)] and *S. griseus* [produces the antibiotic, streptomycin (Ohnishi *et al.*, 2008)]. As of August 2015, more than 160 *Streptomyces* genome sequences are available in the NCBI genome database.

1.2.2 Gene regulation in *Streptomyces*

Between their complex life cycle and their metabolic diversity, it is not surprising that a large portion of the streptomycete genome is dedicated to regulation. For example, over 12.3% of protein-encoding genes in *S. coelicolor* are predicted to have a regulatory role, with sigma factors being highly abundant (>60) within these regulatory proteins (Bentley *et al.*, 2002; Ohnishi *et al.*, 2008). More recently, the regulatory repertoire of *Streptomyces* has been extended beyond protein regulators to include non-coding RNAs (Pánek *et al.*, 2008; Swiercz *et al.*, 2008; Vockenhuber *et al.*, 2011; Moody *et al.*, 2013).

1.2.2.1 Non-coding RNAs

Small non-coding RNAs (sRNAs) are now established as important post-transcriptional regulators of gene expression in bacteria (Storz *et al.*, 2011). These small [50-300 nucleotides (nt) (Storz *et al.*, 2011)] untranslated RNAs regulate a diverse range of cellular processes, including iron homeostasis (Massé and Gottesman, 2002; Reinhart *et al.*, 2015), biofilm formation (Thomason *et al.*, 2012; Papenfort *et al.*, 2015) and quorum sensing (Tu and Bassler, 2007). Bacterial sRNAs typically act by base-pairing with one or more a target mRNAs, although there are several examples of sRNAs whose regulatory targets are proteins (Waters and Storz, 2009).

The vast majority of characterized sRNAs target and influence the translatability or stability of an mRNA transcript (Waters and Storz, 2009). Specifically, binding of an sRNA to its target mRNA [usually in the 5' untranslated region (UTR)] can occlude the ribosome binding site (RBS) and prevent translation of its target (Gottesman and Storz, 2011). sRNAs can also target mRNAs for degradation by ribonucleases (RNases), thereby impacting mRNA stability (Waters and Storz, 2009). Cases where sRNAs act positively on their mRNA target [*i.e.* promoting translation (Prévost *et al.*, 2007) or increasing transcript stability (Opdyke *et al.*, 2004)] have also been observed. Base-pairing sRNAs, in general, are classified as either *cis*-encoded or *trans*-encoded, based on where they are located in the genome relative to their target mRNA (Waters and Storz, 2009). *cis*-encoded sRNAs are found on the opposite strand of their mRNA target and are, therefore, entirely complementary to their target (Brantl, 2007). In contrast, *trans*-encoded sRNAs share more limited complementarity with their target RNAs, since the transcripts are expressed from distinct locations of the genome (Gottesman and Storz, 2011). Due to this limited complementarity, sRNAs encoded *in trans* often function in conjunction with an RNA chaperone (*e.g.* Hfq, see section 1.4.2) to facilitate base pairing with a target mRNA (Waters and Storz, 2009). While *trans*-encoded sRNAs are often expressed from intergenic regions, recent work is revealing diverse origins for sRNAs – *e.g.* sRNAs derived from the 3' UTRs of mRNAs (*e.g.* Chao *et al.*, 2012).

Protein-binding sRNAs exert their regulatory effects by activating, inhibiting or modifying the activity of their target protein (Storz *et al.*, 2011). Often in bacteria, this type of sRNA mimics the structure of other nucleic acids in order to sequester proteins away from their normal targets (Waters and Storz, 2009). For example, the well-characterized *Escherichia coli* 6S RNA adopts a secondary structure similar to an open promoter, which is recognized and bound by the house-keeping form of RNA polymerase (σ^{70}-RNAP) (Cavanagh and Wassarman, 2014). In stationary phase, the binding of σ^{70}-RNAP to 6S RNA generally inhibits transcription from σ^{70}-dependent promoters (Trotochaud and Wassarman, 2004).

In *Streptomyces,* hundreds of ncRNAs have been identified and experimentally validated (Pánek *et al.*, 2008; Swiercz *et al.*, 2008; D'Alia *et al.*, 2010; Vockenhuber *et al.*, 2011; Moody *et al.*, 2013); however, to date, only four have been characterized (Vockenhuber and Suess, 2012; Hindra *et al.*, 2014; Kim *et al.*, 2014; Mikulík *et al.*, 2014). The collection of characterized sRNAs in *Streptomyces* effectively illustrates the diversity of these regulators – it currently includes a protein-binding 6S homolog [scr3559 (Pánek *et al.*, 2008)], a novel antisense sRNA [scr4677 (Hindra *et al.*, 2014)] and two *trans*-encoded sRNAs [scr5239 (Vockenhuber and Suess, 2012) and s-SodF (Kim *et al.*, 2014)].

1.3 Bacterial transcription

Transcription of DNA into a functional RNA transcript is the first step in bacterial gene expression. Transcription itself is a highly regulated process that can be divided into three distinct stages: initiation, elongation and termination (von Hippel, 1998). Transcription begins when RNA polymerase (RNAP), in complex with a sigma factor, recognizes a promoter sequence on the template DNA and starts unwinding the DNA duplex

(Browning and Busby, 2004). Following transcription initiation, the RNAP transitions into an elongation complex (Browning and Busby, 2004), which extends the nascent RNA until its encounters a transcription terminator (Vassylyev *et al.*, 2007). At these termination sites, the elongation complex releases the nascent transcript and rapidly dissociates from the template DNA (Santangelo and Artsimovitch, 2011; Peters *et al.*, 2012).

1.3.1 Transcription termination

RNAP dissociation and RNA release during the final stages of transcription can occur by two distinct mechanisms: intrinsic termination and factor-dependent termination. Since termination is essential for proper bacterial gene expression, both process have been extensively studied (Peters *et al.*, 2011).

1.3.1.1 Factor-dependent terminators

In bacteria, factor-dependent termination is mediated by the auxiliary termination protein, Rho (Boudvillain *et al.*, 2013). This homo-hexameric protein recognizes and binds to C-rich unstructured regions of RNA, termed Rho-utilization (*rut*) sites (Chen *et al.*, 1986; Hart and Roberts, 1991; Nudler and Gottesman, 2002). Despite being important for Rho recognition (Richardson and Richardson, 1996), these *rut* sites display little sequence conservation beyond their C-rich nature (Alifano *et al.*, 1991; Nudler and Gottesman, 2002). Once bound to a nascent transcript, Rho transforms from an open-ring (Skordalakes and Berger, 2003) to a closed-ring structure with the RNA in the center of the hexamer (Skordalakes and Berger, 2006). In this conformation, Rho translocates along the RNA in a 5' to 3' direction in an ATP-dependent manner (Nudler and Gottesman, 2002; Peters *et al.*, 2011). Rho eventually reaches the RNAP elongation complex, where it promotes the release of the nascent transcript (Boudvillain *et al.*, 2013). Termination typically occurs at RNAP pause sites (Lau *et al.*, 1983; Jin *et al.*, 1992), which are found at varying distances (*i.e.* 10-100 nt) from the *rut* site (Boudvillain *et al.*, 2013). A single transcript can have multiple termination sites [spanning up to 100 nt, *e.g.* (Richardson and Richardson, 1996)], resulting in a dispersed pattern of termination (Richardson, 2002).

The exact mechanism underlying Rho-mediated RNAP dissociation is unknown, although a few mechanisms have been proposed. In one model, Rho is predicted to pull the RNA transcript with enough force to disrupt the RNA:DNA hybrid formed within the RNAP (Richardson, 2002). Rho has also been suggested to push the elongating RNAP along the template DNA, preventing further RNA synthesis (Park and Roberts, 2006). More recent evidence has, however, challenged these predictions and instead has supported a model in which Rho promotes a conformational change in RNAP that destabilizes the elongation complex (Epshtein *et al.*, 2010).

Despite being originally identified for its role in terminating λ phage transcription (Roberts, 1969), Rho homologs are broadly distributed throughout the eubacteria, including the streptomycetes (Ingham *et al.*, 1996; Washburn *et al.*, 2001). In *E. coli*, Rho terminates at least 200 transcripts; interestingly, half of these are associated with the 3' ends of genes, while the remaining are found intragenically (Peters *et al.*, 2009). *rut* sites are expected to

be prevalent within protein-coding regions (Nudler and Gottesman, 2002); however, translating ribosomes typically prevent Rho from binding at these sites (Adhya and Gottesman, 1978; Boudvillain *et al.*, 2013). If there is a nonsense mutation within a gene in an operon, these intergenic *rut* sites become important for repressing transcription of the downstream genes (Richardson *et al.*, 1975; Adhya and Gottesman, 1978).

1.3.1.2 Intrinsic (Rho-independent) terminators

In contrast to Rho-dependent termination, intrinsic termination occurs without any auxiliary protein factors (Peters *et al.*, 2011). Instead, a GC-rich RNA stem-loop followed by a U-rich tail is sufficient for transcription termination (Nudler and Gottesman, 2002). These features (GC-rich stem loop; U-rich tail) are characteristic of all canonical intrinsic terminators (Peters *et al.*, 2011). Once transcribed, the conserved terminator U-tract serves to pause transcription, which allows the GC-rich hairpin upstream to form within the RNA exit channel of the RNAP (Gusarov and Nudler, 1999; Peters *et al.*, 2011). Complete hairpin formation is thought to promote termination by either inducing a conformational change in RNAP (Epshtein *et al.*, 2007) or directly pulling the nascent transcript from the RNAP exit channel (Komissarova *et al.*, 2002; Santangelo and Roberts, 2004; Peters *et al.*, 2011). Both mechanisms disrupt the weak rU:dA interactions found within RNAP, and this ultimately causes the elongation complex to dissociate and release the nascent RNA transcript (Gusarov and Nudler, 1999; Komissarova *et al.*, 2002). Notably, intrinsic termination sites are extremely well-defined, unlike Rho-dependent terminators, with termination typically occurring at a single nucleotide (Wilson and Hippel, 1994; Richardson, 2002).

Studies in *E. coli* have recently provided further insight into other terminator features that are required for efficient termination. For example, investigations into the loop joining the two stem ends of the GC-rich hairpin, has revealed that tetraloops are favoured (Peters *et al.*, 2011; Chen *et al.*, 2013), likely because of their increased stability relative to other loop sizes (Varani, 1995). There also appears to be a bias for G:C base pairs at the base of the terminating hairpin (Peters *et al.*, 2011; Chen *et al.*, 2013); these are thought to provide the energy necessary to destabilize the RNA:DNA hybrid within the RNAP (Komissarova *et al.*, 2002; Larson *et al.*, 2008).

While approximately 80% of mRNA transcripts are terminated by canonical intrinsic terminators in *E. coli* (Peters *et al.*, 2009; Santangelo and Artsimovitch, 2011), these structures are not abundant in other bacteria (Mitra *et al.*, 2009). Specifically, many genes in bacteria with high genomic GC content (*e.g. Streptomyces* and *Mycobacterium*) appear to lack canonical terminators (Mitra *et al.*, 2009). This observation has led to the suggestion that non-canonical structures may contribute to transcription termination. These non-canonical structures can include I-shaped (hairpin lacking a U-tract), V-shaped (two hairpins directly in tandem) and U-shaped (two hairpins separated by <50 nt) terminators (Mitra *et al.*, 2011). An intricate branched structure has also been recently identified as a Rho-independent terminator in *Enterococcus faecalis* (Johnson *et al.*, 2014), supporting a terminator role for such non-canonical structures. However, given the current model of

transcription termination, it is unclear how these structures would promote RNAP dissociation. Interestingly, the absence of a U-tract does not appear to impede termination at some non-canonical structures (Ingham *et al.*, 1995; Unniraman *et al.*, 2001). In these cases, RNAP pausing may be facilitated by another factor [*e.g.* the downstream DNA duplex (Lee *et al.*, 1990)] or may not be required at all [*e.g.* when RNAP elongation is slow (Mitra *et al.*, 2009)]. The recent suggestion that some non-canonical terminators may be an artifact of RNA processing (Czyz *et al.*, 2014), has highlighted the need for further investigation of this process.

1.3.2 Other roles for transcription termination

In addition to defining gene boundaries, transcription terminators also serve several other important regulatory roles within the cell (Santangelo and Artsimovitch, 2011). For example, both intrinsic and Rho-dependent terminators can regulate gene expression when located in 5' UTRs of genes (Richardson, 2002; Naville and Gautheret, 2009; Santangelo and Artsimovitch, 2011). In these instances, terminator activity is typically controlled by an environmental cue [*e.g.* the presence of a metabolite or ion (Mandal and Breaker, 2004; Hollands *et al.*, 2012)]. More globally, Rho-dependent termination has been found to be important for maintaining chromosomal integrity (Washburn and Gottesman, 2011), inhibiting antisense transcription (Peters *et al.*, 2012), preventing genome-wide R-loops (DNA:RNA hybrid) (Leela *et al.*, 2013) and suppressing the expression of foreign DNA (Cardinale *et al.*, 2008; Menouni *et al.*, 2013).

1.4 RNA-binding proteins

1.4.1 RNA-binding proteins: mechanisms of action

Not surprisingly, RNA-binding proteins (RBPs) like Rho and the RNA chaperone Hfq, are important post-transcriptional regulators. Like sRNAs (Section 1.2.2.1), RBPs can influence their targets through a variety of different mechanisms, including modulating RNA elongation, stability and translatability. Rho is a classic example of a RBP that inhibits transcript elongation by directly displacing RNAP (see 1.3.1.1). Unlike Rho, other RBPs with this function typically act indirectly. For example, the *trp* RNA-binding attenuation protein (TRAP) binds to the 5' UTR of its target mRNA, stabilizing an intrinsic terminator and promoting termination (Babitzke, 2004). RNases represent another prominent class of RBPs that directly influence RNA stability (Arraiano *et al.*, 2010). These proteins recognize and cleave specific RNA sequences, ultimately promoting RNA degradation (Arraiano *et al.*, 2010; Assche *et al.*, 2015). RBPs can also regulate transcript stability indirectly by modifying the accessibility of their targets to RNases (Assche *et al.*, 2015). For instance, protein S1 prevents RNase E cleavage by binding to the AU-rich regions typically recognized by this RNase (Komarova *et al.*, 2005). Proteins can also affect RNA translatability by promoting or occluding ribosomes from binding to their target transcript (Assche *et al.*, 2015). An example of this is CsrA, which binds to the RBS of its mRNA targets, preventing their translation (Liu *et al.*, 1995; Gutiérrez *et al.*, 2005). Finally, these regulatory effects can also

be mediated by RNA chaperones, which facilitate the interaction of mRNAs with other proteins or sRNAs (Assche *et al.*, 2015).

1.4.2 Hfq and other RNA chaperones

The best-studied RNA chaperone is Hfq, which plays a critical role in sRNA-mediated regulation in many bacteria (Sauer, 2013). Hfq forms a homo-hexameric ring having three RNA binding surfaces: a proximal face, a distal face and a lateral site (Vogel and Luisi, 2011; Sauer *et al.*, 2012). Recent work has suggested that sRNAs can associate with both the proximal face and lateral site of the Hfq hexamer, while the opposite (distal) face is implicated in mRNA binding (Sauer, 2013). The current model of Hfq activity is that canonical (U-tailed) sRNA intrinsic terminators associate with the proximal face of Hfq (Sauer and Weichenrieder, 2011), the U-rich sRNA effector sequences are stabilized by interactions with the lateral site (Sauer *et al.*, 2012) and the distal face interacts with A-rich regions of mRNAs (Soper and Woodson, 2008; Salim and Feig, 2010; Sauer, 2013).

Initial *in vitro* experiments demonstrating that Hfq facilitates sRNA-mRNA interactions (Møller *et al.*, 2002) were complemented by *in vivo* work (in *E. coli*) showing that *trans*-encoded sRNAs require Hfq to regulate their targets (Waters and Storz, 2009). As a result, several – not mutually exclusive – mechanisms have been proposed to explain Hfq's ability to facilitate sRNA-mRNA interactions: (1) Hfq binds both the sRNA and its target mRNA simultaneously, bringing them close enough to interact (Soper *et al.*, 2011) and (2) Hfq remodels the secondary structure of the sRNA or the mRNA (or both) to promote duplex formation (Soper *et al.*, 2011; Vogel and Luisi, 2011). Often after formation, sRNA-mRNAs complexes are targeted for rapid degradation by the Hfq-associated endoribonuclease, RNase E (Morita *et al.*, 2005; Bandyra *et al.*, 2012).

An Hfq-encoding gene has been identified in many (~50%) bacterial genomes that have been sequenced (Sun *et al.*, 2002; Sobrero and Valverde, 2012). While Hfq is required for the activity of *trans*-encoded ncRNAs in organisms with high (≥ 50%) GC-content (e.g. *Caulobacter crescentus,* 67%; *Pseudomonas aeruginosa,* 66%; *E. coli,* 50%), it appears to be dispensable for sRNA function in organisms with lower GC-contents (e.g. *Bacillus subtilis,* 43%; *Staphylococcus aureus,* 32%) (Jousselin *et al.*, 2009). The streptomycetes and their actinobacterial relatives appear to be an exception, having a high genomic GC content (>70%), but no known Hfq homolog (Sun *et al.*, 2002; Swiercz *et al.*, 2008; Jousselin *et al.*, 2009).

Notably, the highly conserved RNA-binding endoribonuclease, YbeY, appears to affect gene expression in *Sinorhizobium meliloti* in a similar manner to Hfq, suggesting a possible role in sRNA-mediated regulation (Pandey *et al.*, 2014; Vercruysse *et al.*, 2014). While other putative RNA chaperones have been identified [*e.g.* FinO (Arthur *et al.*, 2003) and FbpB (Smaldone *et al.*, 2012)], these typically have a small number of defined binding partners (*e.g.* FinP and FsrA, respectively), unlike Hfq and potentially, YbeY.

1.5 Aims of this work

Few transcription terminators have been characterized in *Streptomyces*. To address this fundamental gap in our understanding of this essential process, we employed *in silico* approaches to identify putative intrinsic terminators in these bacteria, as described in Chapter 3. Several identified terminators were prioritized for *in vivo* characterization, with the goal being to gain insight into the terminator features required for effective termination. A secondary goal of these experiments was to identify strong terminators, which could be exploited in synthetic constructs.

To address the lack of a known RNA chaperone, and general absence of target information for most sRNAs in the streptomycetes, we set out to probe sRNA-mediated regulation in Chapter 4. We designed an RNA affinity purification system that will allow us identify any proteins or RNAs that interact with a given sRNA transcript. Of particular interest are RNA chaperones, which would provide insight into how *trans*-encoded sRNAs function in *Streptomyces* bacteria.

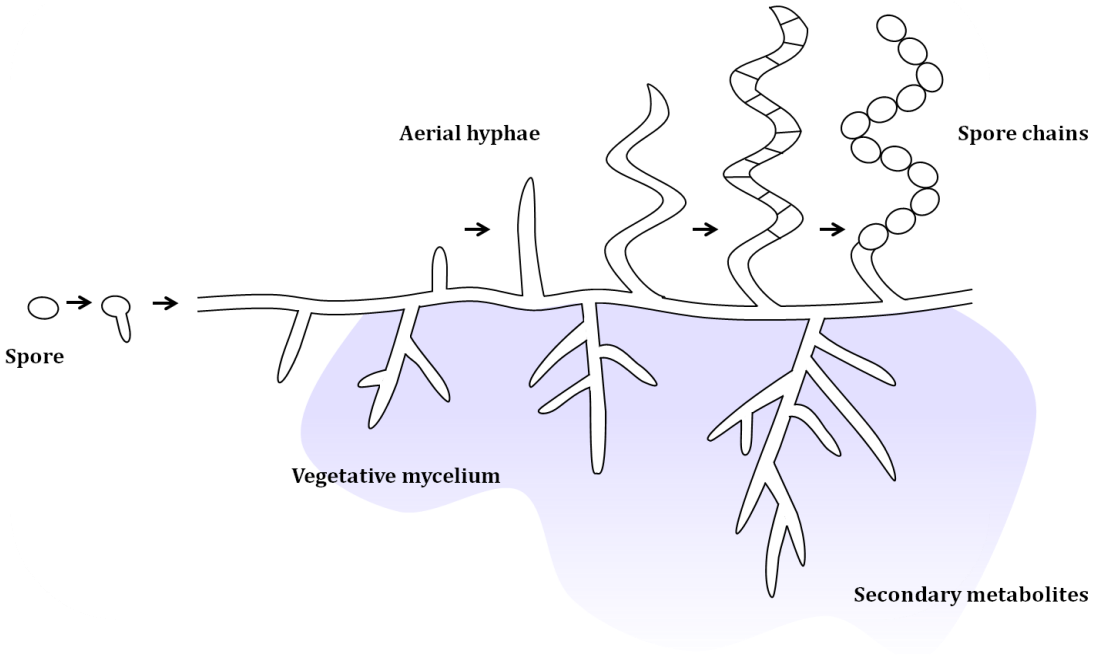

Figure 1.1 *Streptomyces* life cycle. In favourable conditions, a spore germinates and produces germ tubes, which grow by tip extension and branching to form a vegetative mycelium. When conditions become unfavourable, aerial hyphae extend from these vegetative cells into the air. The transition from vegetative growth to aerial development coincides with the production of many secondary metabolites (blue). Aerial hyphae then subdivide into pre-spore compartments, which eventually mature into reproductive spores. Adapted from (Swiercz and Elliot, 2012).

Chapter 2: Materials and Methods

2.1 Bacterial strains and culturing

2.1.1 Bacterial strains and plasmids

All bacterial (*E. coli* and *Streptomyces*) strains and plasmids used in this work are described in Tables 2.1 and 2.2. *Streptomyces* strains were stored as spore stocks at -20°C. *S. venezuelae* spores were harvested from cultures grown on maltose-yeast extract-malt extract (MYM) agar medium (Yang *et al.*, 2001) for 2-3 days, using a protocol adapted from (Kieser *et al.*, 2000). Specifically, spores were transferred to 10 mL of sterile water, creating a spore suspension which was sonicated in a water bath and then passed through a cotton syringe filter. Filtered spores were pelleted by centrifugation, and then resuspended in 40% glycerol. *S. coelicolor* spores stocks were created, as just described, from cultures that had been grown for 4-5 days on mannitol-soy flour (MS) agar medium (Hobbs *et al.*, 1989) overlaid with cellophane discs.

E. coli strains were stored at -80°C as glycerol stocks, which were created by mixing equal volumes of liquid overnight cultures with 40% glycerol.

2.1.2 Bacterial growth conditions

S. venezuelae strains were routinely grown at 30°C in MYM liquid medium or on solid MYM agar; however, MS solid agar was used to culture *S. venezuelae* during conjugations with *E. coli*. *S. coelicolor* strains were typically grown at 30°C on MS agar medium or in a 1:1 mixture of YEME (yeast extract-malt) and TSB (tryptone soya broth) liquid medium (Kieser *et al.*, 2000). Liquid cultures were grown with (*S. coelicolor*) or without (*S. venezuelae*) a steel spring, in an incubator while shaking at 200 rpm. When required, media was supplemented with the following antibiotics at the given final concentrations (μg/mL): apramycin (50), hygromycin B (50), nalidixic acid (25) and thiostrepton (50).

E. coli strains were primarily grown in Luria Bertani (LB) liquid medium or on LB agar medium (Miller, 1972). When using salt-sensitive antibiotics (*e.g.* hygromycin B), *E. coli* strains were cultured in Super Optimal Broth (SOB) (Hanahan, 1983) or on Difco nutrient agar (DNA). All *E. coli* strains were grown at 37°C, unless otherwise indicated. Where appropriate, antibiotics were added to the media at the following final concentrations (μg/mL): ampicillin (100), apramycin (50), chloramphenicol (25), hygromycin B (50) and kanamycin (50).

2.2 Oligonucleotides

All oligonucleotides used in this work are detailed in Table 2.3.

2.3 Molecular biology techniques

2.3.1 Introducing DNA into and extracting DNA from *E. coli*

DNA was introduced into *E. coli* by transformation or electroporation, as described by (Sambrook and Russell, 2001). For transformations, DNA was incubated on ice with either chemically competent cells [prepared using calcium chloride, as per (Sambrook and Russell, 2001)] or Subcloning Efficiency™ DH5α™ competent cells (Invitrogen). Competent cells were heat shocked briefly 20 to 60 seconds (sec) at 42°C (chemically competent cells) or 37°C (commercially competent cells), before being rapidly cooled on ice. Cells were grown in liquid medium without selection for 60 minutes (min), before being spread onto solid media containing the appropriate antibiotics. For electroporations, DNA was mixed with fresh electrocompetent *E. coli* cells, prepared as described in (Gust *et al.*, 2003). Mixtures were transferred to 2 mm electroporation cuvettes (VWR), which then received a 2.5 kV pulse from a Micropulser™ (Bio-Rad) electroporator as per (Gust *et al.*, 2003). Following electroporation, cells were treated as described above (short growth in non-selective medium before growth on solid, selective medium).

Plasmid DNA was extracted from *E. coli* using a standard alkaline lysis (with SDS) protocol, followed by phenol:chloroform extraction and ethanol precipitation as detailed in (Sambrook and Russell, 2001). Following precipitation, nucleic acids were resuspended in nuclease free water. Contaminating RNA was removed by a subsequent RNase A (Roche) treatment. To obtain sequencing-grade DNA, plasmids were isolated from *E. coli* using the PureLink Quick Plasmid Miniprep kit (Invitrogen) or the GeneJET Plasmid Miniprep kit (Thermo Scientific), as per the manufacturer's instructions.

2.3.2 Introducing DNA into and extracting DNA from *Streptomyces*

All plasmid DNA was introduced into *Streptomyces* by conjugation from *E. coli*. Specifically, plasmid DNA was first electroporated (see section 2.3.1) into a methylation-deficient strain of *E. coli* (ET12567), which harbours the mobilizing plasmid, pUZ8002. Plasmids introduced into ET12567/pUZ8002 were then transferred to *Streptomyces* by conjugation, as described by (Gust *et al.*, 2003). For *S. coelicolor* conjugations, spores were heat shocked for 10 min at 50°C in 2 × YT (yeast extract-tryptone) to promote germination. Pre-germinated spores and exponentially growing *E. coli* cells were then mixed, plated onto MS agar medium (lacking $MgCl_2$) and incubated at 30°C (Gust *et al.*, 2003). After 16-20 hours (h) of incubation at 30°C, conjugations were overlaid with an antibiotic solution (*i.e.* water containing nalidixic acid and either apramycin or hygromycin B) to select for streptomycetes that had received the plasmid. Conjugation into *S. venezuelae* was carried out as just described, except that spores were not pre-germinated and conjugations were incubated at room temperature overnight.

In order to isolate *S. venezuelae* genomic DNA, a single colony (isolated from solid MYM agar) was inoculated into 10 mL of MYM liquid media and grown for 24 h. Genomic DNA was isolated from 1 mL of this culture, using the Bacterial Genomic DNA Isolation Kit

(Norgen Biotek), as per the manufacturer's instructions. Genomic DNA was eluted in nuclease free water.

2.3.3 Molecular cloning

2.3.3.1 PCR amplification and purification of DNA

Routine polymerase chain reactions (PCRs) (*i.e.* PCR checks) were carried out with *Taq* DNA polymerase (GeneDirex), using the reaction components and cycling conditions outlined in Tables 2.4 and 2.5. When high fidelity amplification was required (*e.g.* gene amplification), PCRs were carried out using Phusion® High-Fidelity DNA polymerase (New England Biolabs), as per Tables 2.6 and 2.7.

PCR amplification products were typically purified using the PureLink® PCR purification kit (Invitrogen), as per the manufacturer's protocol. When more than one amplification (or digestion) product was produced, the DNA fragment of interest was purified using either the PureLink® quick gel extraction kit (Invitrogen) or the E.Z.N.A.® gel extraction kit (OMEGA bio-tek).

2.3.3.2 Restriction digestion of DNA

Digestions were typically performed in 50 µL reactions, as described by Thermo Scientific for Fast-Digest™ enzymes (*Bgl*II, *Eco*RI, *Hin*dIII, *Nde*I, and *Xba*I) and New England Biolabs for CutSmart™ enzymes (*Bam*HI-HF®, *Kpn*I-HF®, *Nhe*I-HF®, *Spe*I, *Xba*I and *Xho*I).

2.3.3.3 Phosphorylation and dephosphorylation of DNA

DNA fragments or oligonucleotides were phosphorylated using T4 kinase (Invitrogen), as instructed by the manufacturer. Phosphorylated DNA was recovered using the PureLink® PCR purification kit (Invitrogen) or by phenol:chloroform extraction and ethanol precipitation. Digested plasmid DNA was dephosphorylated using Calf Intestinal Alkaline Phosphatase (Invitrogen), as per the manufacturer's instructions, and then recovered as described above.

2.3.3.4 Ligation of DNA

DNA ligations were performed using T4 DNA ligase (Invitrogen) in 10 µL reactions, as described by the manufacturer. Phosphorylated inserts were mixed with dephosphorylated vectors at a ratio of 3:1, for cohesive ends, or 5:1, for blunt ends. Ligation reactions were incubated at 16°C overnight, before being transformed into Subcloning Efficiency™ DH5α™ cells (see section 2.3.1).

2.3.4 Creating a fluorescent terminator reporter construct

The monomeric Cherry (mCherry) and monomeric, superfolder green fluorescent protein (msfGFP) genes were introduced into the pIJ6902 integrative vector (Huang *et al.*, 2005) to create the fluorescent terminator reporter construct, pMC251 (Figure 2.1) Specifically, *Streptomyces* codon-optimized *mCherry* was PCR amplified from pUC57 + *mCherry* using primers mCherry F NdeI and mCherry R XbaI (Table 2.3). The *Nde*I and *Xba*I

sites introduced by these primers were used to clone the amplified fragment into the corresponding sites of pIJ6902, downstream of the thiostrepton-inducible promoter P_{tipA}. Similarly, codon-optimized *msfGFP* was amplified from pUC57 + *msfGFP* using primers msfGFP F EcoRI RBS and msfGFP R BglII (Table 2.3). The amplified product was digested with *Eco*RI and *Bgl*II and then ligated into the same sites downstream of *mCherry* in pIJ6902. The forward primer (msfGFP F EcoRI RBS) also included a short sequence encoding a RBS, in order to introduce a RBS immediately upstream of *msfGFP*. The final reporter construct was verified by sequencing with the M13 forward, TipA, and msfGFP R sequencing primers (Table 2.3).

Terminators were generated by oligonucleotide annealing and the resulting double stranded DNA was then cloned into pMC251, as follows: 200 pmol of each overlapping oligonucleotide (Table 2.3) was first phosphorylated separately, as described in 2.3.3.3. After phosphorylation, complementary oligonucleotides (1 pmol each) were mixed and then annealed in a thermocycler heated to 95°C for 2 min and then slow-cooled to 4°C (at a rate of 1.4°C/min). The resulting DNA fragments, which contained *Bam*HI- and *Xba*I-compatible overhangs, were cloned into pMC251 digested with *Bam*HI/*Xba*I and dephosphorylated. All terminator-containing constructs were sequenced with the msfGFP R sequencing primer, before being transferred into *E. coli* ET12567 via electroporation (see section 2.3.1). To facilitate P_{tipA}-driven gene expression an integrative plasmid carrying *tipA* (pMS82-*tipA*) was first conjugated into wild-type *S. venezuelae*, as described in 2.3.1. Each terminator reporter construct (including empty pMC251) was then conjugated into this *S. venezuelae* strain and wild-type *S. coelicolor*. mCherry and msfGFP fluorescence for each of these *Streptomyces* strains was followed using fluorescence spectroscopy (please see below, in section 2.5.3).

2.3.5 Creating tagged sRNAs

2.3.5.1 Creating tRSA-tagged sRNAs

To tag each sRNA of interest, the tRNA-scaffolded streptavidin aptamer (tRSA) was PCR amplified from pCDNA3-tRSA (Iioka *et al.*, 2011) using tRSA F EcoRI and tRSA R XhoI oligonucleotides, which were phosphorylated prior to amplification. The resulting PCR product included the tRSA tag flanked by an *Eco*RI site and an *Xho*I site. This fragment was digested with *Eco*RI and then directionally cloned into the *Eco*RI/*Eco*RV sites downstream of the constitutive *ermE** promoter (P_{ermE*}) in the sRNA overexpression vector, pMC500 (Elliot Lab). This vector also contains a terminator sequence downstream of the multiple cloning site (MCS), which will terminate transcription of the tRSA tag (or any tRSA-tagged sRNAs lacking an endogenous terminator).

sRNAs of interest were then introduced into pMC255 (Figure 2.2), the tRSA aptamer vector created as described above, using traditional ligation-based cloning. In particular, sRNAs were PCR amplified, along with 108-161 nucleotides of the downstream sequence (to capture the native sRNA terminator), from *S. venezuelae* genomic DNA. The primers used to amplify each sRNA (Table 2.3) contained *Xho*I and *Bam*HI cleavage sites to facilitate cloning

into the corresponding sites of pMC255, immediately downstream of the tRSA tag (note that the *Xho*I site was introduced into pMC500 along with the tRSA tag). All pMC255 derivatives were verified by sequencing with the M13 forward primer (Table 2.3).

Each P_{ermE*}-*tRSA*-(*sRNA*) construct was removed from the corresponding pMC500 intermediate by *Bgl*II digestion, and inserted into the *Bam*HI site of the integrating vector pIJ82. The resulting vectors, termed pMC260-264 (Table 2.2), were passed through *E. coli* ET12567 prior to introduction into wild-type *S. venezuelae*. Chromosomal integration of each pMC260 derivative was confirmed by PCR using genomic DNA isolated from each strain. *In vivo* expression of each tRSA-tagged sRNA was tested by northern blotting, as detailed in 2.4.2.

2.3.5.2 Creating MS2-tagged sRNAs

A MS2 aptamer vector and a novel sRNA overexpression vector were generated from the multipurpose vector, pMC600 (Figure 2.3). This vector, derived from pUC57-mini (GenScript), includes the strong, constitutive SF14 promoter (Labes *et al.*, 1997) upstream of a theophylline riboswitch (Rudolph *et al.*, 2013), two tandem MS2 tags (Said *et al.*, 2009) and a MCS. To remove the theophylline riboswitch, pMC600 was digested with *Xba*I and *Spe*I and then re-ligated following gel extraction. The resulting MS2 aptamer vector (pMC601) contained only the 2 × MS2 tag and the MCS downstream of the SF14 promoter. The novel sRNA overexpression construct (pMC602) included only the SF14 promoter and MCS, and was created as above by digesting (with *Xba*I and *Nhe*I) and re-ligating pMC600.

Each sRNA of interest was PCR amplified from the corresponding pMC255 derivative (pMC256-259), using the same sRNA-specific primers described in 2.3.5.1 (Table 2.3). All amplified PCR products were phosphorylated and then cloned into *Eco*RV-digested pMC601, in order to introduce the 2 × MS2 tag to the 5' end of each sRNA. Control vectors (lacking the MS2 tag) were created by ligating the same phosphorylated sRNA fragments into pMC602 digested with *Eco*RV and dephosphorylated. When PCR products were introduced into pMC601/pMC602 in tandem, the extra sRNA fragment was removed by digestion with *Xho*I, with the vector containing the now single insert subsequently being re-ligated back together.

Following sRNA introduction, all pMC601/pMC602 derivatives were digested with *Kpn*I and *Hind*III to liberate each P_{SF14}-(*MS2*)-(*sRNA*) construct. These fragments were cloned into the corresponding sites of the integrating vector pMS82 (Gregory *et al.*, 2003), creating pMC266-269 and pMC271-274. All vectors were conjugated from *E. coli* ET12567 into wild-type *S. venezuelae*.

2.4 RNA techniques

2.4.1 Isolating RNA from *S. venezuelae*

S. venezuelae overnight cultures were subcultured into 300 mL liquid MYM medium to an OD_{600} of 0.05. These cultures were grown in 2 L flasks at 30°C, with shaking. Cells were

harvested during vegetative growth and fragmentation (8 and 18 h post-inoculation, respectively) and then stored at -80°C until isolation. RNA was isolated from the frozen cell pellets using a guanidium thiocyanate-based protocol (Chomcyzynski and Sacchi, 1987), as described by Moody *et al.* (2013). Extensive bead-beating in a guanidium thiocyanate solution was used to lyse the cells. Following lysis, nucleic acids were recovered by phenol-chloroform extraction (repeated three times) and precipitation in sodium acetate and isopropanol at -20°C overnight. Nucleic acids were pelleted, washed with 70% ethanol and then re-suspended in nuclease free water. Co-extracted DNA was removed using Turbo DNase (Ambion). RNA quantity was determined using a NanoDrop 1000 spectrophotometer, while quality was assessed by agarose gel electrophoresis.

2.4.2 RNA detection by northern blotting

Northern blotting was carried out as described previously (Haiser *et al.*, 2008; Swiercz *et al.*, 2008; Moody *et al.*, 2013). Briefly, 10 µg of total *S. venezuelae* RNA was separated on a denaturing polyacrylamide gel (6%$_{w/v}$ Gene-Gel, Bioshop), before being transferred to a nylon Zeta-Probe membrane (Bio-Rad) with a Trans-Blot® semi-dry Transfer apparatus (BioRad). RNA was chemically cross-linked to the membrane using a 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide cross-linking solution (Pall and Hamilton, 2008), as outlined by Moody *et al.* (2013). Membranes were then pre-hybridized with ULTRAhyb®-Oligo buffer (Ambion) in a hybridization oven (HB-1D, Techne) for 30 min at 42°C. Simultaneously, oligonucleotides (2 pmol each; Table 2.3) were 5'-end-labelled with [γ-^{32}P] ATP (Perkin Elmer) using T4 kinase (Invitrogen), as per the manufacturer's instructions. Labeled probes were recovered using the MinElute Reaction Cleanup Kit (Qiagen) and then hybridized with the membrane overnight at 42°C. To remove unbound probe, membranes were washed two or three times with low-stringency buffer (300 mM sodium chloride, 30 mM trisodium citrate dehydrate, 0.1%$_{w/v}$ sodium dodecyl sulfate) for 5 min at 42°C, before being exposed to a storage phosphor screen (Molecular Dynamics) and imaged using a phosphorimager (Storm 820, Molecular Dynamics). Membranes were stripped with high-stringency buffer (30 mM sodium chloride, 3 mM trisodium citrate dehydrate, 0.1%$_{w/v}$ sodium dodecyl sulfate) at 70°C for at least 10 min in a hybridization oven. To ensure that all radiolabelled probe had been removed, membranes were re-exposed to a phosphor screen before re-probing. A 25 bp DNA ladder (Invitrogen), detected with 5'-end labelled complementary DNA, was also included to assess transcript size.

2.4.3 *In vitro* transcription of RNA

In vitro transcription of tRSA-tagged sRNAs was carried out using the MEGAshortscript™ T7 Kit (Ambion), as described by the manufacturer. Template DNA was PCR-amplified from pMC255-259, using T7 + tRSA aptamer (containing the T7 promoter sequence) and sRNA-specific primers (Table 2.3). Transcription reactions were assembled with 25-85 nM of template DNA (purified by agarose gel extraction), which was degraded with 2-8 U of TURBO DNase (Ambion) following 4 h of incubation at 37°C. Synthesized RNA was recovered by phenol:chloroform extraction and isopropanol precipitation (Sambrook

and Russell, 2001) Recovered RNA was quantified using a NanoDrop 1000 spectrophotometer, whereas RNA purity and quality was assessed by polyacrylamide gel electrophoresis on a denaturing 6%$_{w/v}$ GeneGel (BioShop). To confirm the size of each synthesized transcript, RNA was mixed with xylene cyanol loading buffer (Sambrook and Russell, 2001) and then separated on a 2% agarose gel alongside template DNA (diluted in the same loading dye).

Following synthesis, RNA transcripts were re-folded using four distinct protocols [adapted from (Iioka *et al.*, 2011), (Tahiri-Alaoui *et al.*, 2002), (Leppek and Stoecklin, 2013) and (Hnilicova *et al.*, 2014)]. As described by Iioka *et al.* (2011), 10 µg of RNA was re-folded in a 10 µL reaction containing 10 mM $MgCl_2$ and 10 mM HEPES (pH 7.0). The reaction was heated in a thermocycler to 65°C for 5 min and then slow cooled to 20°C over a period of 45 min. In contrast, the protocol adapted from Tahiri-Alaoui *et al.* (2002) involved denaturing 5 µg of RNA in water at 95°C for 3 min and then re-folding the RNA in binding buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 50 mM KCl, and 10 mM $MgCl_2$) at 20°C for 20 min. Synthesized transcripts were also renatured in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1.5 mM $MgCl_2$ and 2 mM DTT, as per (Leppek and Stoecklin, 2013). Here, 5 µg RNA was incubated at 56°C for 5 min, 37°C for 10 min and then 23°C for 5 min. Finally, 5 µg of synthesized RNA was heated to 90°C for 2 min and then incubated at 23°C for 20 min in folding buffer (100 mM KCl, 10 mM $MgCl_2$, 10 mM Tris-HCl pH 7.0), as detailed in (Hnilicova *et al.*, 2014).

To test whether the *in vitro* transcribed tRSA-tagged sRNAs had re-folded correctly and could bind streptavidin, a streptavidin-bead binding assay was performed. Specifically, RNA re-folding reactions (described above) were adjusted to 85 µL using TEN_{100} buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM NaCl). To prepare the streptavidin-coated magnetic particles (Roche), 30 µL of the particle suspension (Roche, 10 mg particles per mL) were transferred to a 1.5 mL microfuge tube and then pelleted using a 6 Tube Magnetic Stand (Ambion). The streptavidin particles were subsequently washed 3 times with 100 µL of TEN_{100} buffer and resuspended in 75 µL (60-120 ng/µL) of the RNA mixture. The RNA-particle solution was incubated at room temperature, with gentle shaking for 1 h. After 10 and 30 min of incubation, the particles were pelleted and 10 µL of the supernatant was collected. The particles were then resuspended in the remaining supernatant. At 60 min, all of the supernatant was removed and the particles were re-suspended in 30 µL of fresh TEN_{100} buffer before being heated to 95°C for 5 min. Streptavidin-bound RNA was recovered in the supernatant following particle boiling. Supernatant fractions (10 µL each) – collected each time the particles were pelleted – were mixed with xylene cyanol loading buffer and then separated on a 2% agarose gel. RNA was visualized using ethidium bromide staining. Biotinylated, *chpH* promoter DNA (amplified using SVEN CHPH BIOTIN 1 and SVEN CHPH PROM-2, Table 2.3) was included as a positive control for streptavidin binding. In this case, only 500 ng of purified DNA was incubated with 40 µL of streptavidin magnetic particles.

2.5 Protein techniques

2.5.1 Preparing crude cell extracts for *S. coelicolor* and *S. venezuelae*

Streptomyces strains were grown in the appropriate liquid medium at 30°C for up to 2 days before cells were harvested (by centrifugation at 2,218 × g) and the cell pellets were then stored at -80°C prior to lysis. Cell pellets were resuspended in filter sterilized lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM KCl, 50 mM NaCl, 1 mM DTT) containing lysozyme (1 mg/mL) and EDTA-free protease inhibitor [1 cOmplete™ tablet (Roche) per 50 mL] and then incubated at 37°C for 1 h. Cells were lysed by sonication using the Sonifier Cell Disruptor 350 (Branson) at 40% duty for 4-10 cycles of 6 seconds (s) of sonication, followed by 10 s on ice. Lysates were then cleared by centrifugation at ~20,200 × g for 30 min at 4°C.

Crude cell extracts were separated by SDS polyacrylamide gel electrophoresis (SDS-PAGE) and then stained with Coomassie Brilliant Blue R-250 for visualization. Bradford assays (using bovine serum albumin as the standard control) were used to measure total protein content (Bradford, 1976). When lysates were too dilute, cell extracts were concentrated by centrifugation (at 2,218 × g) in an Amicon Ultra-4/15 centrifugal filter (Millipore) that had a cut-off of either 3 kDa.

2.5.2 Fluorescence (mCherry and msfGFP) reporter assays

Cultures of *S. venezuelae* and *S. coelicolor* reporter strains were grown in 10 mL of the appropriate liquid medium (*i.e.* MYM or YEME:TSB, respectively) in a universal vial at 30°C overnight. These overnight cultures were then subcultured into 10-100 mL of the same liquid medium (in a universal vial, a 250 mL flask or a 500 mL flask) and grown at 30°C. To induce transcription of the *tipA* promoter, exponentially growing cultures (typically 4-20 h post-sub-culture) were induced with thiostrepton (added to a final concentration of 50 µg/mL). Cultures were incubated at 30°C for an additional 30 min, before 150 µL of each culture was transferred in triplicate into a black, Microfluor™ 96-well plate (Thermo Scientific). Fluorescence was measured using a Cytation Multi-mode Reader (Biotek), as follows: msfGFP fluorescence was measured at 530 nm (after excitation at 488 nm) immediately before mCherry fluorescence was measured at 615 nm (following excitation at 589 nm). Raw fluorescence measurements were normalized to OD_{600} (*S. venezuelae*) or dry cell weight (*S. coelicolor*) and then compared to the negative control strains, carrying empty pIJ6902. *S. coelicolor* expressing eGFP from the constitutive *ermE** promoter [M145 + pIJ8660-*ermE** (Sexton, unpublished)] was included as positive control for GFP fluorescence.

Alternatively, *Streptomyces* fluorescent reporter strains were grown in liquid medium for 8-40 h before thiostrepton was added, as above. Following 2-4 h of induction, cells were harvested and then stored at -80°C, before being lysed by sonication (as described in 2.5.1). One hundred and fifty µL of each resulting cell lysate was transferred in triplicate into a black, 96-well plate to measure fluorescence. Here, fluorescence measurements were collected at 610 nm for mCherry (587 nm excitation) and then at 510

nm for msfGFP (485 nm excitiation). Fluorescence was normalized to lysate protein content, measured using a standard Bradford assay (see 2.5.1).

2.5.3 Affinity purification assays with *in vitro* transcribed sRNAs

Following *in vitro* transcription, tRSA-tagged sRNAs (10 µg each) were re-folded in a 25 µL reaction using the protocol adapted from (Hnilicova *et al.*, 2014), as described in 2.4.3. Renatured RNA was diluted with 285 µL lysis buffer containing RNaseOUT™ recombinant ribonuclease inhibitor (200 U per mL buffer; Invitrogen). Streptavidin-coated magnetic particles were washed (see 2.4.3) twice with 100 µL of lysis buffer with RNaseOUT™, before being resuspended in 300 µL of the above RNA solution. The tRSA-tagged sRNAs were tethered to the streptavidin particles by gentle mixing at 4°C for 1 h. To remove all unbound RNA, the particles were washed twice with fresh lysis buffer containing RNaseOUT™.

Simultaneously, *S. venezuelae* cell lysates were generated from 6 and 16 h liquid cultures, as outlined in 2.5.1. To block endogenous streptavidin-binding proteins, lysates were incubated with egg white avidin (10 µg per mg lysate protein; EMD Millipore) for 30 min at 4°C with gentle agitation. After blocking, lysates were centrifuged (~20,200 × g for 10 min) before the supernatant was mixed with RNaseOUT™ (200 U per mL lysate).

The RNA-coated streptavidin particles were resuspended in the final pre-cleared lysates, and then incubated at 4°C with shaking for 1.5 h. Particles were washed five times with 50 µL of lysis buffer with RNaseOUT™, re-suspended in 24 µL SDS loading buffer (St-Onge *et al.*, 2015) and then heated to 95°C for 5 min. Elution (10 µL) and wash (20 µL) fractions, diluted in SDS loading buffer, were separated on a 12% SDS-PAGE and then visualized by Coomassie staining. As a control, the affinity purification protocol was carried out with streptavidin particles coated in biotinylated *chpH* promoter DNA [prepared as above with 1.5 µg of amplified DNA (see 2.4.3)] and nucleic acid-free particles.

2.6 Bioinformatic analyses

2.6.1 Identifying and characterizing potential transcription termination sites

Modified BED (Browser Extensible Data) files, which contain the read depth of each nucleotide on the positive and negative strand for two distinct RNA libraries (a full-length transcript library and short transcript library), were obtained for *S. avermitilis*, *S. coelicolor* and *S. venezuelae* (Moody *et al.*, 2013). A custom Perl script was written to screen the full-length transcript library of each species for drastic decreases in transcript levels (Appendix A). In particular, the average read depth was calculated for 10 adjacent nucleotides, and then compared to the average read-depth of the 10 nucleotides upstream (Figure 2.4). A termination signal was identified if the average read depth had decreased by at least 2-fold within the 20 nucleotide window. Every possible 20 nt window in each genome was screened; however, only terminators separated by 20 nt were considered for downstream analysis. In cases where more than one terminator was identified in a particular 20 nt window, only the terminator with the smallest genomic position was analyzed. To avoid identifying terminators in lowly transcribed or extremely highly expressed [*e.g.* ribosomal

RNA (rRNA)] regions, upstream averages were required to be above the average coverage per nucleotide of coding regions (min), but below the maximum (max). These parameters were calculated directly from modified BEDTools (Quinlan and Hall, 2010) output files (Moody *et al.*, 2013) for each species: *S. avermitilis* (min: 25, max: 4300), *S. coelicolor* (min: 50, max: 8200) and *S. venezuelae* (min: 20, max: 4600).

Identified terminators were classified as either intragenic (within a coding region), antisense (on the strand opposite of an annotated gene) or intergenic (between coding regions), using current genome annotations for each species (NC_003155.4, NC_003888.3 and FR845719.1) and a custom Perl script (Appendix B). To ensure the accuracy of the identification script, predicted termination sites were visualized manually using Integrated Genomics Viewer (IGV) Version 2.3 (Thorvaldsdóttir *et al.*, 2013). Interspecies comparisons of terminator locations (and genetic contexts) were carried out manually using IGV and StrepDB (housing genomic databases for each of these three species).

2.6.2 Comparing identified termination signals with predicted terminators

Terminators were predicted for *Streptomyces* species using three different algorithms: ARNold (Naville *et al.*, 2011), TransTermHP (Kingsford *et al.*, 2007) and WebGeSTer (Mitra *et al.*, 2011). Default parameters were used for each program, except for WebGeSTer where the following parameters were adjusted: stem length (4-30 nt), loop length (3-9 nt), maximum mismatch (3) and maximum distance from ORF (270 nt). RNA sequencing terminators (see 2.6.1) with WebGeSTer terminators predicted within 20 nt were identified manually. Based on these comparisons, identified terminators were annotated as either canonical (I-shaped) or non-canonical (*e.g.* L-shaped) based on their corresponding WebGeSTer terminator. Stem and tail sequence composition was calculated for each shared terminator, using custom Perl scripts (Appendix C and D), and then compared to genome-wide averages. Comparisons of structural features (*e.g.* stem length and bulb length) were made directly from WebGeSTer output files (Mitra *et al.*, 2011).

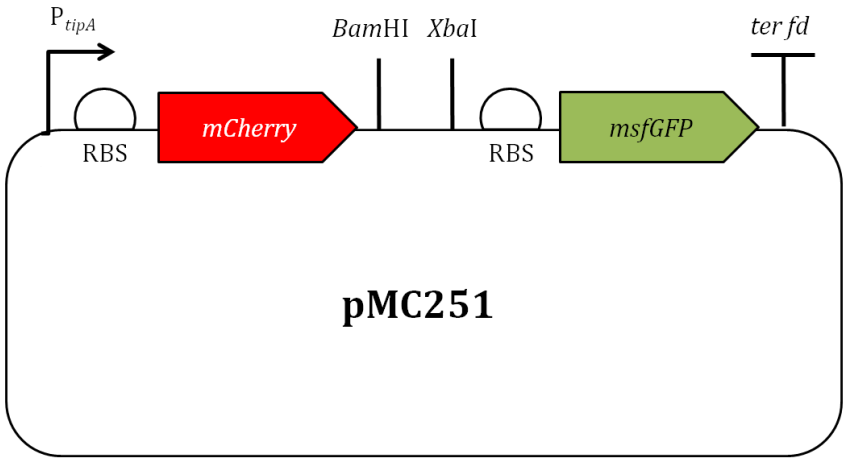

Figure 2.1 Schematic of pMC251. The fluorescent terminator reporter construct was derived from the integrating pIJ6902 vector (Huang *et al.*, 2005), which contains a MCS flanked by a thiostrepton-inducible promoter (P_{tipA}) and a strong transcription terminator from phage fd (*ter fd*). *Streptomyces* codon-optimized *mCherry* and *msfGFP* (including an upstream RBS) were amplified and then introduced into the MCS of pIJ6902, downstream of P_{tipA}. There are *Bam*HI and *Xba*I sites located between *mCherry* and *msfGFP,* which can be used to insert terminators of interest (generated by oligonucleotide annealing).

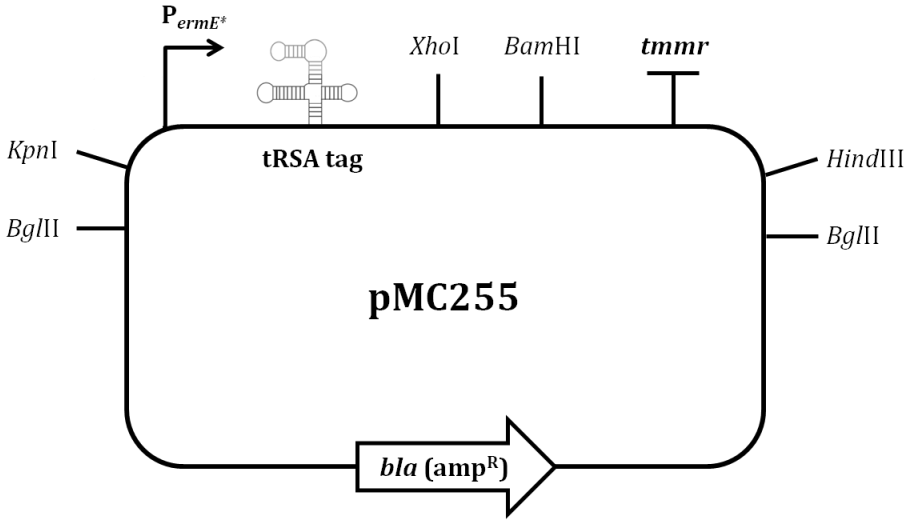

Figure 2.2 Schematic of pMC255. Construct used to tag sRNAs with the tRSA aptamer. Derived from pMC500 (Elliot lab), this vector contains the streptavidin aptamer expressed from P_{ermE*}. There are *Xho*I and *Bam*HI and sites located after the aptamer, which can be used to insert sRNAs of interest upstream of the *tmmr* terminator.

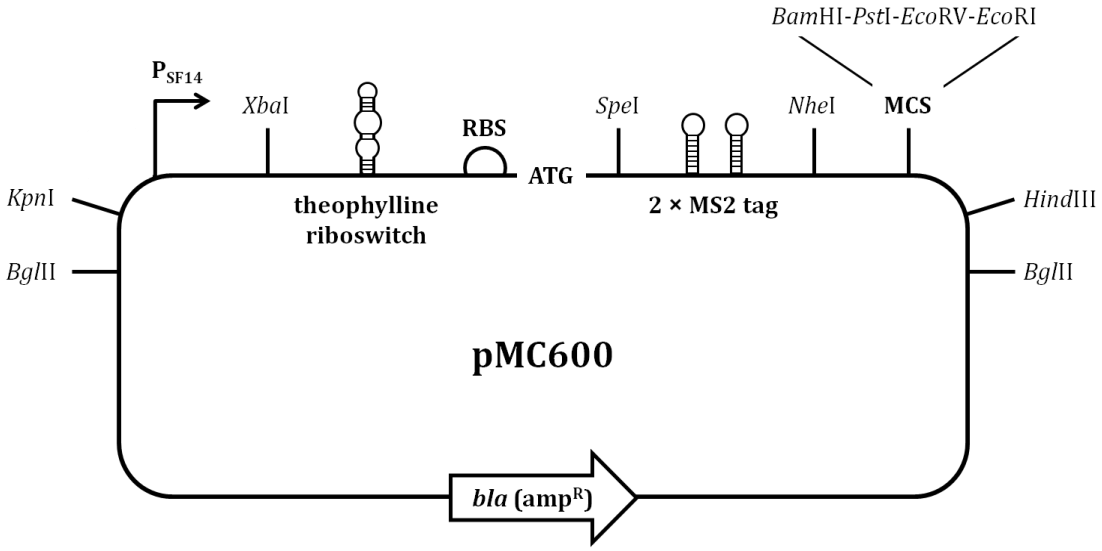

Figure 2.3 Map of pMC600. This multipurpose plasmid contains the strong SF14 promoter (P_{SF14}) (Labes *et al.*, 1997) upstream of a theophylline riboswitch (Rudolph *et al.*, 2013), two tandem MS2 tags (Said *et al.*, 2009) and an extensive MCS. Removal of the theophylline riboswitch (using the *Xba*I and *Spe*I sites) produces an MS2 aptamer cloning vector, which can be used to add the MS2 tag to the 5' end of any sRNA inserted into the MCS. Removal of the riboswitch and the MS2 tag (using the *Xba*I and *Nhe*I sites) creates an overexpression vector, which allows for constitutive gene expression from P_{SF14}. These constructs can be easily transferred into an integrative vector using the *Kpn*I/*Hind*III or *Bgl*II sites, which flank this region. pMC600 is derived from pUC57-mini (GenScript), which carries a gene (*bla*) conferring ampicillin (amp) resistance.

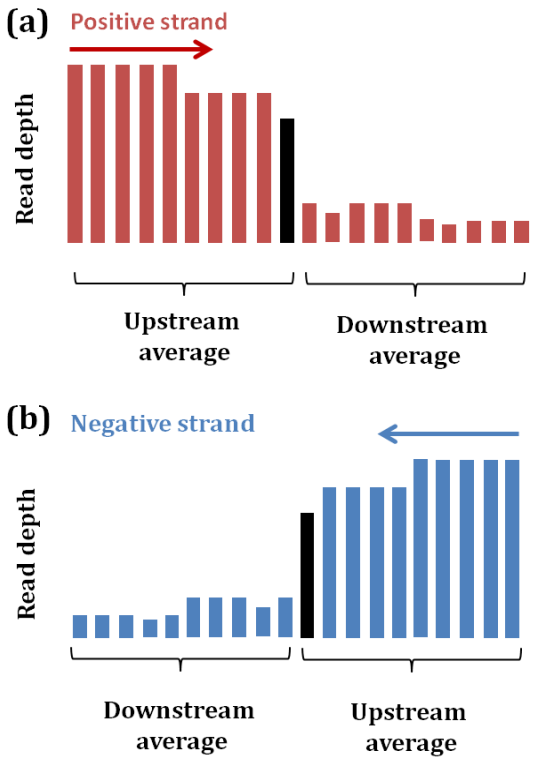

Figure 2.4. Identification of termination sites *in silico*. RNA sequencing data was screened for strong, well-defined terminators on both the (a) positive and (b) negative strand. Average read depth was calculated for 10 adjacent nucleotides (downstream average) and the 10 nucleotides upstream (upstream average). When the downstream average was 2-fold lower than the upstream average, a termination signal (black) was identified, as long as the upstream average was within the boundaries set for each species (see 2.6.1).

Table 2.1 Bacterial strains used in this work

Strain	Genotype, description, or use	Reference
E. coli		
DH5α	Plasmid construction and routine cloning	Invitrogen
ET12567/pUZ8002	Generation of methylation-free DNA and conjugation into *Streptomyces*	(MacNeil *et al.*, 1992; Paget *et al.*, 1999)
S. coelicolor		
M145	SCP1$^-$ SCP2$^-$	(Kieser *et al.*, 2000)
S. venezuelae		
ATCC 10712	wild-type	(Bibb *et al.*, 2012)

Table 2.2 Plasmids used in this work

Plasmid	Description or use	Reference
General vectors		
pUC57	General cloning plasmid	(Yanisch-Perron *et al.*, 1985)
pMC500	pUC57-derived cloning vector containing an *ermE** promoter flanked by transcription terminators	Elliot lab
pMC600	pUC57 derivative with SF14 promoter upstream of a theophylline riboswitch and two tandem MS2 tags	Elliot lab
pIJ82	Integrative plasmid vector (ϕC31 integrase and *attP*)	Gift from H. Kieser
pMS82	Integrative plasmid vector (ϕBT1 integrase and *attP*)	(Gregory *et al.*, 2003)
pIJ6902	Integrative cloning vector (ϕC31) containing a thiostrepton-inducible promoter (P_{tipA})	(Huang *et al.*, 2005)
pIJ8660	Integrative cloning vector (ϕC31) containing *eGFP* without a promoter	(Sun *et al.*, 1999)
Terminator reporter constructs		
pUC57 + *mCherry*	pUC57 + *mCherry*	Gift from M. Buttner
pUC57 + *msfGFP*	pUC57 + *msfGFP* (Q80R, V206K)	Gift from M. Buttner
pMC250	pIJ8660 + *ermE**; eGFP expressed from P_{ermE*}	(Sexton, unpublished)
pMS81 + *tipA*	Integrative vector (ϕBT1) with *tipA*; required for P_{tipA} – driven gene expression in *S. venezuelae*	Gift from L. Servín-González
pMC251	pIJ6902 + *mCherry* + RBS-*msfGFP*; empty fluorescent reporter construct	This work
pMC252	pMC251 + $Term_{sven\_0060}$; *sven_0060* terminator reporter	This work
pMC253	pMC251 + $Term_{sven\_2895}$; *sven_2895* terminator reporter	This work
pMC254	pMC251 + $Term_{sven\_4374}$; *sven_4374* terminator reporter	This work
Constructs used for tagging sRNAs		
pCDNA3-tRSA	Expression vector containing the tRSA (tRNA-scaffolded streptavidin aptamer) tag	(Iioka *et al.*, 2011)
pMC255	pMC500 with P_{ermE*} and tRSA tag	This work
pMC256	pMC255 + *svr1031*; expression of tRSA-tagged svr1031	This work
pMC257	pMC255 + *svr2416*; expression of tRSA-tagged svr2416	This work
pMC258	pMC255 + *svr3329*; expression of tRSA-tagged svr3329	This work
pMC259	pMC255 + *svr5279*; expression of tRSA-tagged svr5279	This work
pMC260	pIJ82 with P_{ermE*} and tRSA tag	This work
pMC261	pMC260 + *svr1031*; expression of tRSA-tagged svr1031	This work
pMC262	pMC260 + *svr2416*; expression of tRSA-tagged svr2416	This work
pMC263	pMC260 + *svr3329*; expression of tRSA-tagged svr3329	This work
pMC264	pMC260 + *svr5279*; expression of tRSA-tagged svr5279	This work
pMC601	pMC600 lacking the theophylline riboswitch (contains P_{SF14} and two tandem MS2 tags)	Elliot Lab
pMC265	pMS82 with SF14 promoter and MS2 tag	This work
pMC266	pMC265 + *svr1031*; expression of MS2-tagged svr1031	This work
pMC267	pMC265 + *svr2416*; expression of MS2-tagged svr2416	This work
pMC268	pMC265 + *svr3329*; expression of MS2-tagged svr3329	This work
pMC269	pMC265 + *svr5279*; expression of MS2-tagged svr5279	This work
pMC602	pMC600 lacking the theophylline riboswitch and the MS2 tag (only contains P_{SF14})	Elliot Lab
pMC270	pMS82 with P_{SF14}	This work
pMC271	pMC270 + *svr1031*; expression of untagged svr1031	This work
pMC272	pMC270 + *svr2416*; expression of untagged svr2416	This work
pMC273	pMC270 + *svr3329*; expression of untagged svr3329	This work
pMC274	pMC270 + *svr5279*; expression of untagged svr5279	This work

Table 2.3 Oligonucleotides used in this work

Name	Sequence (5' to 3')*	Description or use
General oligonucleotides		
M13 forward	GTAAAACGACGGCCAGT	General sequencing
M13 reverse	GCGGATAACAATTTCACACAGG	General sequencing
Oligonucleotides used in terminator-based studies		
mCherry F NdeI	CGTACGCATATGGTCTCCAAGGGCGAG	Cloning *mCherry* into pIJ6902
mCherry R XbaI	GACTACTCTAGATGGGTCTCACTTGTACAGCTC	Cloning *mCherry* into pIJ6902
msfGFP F EcoRI RBS	TGACTAGAATTC*AGAAGGGAG*TACGACGATGTCC AAGGGCGAGGAG	Cloning *msfGFP* into pIJ6902
msfGFP R BglII	TCACATAGATCTGGGTCTCACTTGTACAGCTCG	Cloning *msfGFP* into pIJ6902
msfGFP R sequencing	CTTGTCGGCGGTGATGTAGA	Sequencing terminator constructs
TipA	GGATCGGGGATCTGGGCTGAGGG	Sequencing terminator constructs
sven0060 term fwd	***CTAG***CGTCACGCGGGCCGCCGGCTGCCATGCCGG CGGCCCGCGGGCGTTTCCGGTTC	Creating dsDNA for *sven_0060* terminator
sven0060 term rev	***GATC***GAACCGGAAACGCCCGCGGGCCGCCGGCAT GGCAGCCGGCGGCCCGCGTGACG	Creating dsDNA for *sven_0060* terminator
sven2895 term fwd	***CTAG***AAGGGCCGTCCGCCATCCGTGGTGCGGACG GCCCTTCGCGTGTTTT	Creating dsDNA for *sven_2895* terminator
sven2895 term rev	***GATC***AAAACACGCGAAGGGCCGTCCGCACCACGG ATGGCGGACGGCCCTT	Creating dsDNA for *sven_2895* terminator
sven4374 term fwd	***CTAG***GGGAAGGCCCCGCACCGAAAGGTGCGGGGC CTTCTCGCATGTCCGG	Creating dsDNA for *sven_4374* terminator
sven4374 term rev	***GATC***CCGGACATGCGAGAAGGCCCCGCACCTTTC GGTGCGGGGCCTTCCC	Creating dsDNA for *sven_4374* terminator
Oligonucleotides used for the sRNA-based work		
tRSA F EcoRI	CATCATGAATCCAAAAAAAAAAAAAAGCCCGGATA GCTCAGTC	Amplifying the tRSA aptamer for insertion into pMC500
tRSA R XhoI	CTCGAGTTTTTTTTTTTTCTGCAGTGGCGCCCG	Amplifying the tRSA aptamer for insertion into pMC500; *in vitro* transcription of *tRSA*
svr1031 F XhoI	CATCATCTCGAGTGTGCGGCTTTTCGGAGC	Cloning *svr1031* into aptamer vectors
svr1031 R BamHI	CATCTAGGATCCCCACAACCGTCTTCGCATGG	Cloning *svr1031* into aptamer vectors
svr2416 F XhoI	CTACTACTCGAGCTCCTGCTCGTGATCGTCTTC	Cloning *svr2416* into aptamer vectors
svr2416 R BamHI	CATAATGGATCCTCGCCTTCTTCGGTCCCG	Cloning *svr2416* into aptamer vectors
svr3329 F XhoI	CATCTACTCGAGAGGCCCGCGAGACCGAG	Cloning *svr3329* into aptamer vectors
svr3329 R BamHI	CATCATGGATCCGTCGAGCTGGGTGTCCGC	Cloning *svr3329* into aptamer vectors
svr5279 F XhoI	CATTATCTCGAGCTCCGGCCGCGCCGGAC	Cloning *svr5279* into aptamer vectors
svr5279 R BamHI	CATCATGGATCCGTTATGCGCCAGAGATTCGCG	Cloning *svr5279* into aptamer vectors
tRSA probe	ACTATCTTACGCACTTGCATGATTCTGG	Northern probe to detect the tRSA aptamer
Conserved sRNA1	GCCCGGTGAAGGTTGAGAAGACGATCACGA	Northern probe to detect svr2416
Conserved sRNA2	GGGGGAGCCGAGTCGGGCAGTTCGGGA	Northern probe to detect svr5279

T7 + tRSA aptamer	TAATACGACTCACTATAGGG**AAAAAAAAAAAAAG CCCGGATAGCTCAGTC**	*In vitro* transcription of tRSA-tagged sRNAs
svr1031 end R EcoRI	GCAGTGAATTCGCAGAACGAAGCCCCCGAT	*In vitro* transcription of *tRSA-svr1031*
svr2416 end R EcoRI	GCAGTGAATTCTGGAAGAACCCCCGCGATT	*In vitro* transcription of *tRSA-svr2416*
svr3329 end R EcoRI	CGTGAGAATTCAAGGGTTGCGGCGGCAC	*In vitro* transcription of *tRSA-svr3329*
svr5279 end R EcoRI	CGTGAGAATTCAACCGGCCGCGTCGGGC	*In vitro* transcription of *tRSA-svr5279*
pMC600 seq fwd	TAAGGGCGACACGGAAATGT	Sequencing pMC600
pMC600 seq rev	TTTGTGATGCTCGTCAGGGG	Sequencing pMC600
SF14 promoter seq	CCTTGACCTTGATGAGGCGG	Sequencing pMC600
SVEN CHPH BIOTIN 1	GTTTCATCGACGCCTGCT	Biotinylated at 5' end; amplifying the *chpH* promoter
SVEN CHPH PROM-2	CAGCGACGACCTTCTTGA	Amplifying the *chpH* promoter

***** Restriction enzyme sites (*Bam*HI, *Eco*RI, *Nde*I, *Xba*I and *Xho*I), ***5' overhangs***, *ribosome binding sites (RBS)* and **T7 promoters** are indicated.

Table 2.4 *Taq* polymerase (GeneDirex) PCR reaction conditions.

Reaction component	Final concentration
PCR buffer (GeneDirex) (10×)	1×
DMSO (100%)	5%
Forward primer (5 µM)	0.3 µM
Reverse primer (5 µM)	0.3 µM
dNTPs (10 mM)	200 µM (each nucleotide)
Taq polymerase (5 U/µL)	0.025 U/µL
Template DNA	0.2 ng/µL plasmid DNA
Nuclease free water	Make final volume up to 20 µL
Total volume	20 µL

Table 2.5 *Taq* polymerase (GeneDirex) PCR cycling conditions.

PCR Step	Temperature (°C)	Time	Number of Cycles
Denaturation	95	3 min	1
Denaturation	95	30 s	35
Primer annealing	T_m* - 5	30 s	
Extension	72	1 min/kb	
Final Extension	72	5 min	1

* T_m: primer melting temperature

Table 2.6 Phusion® High-Fidelity polymerase (New England Biolabs) PCR reaction conditions.

Reaction component	Final concentration
Phusion GC buffer (5×)	1×
DMSO (100%)	5%
Forward primer (5 µM)	0.5 µM
Reverse primer (5 µM)	0.5 µM
dNTPs (10 mM)	200 µM (each nucleotide)
Phusion polymerase (2 U/µL)	0.02 U/µL
Template DNA	0.2-1 ng/µL genomic DNA
Nuclease free water	Make final volume up to 50 µL
Total volume	50 µL

Table 2.7 Phusion® High-Fidelity polymerase (New England Biolabs) PCR cycling conditions.

PCR Step	Temperature (°C)	Time	Number of Cycles
Denaturation	98	3 min	1
Denaturation	98	10 s	35
Primer annealing*	T_m + 3	30 s	
Extension	72	15-30 s/kb	
Final Extension	72	5 min	1

* if the T_m was greater than or equal to 72°C, the primer annealing step was excluded.

Chapter 3: Intrinsic termination signals in *Streptomyces* bacteria

3.1 Introduction

3.1.1 Transcription termination

Termination is the final stage of transcription, during which RNAP dissociates from the template DNA and releases a nascent RNA transcript (Peters *et al.*, 2011). This process is essential for both terminating and regulating gene expression in bacteria (Santangelo and Artsimovitch, 2011). Transcription terminators are typically located at the 3' end of genes and function to insulate the expression of individual genes; however, they can also be involved in regulation when found elsewhere (*e.g.* 5' end of genes) (Santangelo and Artsimovitch, 2011).

In bacteria, there are two main classes of termination signals: factor-dependent and intrinsic terminators. Factor-dependent terminators require the regulatory protein, Rho, to mediate the release of the new RNA transcript, whereas termination at intrinsic terminators depend only on a nucleic acid signal (*i.e.* RNA structure) (Peters *et al.*, 2011; Santangelo and Artsimovitch, 2011). Canonical intrinsic terminators are characterized by a GC-rich RNA hairpin followed by U-rich tail (Peters *et al.*, 2011). The U-tract promotes termination by inducing a transient transcriptional pause (Gusarov and Nudler, 1999). This pause allows the upstream hairpin to form, resulting in RNAP dissociation and RNA release (Gusarov and Nudler, 1999; Nudler and Gottesman, 2002).

While canonical intrinsic terminators are abundant in *E. coli* (Peters *et al.*, 2009; Santangelo and Artsimovitch, 2011), they are underrepresented in many other bacterial species (Mitra *et al.*, 2009). As a result, alternative terminator structures have been proposed, including non-canonical structures like I-shaped, V-shaped and U-shaped terminators (Section 1.3.1.2) (Mitra *et al.*, 2011). Although there is conflicting evidence that I-shaped terminators may be functional *in vivo* (Ingham *et al.*, 1995; Unniraman *et al.*, 2001; Czyz *et al.*, 2014), the termination ability of the other predicted structures has not yet been tested. Recently, a unique branching structure (with a trailing U-tract) was shown to function as an intrinsic terminator in *Enterococcus faecalis* (Johnson *et al.*, 2014), supporting the idea that there may be greater flexibility in the structure of intrinsic terminators than was initially thought.

3.1.2 Predicting and verifying terminators

Numerous terminator prediction programs have been designed to date [*e.g.* ARNold (Naville *et al.*, 2011), TransTermHP (Kingsford *et al.*, 2007) and WebGeSTer (Mitra *et al.*, 2011)], and these can facilitate genome annotation and the prediction of transcriptional units. These algorithms have focused largely on the detection of intrinsic terminators, as Rho-dependent termination sites have no identifiable consensus sequence or structure (Alifano *et al.*, 1991; Ciampi, 2006). For the prediction of intrinsic terminators, these programs typically search for motifs that can form short, low-energy hairpins (Kingsford *et al.*, 2007). Additional constraints on terminator stem-length, tail U-content and genetic

context (informed by known intrinsic terminators) are often also imposed (Gardner *et al.*, 2011; Mitra *et al.*, 2011). For example, the TransTermHP algorithm will only identify palindrome sequences that are followed by a (6 nt) tail with at least 3 thymine residues (Kingsford *et al.*, 2007). In contrast, the WebGeSTer prediction program will search for both canonical and non-canonical terminator structures (see 3.1.1), but only at the 3' end (between nucleotides -20 and +270 relative to the stop codon) of annotated genes (Mitra *et al.*, 2011). Notably, only hairpins with a (10 nt) tail containing at least 4 thymine residues are recognized as canonical terminators by WebGeSTer (Mitra *et al.*, 2011). Other statistical learning algorithms also exist, like Erpin which was trained on a set of *E. coli* and *Bacillus subtilis* intrinsic terminators (Gautheret and Lambert, 2001). This algorithm has been combined with a pattern recognition program – RNAmotif (Macke *et al.*, 2001) – to generate ARNold (Naville *et al.*, 2011), which has been used to identify terminators in many different species (Ramos *et al.*, 2014; Fritsch *et al.*, 2015). Like TransTermHP, ARNold does not restrict its search to a particular genetic context (Naville *et al.*, 2011).

While computation approaches have been used successfully to identify millions of intrinsic terminators (Mitra *et al.*, 2011), few have been experimentally validated. To be classified as a terminator, these structures must reduce transcription of the downstream region *in vivo* (Peters *et al.*, 2011). This property is often tested using reverse transcription (RT)-PCR (Arrebola *et al.*, 2012; Fritsch *et al.*, 2015) or a reporter assay (Ingham *et al.*, 1995; Chen *et al.*, 2013). Specifically, RT-PCR of regions upstream and downstream of a potential terminator can be used to detect decreases in transcription (Fritsch *et al.*, 2015). When a potential terminator is introduced upstream of a reporter gene [*e.g.* a fluorescent protein (Chen *et al.*, 2013) or an antibiotic resistance gene (Ingham *et al.*, 1995)], decreases in reporter expression are also indicative of effective termination. The other requirement of potential terminators is that they promote RNAP dissociation during *in vitro* transcription (Peters *et al.*, 2011), although this condition is often not tested.

3.1.3 Intrinsic terminators in the streptomycetes

Few studies have focused on understanding transcription termination in *Streptomyces* bacteria. To date, there have only been four intrinsic terminators characterized in these organisms (Pulido and Jimenez, 1987; Ingham *et al.*, 1996), three of which were validated using *in vitro* transcription experiments (Ingham *et al.*, 1995). Notably, these terminators appear to lack the canonical U-tail associated with most *E. coli* and *B. subtilis* terminators (Pulido and Jimenez, 1987; Ingham *et al.*, 1995). These *in vivo* studies are consistent with bioinformatic studies performed here, which predict few canonical terminators in *Streptomyces* species (Table 3.1). Instead, the streptomycetes appear to encode a large number of non-canonical terminators (Mitra *et al.*, 2011). The low representation of canonical structures (as predicted by ARNold and WebGeSTer) is likely a result of the high *Streptomyces* genomic GC content (Mitra *et al.*, 2011); when tail U-content constraints are relaxed (*i.e.* only 3 Us are required in the tail, as in TransTerm) there is a drastic increase in the number of terminators predicted.

Due to the lack of canonical intrinsic terminators, transcription termination remains poorly understood in these bacteria. We are interested in determining whether the streptomycetes can terminate transcription at non-canonical structures and what terminator characteristics (*e.g.* stem length, loop size and tail composition) contribute to termination efficiency. With these analyses we also hope to expand the collection of characterized *Streptomyces* terminators which can be used to engineer new synthetic constructs and enhance our ability to predict transcriptional units from genomic data.

3.2 Results

3.2.1 *In silico* identification of potential terminators

Given that the number of terminators predicted by the current algorithms varied so dramatically (Kingsford *et al.*, 2007; Mitra *et al.*, 2011; Naville *et al.*, 2011), we chose to develop a novel *in silico* approach that would allow us to identify biologically relevant termination sites. We created an algorithm to search RNA sequencing data for drastic (>50%), well-defined decreases in transcript levels. For each 10 nucleotide window in the genome, this algorithm calculated the average read depth and compared it to the average of the 10 nucleotides upstream (Figure 2.4). Potential termination sites were then identified at genome positions where transcript levels decreased by at least 50% within the 20 nt window. To avoid falsely identifying terminators in lowly expressed regions, the algorithm required the average read depth within any 10 nt window to exceed a certain species-specific threshold (based on the average coverage per nt of all coding regions). Upper limits (based on the maximum read depth of a coding region) were also implemented to eliminate any terminators within the highly expressed rRNA regions, where terminator structures have been predicted previously in *Streptomyces* (Pernodet *et al.*, 1989). Importantly, the algorithm was designed such that both boundaries (upper and lower limits) could be easily adjusted depending on RNA sequencing coverage.

We have applied this algorithm to RNA sequencing data from three divergent *Streptomyces* species (*S. avermilitis, S. coelicolor* and *S. venezuelae*)(Moody *et al.*, 2013) in order to identify candidate termination sequences in each genome. The sequencing data from this study comprises two distinct RNA libraries: (1) a long transcript library created from RNA samples depleted of rRNA and sequenced using a protocol optimized for long transcripts and (2) a short transcript library enriched with small (40-300 nt) transcripts (Moody *et al.*, 2013). For these analyses, we chose to screen RNA sequencing libraries enriched for long transcripts (Moody *et al.*, 2013) using species-specific boundaries set based on the sequencing coverage of coding regions. To avoid small RNA degradation products, all sequencing reads from the short transcript library (Moody *et al.*, 2013) were initially ignored. Using these parameters, this algorithm was used successfully to identify hundreds of termination sites in all three *Streptomyces* species (Table 3.2).

Although terminators were evenly distributed between the positive and negative strands (Table 3.2), there was a clear bias in their genomic context (Figure 3.1). More than

half (63-71%) of the terminators identified in each species were positioned within a coding region. The rest of the terminator sites were located in intergenic regions, either antisense to an annotated coding region or between coding regions (Figure 3.1). Interestingly, a large number (62-86%) of antisense terminators were found opposite the highly expressed rRNA operons in each species.

3.2.2 *In silico* characterization of identified terminators

Since sense-strand intergenic terminators have a more obvious biological function (termination of their upstream sense-genes) than antisense or intragenic terminators, we chose to focus our analyses on *S. venezuelae* where this type of terminator was most prevalent (Figure 3.1). Each intergenic terminator found between coding sequences was first visualized using IGV (Thorvaldsdóttir *et al.*, 2013) to ensure that the algorithm had only detected well-defined termination sites. While most terminators exhibited drastic decreases in transcript levels (*e.g.* Figure 3.2A), we detected several (21) that were associated with significant transcriptional read-through (*e.g.* Figure 3.2B). These termination sites, along with those (20) associated with rRNA operons, were excluded from the downstream analyses.

To determine which termination sites likely resulted from intrinsic RNA structures, we compared the remaining well-defined terminators to those identified by the WebGeSTer algorithm (Mitra *et al.*, 2011). Due to the constraints imposed by WebGeSTer, these comparisons were limited to RNA sequencing terminators identified within 270 nucleotides of an annotated stop codon. As a result, 34 terminators (including 31 sRNA-associated terminators) were excluded from the analysis. Of the remaining 131 eligible terminators, 60% co-localized with a terminator predicted by WebGeSTer (Figure 3.3). The majority (80%) of these terminators aligned with non-canonical I-shaped hairpins, which left the number of canonical terminators unsurprisingly low (20%) (Figure 3.3). However, these L-shaped terminators appear to be overrepresented, as only 7% of all WebGeSTer predicted terminators in *S. venezuelae* have a canonical structure (Table 3.1). The aligned RNA sequencing terminators showed a preference for longer hairpin stems, when compared to the entire set of WebGeSTer terminators (19.3 nt vs. 15.7 nt, Table 3.3). In contrast, there were only slight differences in loop size and hairpin GC content between the two collections of terminators (Table 3.3). Together, these results suggest that *Streptomyces* may terminate transcription efficiently at non-canonical terminators having longer hairpin stems.

Interestingly, over two-thirds (70%) of the aligned terminators exhibited an unexpected expression profile (represented in Figure 3.4A-B). While sequencing reads from the short-transcript library typically aligned nicely with predicted WebGeSTer terminators, very few reads from the long-transcript library were ever detected in these regions. These expression profiles were consistent with the terminator hairpin forming a stable RNA degradation product distinct from the upstream transcript.

3.2.3 *In vivo* validation of potential intrinsic terminators

3.2.3.1 Building a fluorescent reporter construct

To confirm that the intrinsic terminators identified in the *in silico* analyses (Section 3.2.1 and 3.2.2) were able to terminate transcription *in vivo*, we created a specialized reporter construct similar to ones reported earlier (Huang, 2007; Chen *et al.*, 2013) (Figure 2.1). This system consisted of two *Streptomyces* codon-optimized fluorescent reporter genes (mCherry and msfGFP), expressed from a single, thiostrepton-inducible promoter (P_{tipA}) (Huang *et al.*, 2005). Predicted terminators were introduced between mCherry and msfGFP, allowing termination efficiency to be quantified simply by comparing the expression of the two reporters (Chen *et al.*, 2013). For example, highly efficient terminators were expected to reduce expression of the downstream reporter (msfGFP), relative to the upstream reporter (mCherry) (Huang, 2007). Importantly, these analyses would allow us to distinguish between *bona fide* transcription termination sites and sites arising from RNA degradation, as the latter should have no impact on the expression of the downstream reporter.

3.2.3.2 Prioritizing terminators for testing *in vivo*

Of the 79 potential intrinsic terminators (Figure 3.3, aligned terminators), we initially prioritized a small subset for insertion into the fluorescent reporter construct. Specifically, we selected three structurally diverse terminators: 0060, 2895 and 4374. The numerical designations were based on their associated gene numbers in *S. venezuelae* - the terminators were found downstream of *sven_0060*, *sven_2895* and *sven_4374*, respectively (Table 3.4), and for each, we observed strong terminator sites downstream of homologous genes in *S. avermitilis* and *S. coelicolor*. Further analysis of genomic context revealed that terminator 4374 (and that of its homologous genes) was positioned between convergent genes, while the two other terminators are situated between genes orientated in the same direction (Table 3.4).

To determine which terminator characteristics were important for termination, we sought to compare the activity of terminators exhibiting diverse structural features (Table 3.4), focussing on stem length and tail-U content. Since stem length is expected to contribute to termination strength (see 3.2.2), we included one terminator with a relatively long (18 nt) hairpin stem (0060), compared with the shorter 16 nt hairpin stems for 2895 and 4374. A terminator with low tail U-content (4374 - 20% U's) was selected for comparison with terminators having higher (50%) U-content (0600 and 2895) to assess the functionality of non-canonical structures in *Streptomyces* (Table 3.4).

In addition to structural differences, these terminators were also associated with different RNA sequencing expression profiles (Figure 3.4). Notably, the terminator-associated peak in the short read library was only present in conjunction with terminator 0060 and terminator 2895 (compare Figure 3.4A-B with C). It is unclear whether the accumulation of a short stable transcript influences termination, or whether short transcript accumulation is associated with U-tail composition (Figure 3.4C).

3.2.3.3 Measuring terminator strength *in vivo*

To quantify termination efficiency, the selected terminators were cloned into the fluorescent reporter construct described in 3.2.3.1. Each terminator reporter construct was then introduced into *S. venezuelae* expressing TipA, since this protein was required for transcription from the thiostrepton-inducible promoter P_{tipA}. The terminator-less reporter construct (positive control) and the empty cloning vector (negative control) were also introduced into the same *S. venezuelae* strain.

Before beginning the terminator analyses, we wanted to demonstrate *in vivo* expression of the terminator-less fluorescent reporter construct. Initially, we measured the fluorescence of whole *S. venezuelae* cells carrying this construct; however, we were unable to reproducibly detect increases in mCherry and msfGFP expression following thiostrepton-induction. Post-induction fluorescence of this strain was also not consistently higher than the empty vector control. To assess whether this was a consequence of insufficient P_{tipA} activation by TipA, both control constructs were transferred into wild-type *S. coelicolor,* where the thiostrepton-inducible promoter has been used successfully in the past [*e.g.* (Huang *et al.*, 2005)] as a result of the chromosomally-encoded *tipA* gene (Takano *et al.*, 1995). However, even in *S. coelicolor*, we were unable to reliably detect higher fluorescence levels in the induced positive control compared with the un-induced strain or the negative control. These results suggested that P_{tipA} inactivation was likely not responsible for the lack of reporter expression, although these comparisons did not allow us to determine whether other factors were influencing reporter transcription or translation.

Since *Streptomyces* cells exhibit strong autofluorescence (Flärdh and Buttner, 2009), the weak mCherry and msfGFP signals may have been masked in the whole cell assays. To address this possibility, cell lysates were used to measure fluorescence of the terminator-less construct instead. Alternate excitation and emission parameters were also employed (see 2.5.2) to maximize the likelihood of detecting each fluorescent protein. While preliminary experiments in *S. venezuelae* demonstrated a clear increase in msfGFP expression following induction, mCherry levels did not exceed those of the empty vector control (Figure 3.5). Notably, the ~2-fold increase in msfGFP expression does not compare to the 200-fold increase typically observed for P_{tipA}-driven genes (Murakami *et al.*, 1989). This, along with the fact that these trends were never reproduced, suggests that the reporter, if expressed, is not being detected accurately. Despite testing a range of incubation times, induction times and lysis techniques, we were also never able to consistently detect either fluorescent reporter in *S. coelicolor* lysates.

Due to the discrepancies observed with the terminator-less reporter construct, mCherry and msfGFP fluorescence has not yet been quantified for the remaining terminator constructs.

3.3 Discussion

In this work, we have developed an innovative terminator prediction program that leverages transcriptional data to identify intrinsic terminators, and used it to evaluate transcription termination in three divergent *Streptomyces* species.

3.3.1 Predicting transcription terminators using RNA sequencing data

Using our *in silico* prediction algorithm, we successfully identified hundreds of terminators in the streptomycetes. While typical computation methods consider RNA sequence and structure (Kingsford *et al.*, 2007; Mitra *et al.*, 2011; Naville *et al.*, 2011), our analyses were primarily based on transcriptomic data. Using RNA sequencing data to guide terminator identification would be expected to uncover more biologically relevant terminators than strictly computational methods. Due to the lack of well-characterized *Streptomyces* intrinsic terminators, we were unable to compile a training set for our algorithm. Instead, we chose parameters that would identify well-defined termination sites, although the program was designed such that these could be easily modified to search for terminators with different expression levels or patterns.

Applying this novel approach to RNA sequencing data from three *Streptomyces* species resulted in the identification of many well-defined termination sites. In contrast to other computational approaches (*e.g.* TransTermHP and WebGesTer), our algorithm detected relatively few terminators (compare Tables 3.1 and 3.2). This is not entirely surprising as some genes are not well expressed and many do not appear to have a well-defined termination site. Instead, transcript levels typically decrease gradually over hundreds of nucleotides, as determined by manual inspection of RNA sequencing data [using IGV (Thorvaldsdóttir *et al.*, 2013)]. While transcript levels decreased dramatically at most identified termination sites, gradual decreases in transcript levels or transcriptional read-through (Figure 3.2) was observed at ~10% of sites. This high rate of false positives meant that all terminators needed to be manually verified following identification, a property that is not ideal for a high-throughput prediction tool. In the future, more stringent parameters could be integrated into the algorithm to ensure that read depth following a potential terminator is relatively low, thus reducing the need for manual confirmation.

Despite these shortcomings, this algorithm appears to accurately detect intrinsic transcription terminators. In fact, well over half of the termination sites identified downstream of an annotated coding sequence aligned nicely with an intrinsic terminator predicted by the WebGeSTer algorithm. Interestingly, most aligned terminators displayed a characteristic peak in the short transcript library which corresponded to the terminator hairpin. This unique expression profile was consistent with the highly stable terminator hairpin being cleaved from the full-length transcript following termination. This raises the intriguing possibility that these stable terminator fragments may act as small, non-coding RNAs within the cell (Kim *et al.*, 2014). Other sRNAs produced from mRNA 3' UTRs are common in *E. coli* (Chao *et al.*, 2012) and are not unprecedented in *Streptomyces* (Kim *et al.*, 2014).

Of the other terminators identified in these analyses, the majority were found within coding sequences. These intragenic termination sites may have arisen through premature transcription termination or may be an artifact stemming from differential degradation within any given mRNA (Czyz *et al.*, 2014). Transcription attenuation is commonly used to regulate gene expression in bacteria (Henkin and Yanofsky, 2002), including *Streptomyces* where 0.6% of genes are predicted to be impacted by attenuation (Naville and Gautheret, 2009). While premature termination typically occurs within the 5' UTR of genes (Naville and Gautheret, 2009), we cannot exclude the possibility that transcription attenuates within the coding regions of *Streptomyces* genes. It is important to note that incorrect genome annotations could also lead to terminators being inaccurately classified as intragenic.

Since antisense transcription appears to be pervasive in many bacteria [*e.g. Streptomyces* (Moody *et al.*, 2013), *E. coli* and *Salmonella* (Raghavan *et al.*, 2012)], it is not surprising that antisense-associated terminators were also found on the opposite strand of annotated genes. The vast majority of these terminators were associated with antisense RNAs opposite rRNA operons. The role of these antisense RNAs is unknown; however, it is tempting to speculate that they are involved in rRNA processing.

3.3.2 Canonical versus non-canonical terminators in *Streptomyces*

Consistent with predictions made by other computational approaches (*e.g.* WebGeSTer), the majority of intergenic terminators identified here using experimental data as a basis, aligned with predicted non-canonical structures. This type of terminator is thought to be sufficient for termination when RNA elongation is slow, since the hairpins will have time to form even without a U-tract induced pause (Mitra *et al.*, 2009). Evidence for this hypothesis comes from the fact that organisms with slower elongation rates [*e.g.* the *Streptomyces* relative, *Mycobacterium* (Harshey and Ramakrishnan, 1977)] have more non-canonical structures (Unniraman *et al.*, 2001) than *E. coli*, where elongation occurs at a faster rate (Bremer and Yuan, 1968; Shen and Bremer, 1977). Notably, the U-tail is no longer required when *E. coli* RNAPs are static or elongating slowly (Yarnell and Roberts, 1999). Although the RNA elongation rate is unknown in the streptomycetes, all of the previously characterized *Streptomyces* terminators have non-canonical structures (Pulido and Jimenez, 1987; Ingham *et al.*, 1995).

The role of non-canonical terminators in *Mycobacterium* has, however, recently been called into question. Czyz and colleagues (Czyz *et al.*, 2014) reported that non-canonical structures are unable to terminate transcription *in vitro*, and instead concluded that termination at non-canonical terminators is either an artifact of RNA degradation or is due to Rho-dependent termination (Czyz *et al.*, 2014). It will be important to combine our reporter assays with *in vitro* transcription experiments to ensure that termination can occur at non-canonical terminators in *Streptomyces*.

Interestingly, we also identified a few (16) termination sites in the RNA sequencing data which corresponded to canonical intrinsic structures. It is well established that canonical structures can terminate transcription in these bacteria, as many canonical phage

terminators [*e.g.* phage λ's t_o terminator (Scholtissek and Grosse, 1987) and phage fd's *ter fed* terminator (Gentz *et al.*, 1981)] have been used successfully in *Streptomyces* expression vectors (Huang *et al.*, 2005). However, since these canonical terminators are in such low abundance within the genome, their genetic context is likely to be of biological relevance (Chapter 5).

In *E. coli*, terminator strength depends on terminator loop size and stem GC content (as well as tail U-content) (Chen *et al.*, 2013), whereas in *Streptomyces* it only appears to depend on stem length. Specifically, the well-defined terminators identified in our *in silico* analyses showed a strong preference for longer hairpin stems; however, the mechanism by which longer hairpin stems promote efficient termination still remains to be determined. Notably, we were unable to observe a difference in loop size or bulk stem GC content between the strong identified terminators and all of the WebGeSTer predicted terminators. Since strong *E. coli* terminators only exhibit increased GC content at the base of the hairpin stem (Chen *et al.*, 2013), it may be worth investigating positional differences in GC content in the future.

3.3.3 Experimental verification of predicted terminators and future directions

Since our algorithm was unable to distinguish between termination sites and sites resulting from RNA degradation, it is important to combine these *in silico* analyses with experimental approaches. We initially attempted to create a fluorescent reporter construct that would allow us to quantify termination efficiency for any terminator of interest using msfGFP, and account for any transcriptional differences between strains/experiments using mCherry. While a similar reporter construct has been used to study terminator strength in *E. coli* (Chen *et al.*, 2013), it was not successfully implemented in either *S. coelicolor* or *S. venezuelae*. Specifically, we could never reproducibly detect mCherry and msfGFP expression in our positive controls (the terminator-less reporter strains), at levels above background. In the whole cell assays, these fluorescent reporters were likely masked by the strong autofluorescence exhibited by *Streptomyces* (Flärdh and Buttner, 2009); however, it is still unclear why mCherry and msfGFP fluorescence were not reliably detected in crude cell lysates, especially since GFP fluorescence has been previously quantified in *Streptomyces* lysates (Siegl *et al.*, 2013). Absence of measurable fluorescence is likely to result from one or more of: insufficient reporter expression, inefficient folding of the fluorescent proteins, or inefficient reporter detection. To date, we have been unable to distinguish between these possibilities, although we do not expect the lack of fluorescence to be due to unfolded fluorescent proteins, as strains were induced for far longer (2-4 hours) than is required for mCherry [~40 minutes (Merzlyak *et al.*, 2007)] or msfGFP [~4 min (Pédelacq *et al.*, 2006; Khmelinskii *et al.*, 2012)] maturation.

After little success using the fluorescent reporter construct, we have decided to transition to a more classical reporter system (the GUS reporter system, Figure 3.6), which has been optimized for use in the streptomycetes (Myronovskyi *et al.*, 2011). With this system, terminators are introduced upstream of the reporter gene *gusA* (β-glucuronidase),

whose expression can be easily quantified using a spectrophotometric assay (Myronovskyi *et al.*, 2011). Importantly, we will be replacing the thiostrepton-inducible promoter with the constitutive SF14 promoter (Labes *et al.*, 1997), to ensure expression of the reporter system. Initial characterization of the terminators described above (Section 3.2.3.2) will be important for expanding the collection of terminators available for use *in vivo* (for synthetic purposes) and *in silico* (for algorithm training) (Chapter 5). Further mutational analyses of these structures will then be used to examine the effect of tail composition, stem length, and loop size on terminator efficiency (Chapter 5). These analyses will be critical for understanding how transcription terminates in *Streptomyces* and other related bacteria (*e.g. Mycobacterium*), and will ultimately allow us to better understand transcriptional dynamics, and predict – and manipulate – transcriptional units in these bacteria.

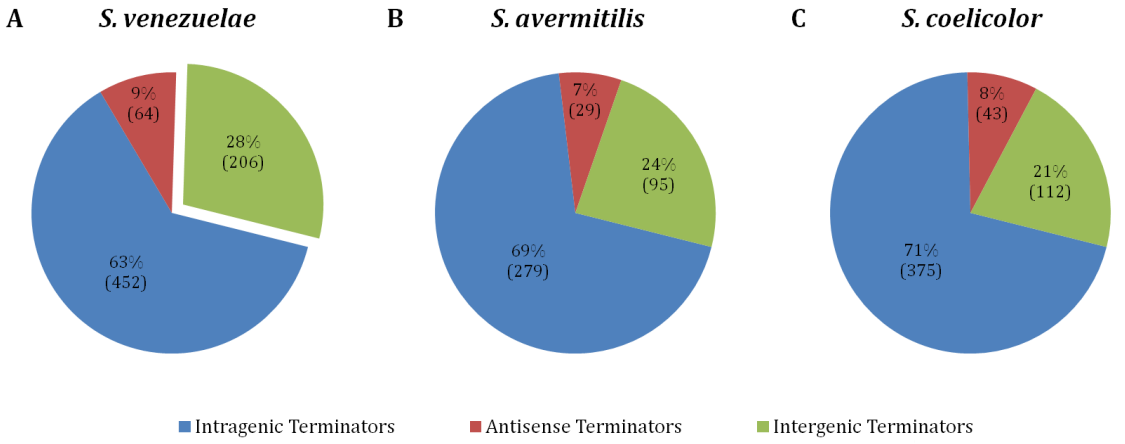

Figure 3.1 Location of termination sites identified from long read RNA sequencing data. Terminators identified in (A) *S. venezuelae,* (B) *S. avermitilis* and (C) *S. coelicolor* were classified as either intragenic (within a coding region), antisense (on the opposite strand of an annotated gene) or intergenic (between coding regions). Similar distributions were observed for all three species.

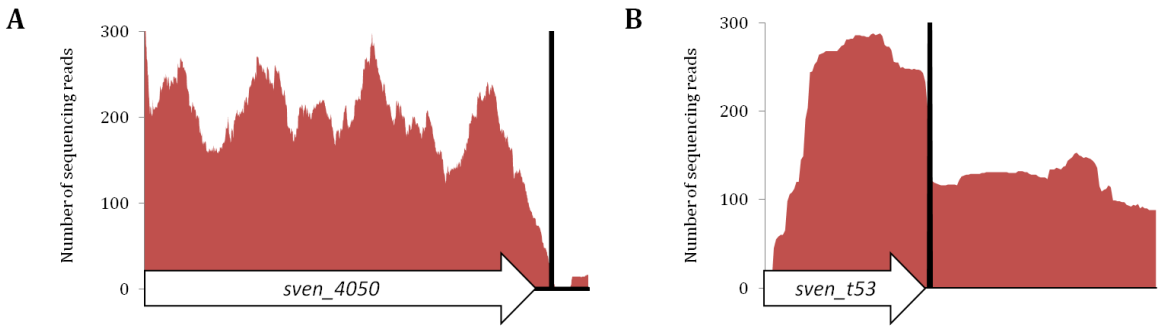

Figure 3.2 Read coverage of termination sites identified in *S. venezuelae*. Expression profiles of (A) *sven_4050* and (B) *sven_t53*, with corresponding transcription termination sites indicated in black. Note that transcript levels on the positive strand (red) decrease rapidly following the terminator identified in (A), while in (B) they remain relatively high, even with a drop of >50% in transcript levels (as dictated by the algorithm parameters used for the *in silico* analyses).

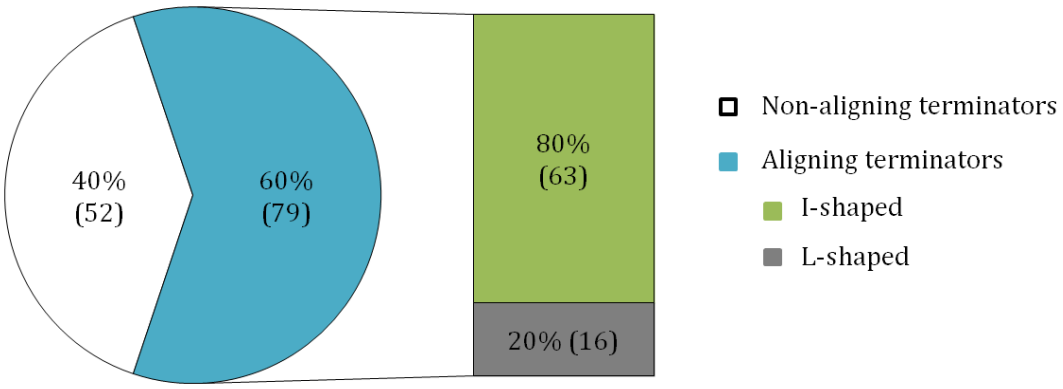

Figure 3.3 Distribution of aligned terminators. Well-defined intergenic terminators in *S. venezuelae* were compared to terminators predicted by the WebGeSTer algorithm (Mitra *et al.*, 2011). The RNA sequencing terminators with a WebGeSTer hairpin predicted within 20 nucleotides were considered aligning terminators. Aligned terminators were further classified as either I-shaped (non-canonical) or L-shaped (canonical), based on the structure of the corresponding WebGeSTer terminator.

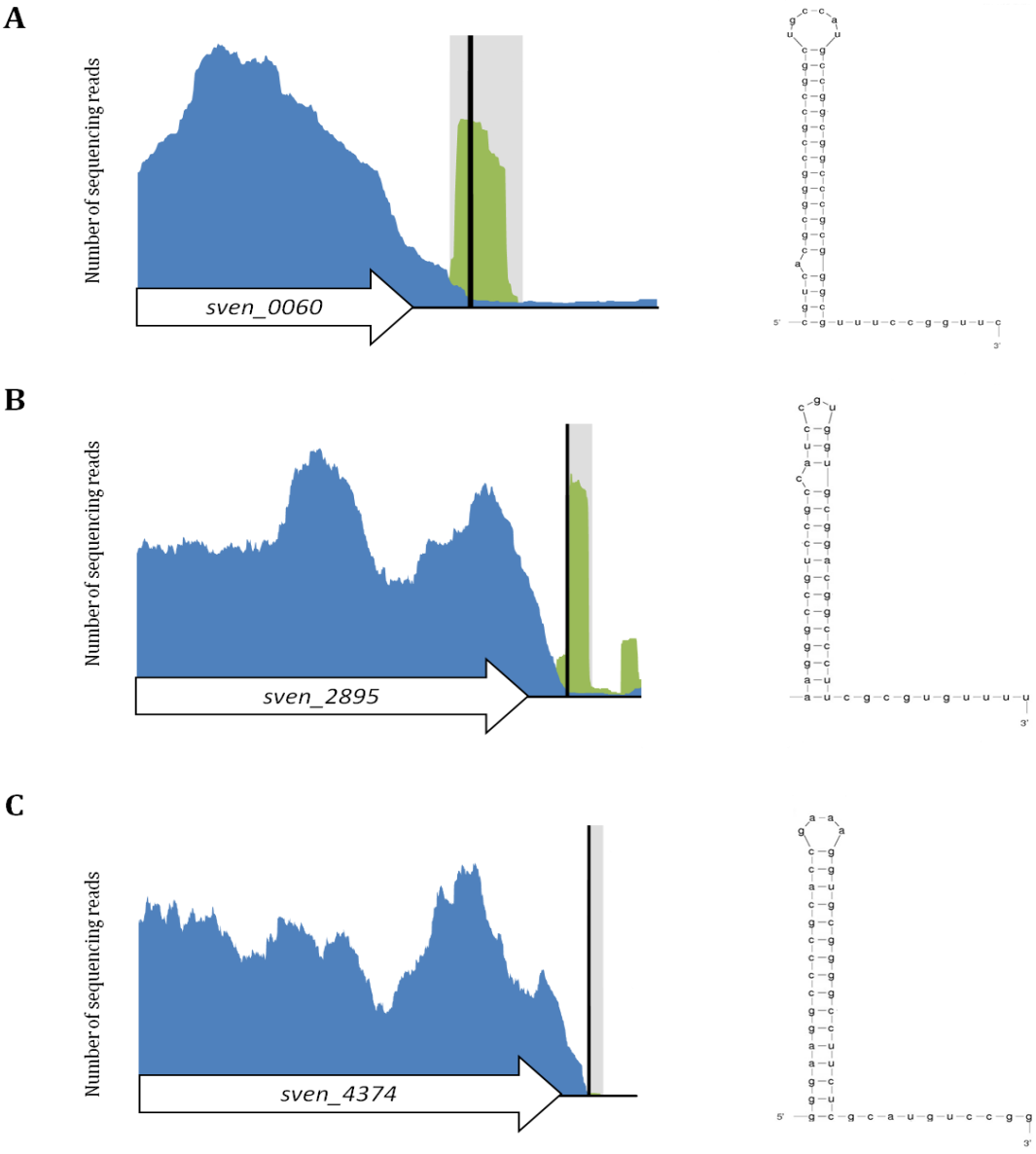

Figure 3.4 Comparing expression and structure of prioritized terminators. Expression profile for each selected terminator: (A) 0060, (B) 2895 and (C) 4374. Read coverage is shown for both the long transcript (blue) and short transcript (green) libraries. The location of each identified termination site is denoted by a black line. Terminators predicted by WebGeSTer (Mitra *et al.*, 2009) are shown in grey, with the corresponding RNA structures [illustrated using Mfold (Zuker, 2003)] on the right.

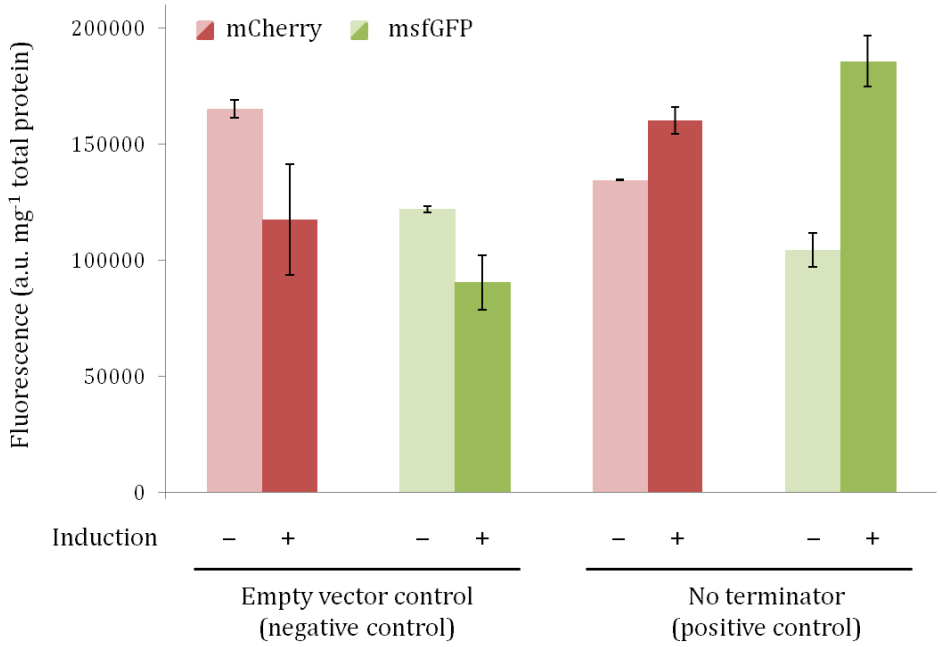

Figure 3.5 *In vivo* expression of fluorescent terminator reporter construct. mCherry and msfGFP fluorescence were measured in cell lysates harvested from *S. venezuelae* strains expressing mCherry and msfGFP from a thiostrepton inducible promoter (No terminator; pMS81-*tipA*/pMC251). As a control, fluorescence was also measured from lysates prepared from *S. venezuelae* containing the empty cloning vector (pMS81-*tipA*/pIJ6902). All strains were grown in 50 mL of MYM liquid medium for 8 hours before thiostrepton (or an equal volume of DMSO) was added. Two hours post-induction, cells were harvested and lysed. All fluorescence measurements were normalized to lysate protein content. Values are shown as the means of two biological replicates with standard errors.

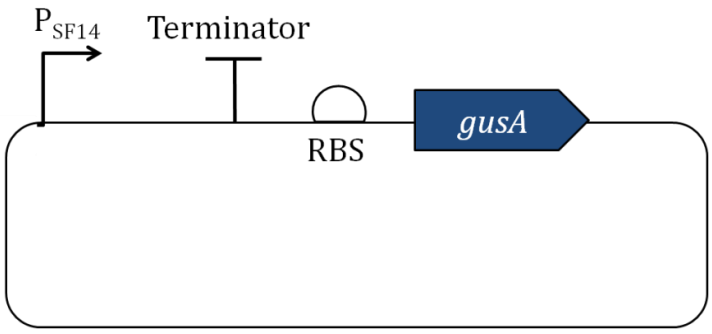

Figure 3.6. Schematic of new terminator reporter construct. With the new GUS reporter system, terminators were inserted upstream of *gusA* (which encodes β-glucuronidase) and its associated RBS. The entire construct will be expressed from the constitutive SF14 promoter.

Table 3.1 Terminators predicted in three divergent *Streptomyces* species

		Number of terminators predicted		
	Number of Genes	ARNold[1]	TransTermHP[1]	WebGeSTer[1,2]
S. avermitilis	7573	640	3431	5124 (440)
S. coelicolor	7846	532	2655	4983 (320)
S. venezuelae	7455	–	–	4927 (338)

[1] Terminators were predicted using three algorithms: ARNold (Naville *et al.*, 2011), TransTermHP (Kingsford *et al.*, 2007) and WebGeSTer (Mitra *et al.*, 2011).
[2] The total number of predicted terminators is shown first, followed by the subset of terminators that exhibit a canonical structure in brackets.

Table 3.2 Identification of transcription termination sites from RNA sequencing data

	Number of terminators identified		
	Positive strand	Negative strand	Total
S. avermitilis	199	204	403
S. coelicolor	280	250	530
S. venezuelae	375	347	722

Table 3.3 Bulk characteristics of WebGeSTer predicted terminators

	All WebGeSTer Terminators[a]	Aligned WebGeSTer Terminators[b]
Stem Length (nt)	15.7	19.3
Loop size (nt)	5.2	4.6
Hairpin GC content (%)	82.2	79.4
Tail U-content (%)	15.2	23.2

[a] All terminators predicted by WebGeSTer (Mitra *et al.*, 2011) for *S. venezuelae* (n = 4927).
[b] Only WebGeSTer terminators that co-localized with terminators identified in the RNA sequencing data (n= 79).

Table 3.4 Characteristics of terminators prioritized for *in vivo* characterization

	Terminator Number		
	0060	2895	4374
Gene upstream of terminator (and associated gene product)	*sven_0060* (cold shock protein)	*sven_2895* (NADH dehydrogenase)	*sven_4374* (Translation elongation factor Tu)
Stem Length (nt)	18	16	16
Loop size (nt)	6	3	4
Hairpin GC content (%)	94.6	72.7	78.1
Tail U-content (%)	50	50	20
Orientation of downstream gene	same	same	convergent

Chapter 4: Identifying RNA-binding proteins in the streptomycetes

4.1 Introduction

4.1.1 RNA-binding proteins and methods for their identification

Bacterial gene expression is subject to multiple levels of regulation. At the post-translational level, gene expression is rapidly controlled by two distinct regulators: small RNAs and RNA-binding proteins (RBPs) (Assche *et al.*, 2015). sRNAs exert their regulatory effects by interacting with either a protein or mRNA target (Waters and Storz, 2009). Protein-binding sRNAs modulate protein activity, most commonly by sequestering proteins away from their normal targets (Storz *et al.*, 2011). In contrast, sRNAs that interact with mRNA targets typically affect transcript stability or translatability. Due to the limited complementarity between sRNAs and mRNAs encoded *in trans*, these interactions often require the RNA chaperone, Hfq (Waters and Storz, 2009).

RBPs can also serve as important post-transcriptional regulators and are involved in a variety of cellular processes. Much like sRNAs, RBPs can modulate the stability and translation of target RNAs (Assche *et al.*, 2015), either directly or by recruiting other interaction partners (*e.g.* sRNAs or proteins). One of the best characterized post-transcriptional regulatory proteins is Hfq. This global RNA chaperone facilitates sRNA-mediated regulation by promoting the interaction of sRNAs and mRNA targets (Vogel and Luisi, 2011). As a result, Hfq mutants generally exhibit increased sensitivity to environmental stresses and decreased virulence (Tsui *et al.*, 1994; Robertson and Roop, 1999; Christiansen *et al.*, 2004; Liu *et al.*, 2010; Wang *et al.*, 2014). A similar phenotype has also been observed in strains where the endoribonuclease, YbeY, has been mutated (Pandey *et al.*, 2011; Vercruysse *et al.*, 2014). In fact, recent evidence has suggested that YbeY may mediate sRNA-based regulation in a similar manner to Hfq (Pandey *et al.*, 2014; Vercruysse *et al.*, 2014). While both Hfq and YbeY interact with diverse RNA transcripts, there are examples of other RNA chaperones that have smaller interaction networks [*e.g.* FinO (Arthur *et al.*, 2003) and FbpB (Smaldone *et al.*, 2012)].

The vast majority of RBPs identified to date have been discovered serendipitously. For example, Hfq was initially identified as a bacteriophage host factor during the characterization of the phage Qβ replicase (Franze de Fernandez *et al.*, 1968; Kondo *et al.*, 1970), whereas CsrA was identified as a regulator of glycogen biosynthesis by transposon mutagenesis (Romeo *et al.*, 1993). More recently, however, direct approaches for identifying transcript-specific RBPs have been developed. The first targeted searches used *in vitro* synthesized and chemically labelled (*e.g.* biotinylated) RNAs for affinity purification of bound RBPs (Iioka *et al.*, 2011). This approach led to the successful identification of several eukaryotic RBPs (Rouault *et al.*, 1989; Ross *et al.*, 1997; Mehta and Driscoll, 1998). To avoid complications associated with chemically labeled RNAs (*i.e.* compromised secondary structure or low target affinity), these labels have now largely been replaced with RNA aptamers (Iioka *et al.*, 2011). These aptamers are short RNA sequences that bind to target molecules with high specificity and affinity (Walker *et al.*, 2008). Aptamer tags specific for

streptavidin (Srisawat and Engelke, 2001; Iioka *et al.*, 2011; Leppek and Stoecklin, 2013), streptomycin (Dangerfield *et al.*, 2006; Windbichler and Schroeder, 2006), tobramycin (Hartmuth *et al.*, 2004) and the MS2 phage coat protein (Said *et al.*, 2009; Desnoyers and Massé, 2012) have all been successfully introduced into target RNAs and used to isolate specific RBPs. This method of RNA tagging is advantageous because it can be used to purify RNA-protein complexes that are assembled either *in vitro* or *in vivo* (Said *et al.*, 2009; Iioka *et al.*, 2011).

Since RNA aptamers can also impact transcript stability and secondary structure, complementary oligonucleotides have been used as an alternative. With this approach, biotinylated oligonucleotides that are complementary to the target RNA are first immobilized on an affinity matrix (Hartmann *et al.*, 2005). Following hybridization of the native target RNA to the bait oligonucleotide, RBPs interacting with the target can then be isolated under either denaturing or non-denaturing conditions (Walker *et al.*, 2008). This method has been most effective for purifying RBPs that associate with RNAs having unstructured/accessible regions (Blencowe *et al.*, 1989; Lingner and Cech, 1996); however, this technique is much less effective for highly structured target transcripts (Walker *et al.*, 2008).

4.1.2 RNA-binding proteins in *Streptomyces* bacteria

Streptomyces bacteria encode homologs of many well-characterized RBPs, including PNPase (Jones and Bibb, 1996), RNase III (Taverniti *et al.*, 2011; Jones *et al.*, 2014), RNase E (Lee and Cohen, 2003) and RNase J (Jones *et al.*, 2014). In other bacteria, these proteins play an important role in sRNA-based regulation (Viegas *et al.*, 2007; Durand *et al.*, 2015; Liu *et al.*, 2015); however, the cellular targets of these RBPs or their role in post-transcriptional regulation remains largely unknown in the streptomycetes. Recent work has demonstrated that RNase III and RNase J, in particular, are required for normal development and antibiotic production, although their direct targets remain elusive (Sello and Buttner, 2008; Bralley *et al.*, 2014; Jones *et al.*, 2014).

Surprisingly, the streptomycetes do not encode a homolog of the pervasive RNA chaperone, Hfq (Sun *et al.*, 2002; Swiercz *et al.*, 2008). Hfq has been reported to be required for the activity of *trans*-encoded sRNAs in bacteria with high genomic GC-content (Jousselin *et al.*, 2009); however, the extremely GC-rich (>70%) *Streptomyces* bacteria are an exception, as they encode hundreds of *trans*-acting sRNAs but no Hfq homolog (Swiercz *et al.*, 2008; Moody *et al.*, 2013). Interestingly, an YbeY homolog has been identified in *Streptomyces* (Davies and Walker, 2008), although its RNA-binding potential has not yet been determined (Assche *et al.*, 2015). Without an obvious RNA chaperone, it is unclear how sRNA-mRNA interactions are mediated in the streptomycetes. We were interested in determining whether *Streptomyces* encode a novel RNA chaperone or whether their sRNAs can function independently of a chaperone. This type of analysis will be important for expanding our understanding of sRNA-mediated regulation in streptomycetes and actinobacteria as a whole.

4.2 Results

4.2.1 Selecting sRNAs for affinity purification

To identify proteins involved in sRNA-mediated regulation (*e.g.* RNA chaperones), we chose to employ an affinity purification strategy using *Streptomyces* sRNAs of interest. Specifically, we selected three *S. venezuelae* sRNAs (svr1031, svr2416 and svr5279) that were also conserved in *S. coelicolor* and *S. avermitilis* (Moody *et al.*, 2013). These sRNAs are highly structured transcripts that range in size from 73 to 90 nucleotides (Moody *et al.*, 2013). Previous northern blotting experiments revealed these sRNAs to be highly expressed during at least two distinct developmental stages (Moody *et al.*, 2013). While little is known about the biological function of svr1031 and svr5279, svr2416 negatively regulates the expression of a nickel-containing superoxide dismutase (encoded by *sodN*) (Kim *et al.*, 2014). svr2416 (also termed s-sodF) is processed from the 3' untranslated region (UTR) of the *sodF* transcript before it binds to and inhibits the translation of the *sodN* mRNA (Kim *et al.*, 2014). There is currently no indication that a protein factor is required for this interaction (Kim *et al.*, 2014); however, many 3'UTR-derived sRNAs bind Hfq in *Salmonella* (Chao *et al.*, 2012).

Another sRNA, svr3329, was included as a positive control for RBP isolation. svr3329 and its *S. coelicolor* homolog, scr3559, are predicted to be *Streptomyces* 6S RNAs (Pánek *et al.*, 2008). Classically in *E. coli*, 6S RNA interacts with and sequesters the housekeeping form of RNAP (*i.e.* σ^{70}-containing RNAP). This interaction inhibits σ^{70}-dependent transcription, allowing for increased transcription from promoters dependent on alternative sigma factors (*e.g.* σ^{s}) (Cavanagh and Wassarman, 2014). The *Streptomyces* 6S RNA appears to directly bind RNAP complexed with its principal sigma factor, HrdB (Mikulík *et al.*, 2014).

4.2.2 Tagging sRNAs with the tRSA aptamer

Since all the *Streptomyces* sRNAs of interest are highly structured, we chose to carry out the affinity purification using sRNAs tagged with a RNA aptamer. We initially selected a streptavidin aptamer that had been used previously to identify RBPs in *Salmonella* and *Staphylococcus aureus* (Niemann *et al.*, 2013; Zhang *et al.*, 2015). This aptamer was originally identified by *in vitro* selection (Srisawat and Engelke, 2001), and was later modified to include a tRNA scaffold (Figure 4.1A) (Iioka *et al.*, 2011), which serves to stabilize aptamer RNA (and sRNA) secondary structure, and increase streptavidin binding efficiency (Iioka *et al.*, 2011); it has been termed the tRNA scaffolded streptavidin aptamer or tRSA. To avoid interfering with the native transcription terminators, we attached the tRSA aptamer to the 5' end of each *S. venezuelae* sRNA of interest.

4.2.2.1 *In vivo* expression of tRSA-tagged sRNAs

As sRNA-protein complexes assembled *in vivo* are expected to be more biologically relevant (Said *et al.*, 2009; Änkö and Neugebauer, 2012), we initially chose to express each

tRSA-tagged sRNA in *S. venezuelae*. With this approach, tagged sRNA-protein complexes formed *in vivo* could be purified directly from *S. venezuelae* lysates using streptavidin-coated beads (Figure 4.2A). Following affinity purification, bound RBPs would be isolated and analyzed by SDS-PAGE and mass spectrometry.

Each tRSA-tagged sRNA was cloned downstream of the constitutive *ermE** promoter (Schmitt-John and Engels, 1992) before being introduced into *S. venezuelae* on a integrating plasmid vector. Northern blotting experiments were conducted to verify that each tagged sRNA was stably expressed *in vivo.* For these analyses, RNA was harvested from the empty vector strain (carrying pIJ82), the tRSA control strain (carrying pMC260), and each strain carrying a tRSA-sRNA construct (pMC261-264). To determine if the tagged sRNAs were differentially expressed throughout the *S. venezuelae* life cycle, RNA was extracted from each strain during vegetative growth (8 h) and fragmentation (18 h). When total RNA was probed with an aptamer-specific oligonucleotide, no transcripts were detected in any of the *S. venezuelae* strains at either time point. While this may suggest that the tRSA-tagged sRNAs were not expressed *in vivo*, it may also just be a result of inefficient probe hybridization. To distinguish between these two possibilities, we re-probed the northern blots with two sRNA-specific oligonucleotides. As expected, the svr2416- and svr5279-specific probes bound the native sRNAs in all strains; however, there was no evidence of a larger transcript that would correspond to the tagged sRNA, in strains carrying the corresponding tRSA-sRNA construct (Figure 4.3). These results implied that the tRSA-tagged sRNAs were not stably expressed in their *S. venezuelae* host. As a result, we did not elect to carry out the affinity purification of *in vivo* assembled sRNA-protein complexes, as described above.

4.2.2.2 *In vitro* expression of tRSA-tagged sRNAs

As an alternative to expressing the tRSA-tagged sRNAs *in vivo*, we synthesized the transcripts *in vitro*. In this case, *in vitro* transcribed sRNAs would be tethered to streptavidin-coated particles and these incubated with *S. venezuelae* cell lysates (Figure 4.2B). Bound RBPs would then be isolated and identified, as described in section 4.2.2.1. While *in vitro* approaches have been used to identify RBPs (Iioka *et al.*, 2011; Leppek and Stoecklin, 2013; Niemann *et al.*, 2013; Hnilicova *et al.*, 2014), they are only successful if the RNA aptamer adopts the correct conformation and can bind the affinity matrix.

To use this protocol to identify *S. venezuelae* RBPs, tRSA-tagged sRNAs were PCR amplified (using a primer containing the T7 promoter sequence) and transcribed using T7 RNA polymerase *in vitro.* Following synthesis, RNA transcripts were re-folded by heating to 65°C and then slow cooling to 20°C, as outlined by (Iioka *et al.*, 2011). To ensure that the tRSA aptamer had re-folded correctly, we tested the streptavidin binding efficiency of the *in vitro* transcribed, tagged sRNAs. Specifically, tRSA and tRSA-svr103 RNA were re-folded and incubated with streptavidin magnetic particles for at least 30 min. Despite incubation with the streptavidin particles, we could not detect any streptavidin-bound RNA and there was no observable decrease in the amount of unbound RNA (Figure 4.4A). This was likely a result of improper aptamer re-folding, as we had controlled for the integrity of the beads

and our binding conditions using biotinylated *chpH* promoter DNA, which was observed to be successfully bound the beads (Figure 4.4E).

Several other RNA re-folding protocols were tested [see Section 2.4.3 (Tahiri-Alaoui *et al.*, 2002; Leppek and Stoecklin, 2013)]; however, each yielded similar results as above (Figure 4.4B-C), with the exception of the protocol described by (Hnilicova *et al.*, 2014), where RNA denatured at 95°C and then re-folded at 23°C (in folding buffer containing KCl and $MgCl_2$) appeared to bind weakly to the streptavidin particles (Figure 4.4D). Although we were unable to observe a decrease in unbound RNA, we could clearly detect tRSA and tRSA-svr1031 RNA bound to the streptavidin-coated beads.

4.2.2.3 Affinity purification of tRSA-tagged sRNAs

Given the successful re-folding of tRSA and tRSA-svr1031, we used the protocol described by (Hnilicova *et al.*, 2014) to fold each *in vitro* transcribed, tRSA-tagged sRNA. Re-folded RNA was bound to streptavidin magnetic particles before being incubated with crude *S. venezuelae* lysates. Since we did not know when these sRNAs were biologically active (or would bind an RBP), we pooled cell lysates harvested from cells at two developmental stages: vegetative and fragmentation. Lysates were incubated with egg white avidin, to block biotinylated proteins (Said *et al.*, 2009; Iioka *et al.*, 2011), and RNase inhibitor, to prevent RNA degradation (Niemann *et al.*, 2013). It is important to note that the streptavidin aptamer showed no affinity for avidin (Srisawat and Engelke, 2001).

After incubating the crude lysates with the immobilized tRSA-tagged sRNAs, unbound proteins were removed by extensive particle washing (Figure 4.5A). Streptavidin particles were then boiled in SDS loading buffer to elute any RNA-bound proteins. SDS-PAGE followed by Coomassie staining of eluted proteins failed to detect any sRNA-specific RBPs (Figure 4.5B). In fact, the protein fractions eluted from the tRSA-tagged sRNAs and the tRSA RNA were indistinguishable (Figure 4.5B, compare lanes 3 through 7). Enrichment of the same proteins after affinity purification of the tRSA-tagged sRNAs and the tRSA control RNA suggested that *S. venezuelae* may encode either non-specific RBPs, or RBPs that can interact with the tRSA aptamer. To assess these possibilities, we carried out the affinity purification with nucleic acid-free streptavidin particles and particles coated in biotinylated *chpH* promoter DNA. The eluate samples from this experiment were identical to those obtained from the affinity purification of the tRSA-containing RNAs (Figure 4.5B, lanes 8 and 9), suggesting that all of the detectable proteins were binding directly to the streptavidin particles, despite the avidin pre-incubation.

4.2.3 Tagging sRNAs with the MS2 aptamer

Without success using the tRSA aptamer, we have opted instead to replace this tag with an MS2 aptamer. This aptamer is based on an RNA hairpin from the MS2 bacteriophage, which is bound with high affinity by the MS2 coat protein (Carey *et al.*, 1983; Bardwell and Wickens, 1990). The aptamer has been modified to include two MS2 binding hairpins in tandem (Figure 4.1B) (Batey and Kieft, 2007; Said *et al.*, 2009). This approach

has been used successfully to identify RBPs in *E. coli* (Desnoyers and Massé, 2012; Lalaouna *et al.*, 2015) and *Salmonella* (Said *et al.*, 2009).

To tag each *S. venezuelae* sRNA of interest (see Section 4.2.1), we cloned the sRNAs downstream of the modified MS2 aptamer (Said *et al.*, 2009). Introducing the MS2 at the 5'-end of the sRNA transcript ensured that it did not disrupt the native transcription terminator. To facilitate *in vivo* expression, MS2-tagged sRNAs were cloned behind a constitutive promoter, SF14 (Labes *et al.*, 1997), which had been previously used to express a non-coding RNA in *Streptomyces* (Rudolph *et al.*, 2013). The tagged-sRNA constructs were all moved into an integrative vector for transfer into *S. venezuelae*. The expression and stability of these MS2-tagged sRNAs has not yet been tested *in vivo.*

4.3 Discussion

The work described in this chapter has laid the groundwork for a flexible RNA affinity purification protocol that has the potential to be used to identify binding partners (protein or RNA) for any sRNA of interest in *Streptomyces*.

4.3.1 The tRSA aptamer is not ideal for affinity purification

We first attempted to design an affinity purification protocol based on a novel streptavidin aptamer (Iioka *et al.*, 2011). This tRSA aptamer has been used to successfully identify RBPs in other bacteria [*i.e. Salmonella* (Niemann *et al.*, 2013) and *S. aureus* (Zhang *et al.*, 2015)]; however, it did not work well in *Streptomyces.* When tRSA-tagged sRNAs were introduced into *S. venezuelae,* we were unable to detect the transcripts by northern blot analyses. The absence of stable tRSA-sRNA transcripts may be a result of their inefficient expression *in vivo* or their rapid degradation following transcription. The tRSA-tagged sRNA constructs were expressed from the well-studied *Streptomyces* promoter, *ermE** (Schmitt-John and Engels, 1992). This promoter is typically used as a strong, constitutive promoter for protein-encoding genes (Seghezzi *et al.*, 2011), but it has recently been used for the expression of a sRNA in *Streptomyces* - notably the svr2416 orthologue in *S. coelicolor* (Kim *et al.*, 2014). Despite being used to successfully express one sRNA species, we cannot exclude the possibility that P_{ermE*} was not strong enough to promote detectable expression of the tRSA-tagged sRNAs. Alternatively, it is conceivable that the highly structured tRSA-sRNAs were degraded by RNases shortly after transcription. For example, the double-strand-specific RNase III (Nicholson, 2014) in *Streptomyces* may recognize and cleave a hairpin within the structured tRSA aptamer, promoting the further breakdown of the sRNA.

Instead of trying to address these possibilities, we chose to switch to an *in vitro*-based affinity purification protocol since previous studies had suggested that this was successful way to identify RBPs with the tRSA aptamer (Iioka *et al.*, 2011; Niemann *et al.*, 2013; Zhang *et al.*, 2015). However, when *S. venezuelae* tRSA-tagged sRNAs were synthesized and re-folded *in vitro*, they showed almost no affinity for streptavidin. While the tRNA scaffold was originally added to stabilize the streptavidin aptamer conformation (Iioka *et al.*, 2011), it was obviously unable to do so in this case. Even in the presence of

Mg^{2+}, an ion known to stabilize tRNA structures (Agris, 1996), we could not detect proper tRSA-sRNA folding (Figure 4.4A-C). We were only able to detect streptavidin binding for tRSA-tagged sRNAs that had been re-folding using the protocol described by (Hnilicova *et al.*, 2014). Even then, these transcripts only exhibited a weak affinity for streptavidin, which may explain why we were unable to purify any RNA-specific RBPs from *S. venezuelae* cell lysates.

Despite being known to interact with RNAPσ^{HrdB} (Mikulík *et al.*, 2014), we were unable to detect a protein unique to svr3329 using these *in vitro* based assays. Improper folding of the tagged sRNA, which prevented either streptavidin binding or target interaction, is expected to be the cause. This is likely the case for the other tRSA-tagged sRNAs, although we cannot rule out the possibility that they do not interact with a RBP. It is also important to note that we incubated sRNA-coated streptavidin particles with lysates harvested from wild-type *S. venezuelae*. As a result, lysates would have contained endogenous sRNAs that could have sequestered their target away from the streptavidin-immobilized sRNAs. In the future, carrying out the affinity purification using lysates harvested from *S. venezuelae* sRNA mutant strains may be more fruitful.

4.3.2 MS2-tagging sRNAs and future directions

Given the difficulties associated with using the tRSA aptamer, we have selected a modified MS2 tag (Said *et al.*, 2009) to use as an alternative strategy. This aptamer has been used successfully for purifying *in vivo* assembled RNA-protein complexes (Desnoyers and Massé, 2012; Lalaouna *et al.*, 2015), and more recently cellular RNA targets for sRNAs (most recent Masse paper). This type of approach is advantageous, since it is likely to identify biologically relevant interactions that may be missed by *in vitro*-based methods (Said *et al.*, 2009; Änkö and Neugebauer, 2012). For example, transient or weak RNA-protein interactions (like RNase-RNA complexes) are not typically identified in *in vitro* screens, but can be captured by formaldehyde cross-linking *in vivo* (Said *et al.*, 2009). As a result, we hope the MS2-tagged *S. venezuelae* sRNAs will be stable, and properly folded, when expressed *in vivo*.

While each MS2-sRNA construct has been successfully introduced into wild-type *S. venezuelae*, we have not yet determined whether these transcripts are expressed or stable *in vivo*. To promote expression of the MS2-tagged sRNAs in *Streptomyces*, each construct was placed under the control of the strong, constitutive SF14 promoter (Labes *et al.*, 1997). This promoter appears to be much stronger than P_{ermE*}, especially when placed upstream of a non-coding RNA element (Rudolph *et al.*, 2013). As most of the published stable MS2-RNA chimeras have the MS2 tag at their 5' end (Said *et al.*, 2009; Desnoyers and Massé, 2012; Lalaouna *et al.*, 2015), we chose to attach the aptamer sequence to this position for each *S. venezuelae* sRNA. At this site, the tag will not interrupt the native sRNA transcription terminator, or any target recognition sites (Said *et al.*, 2009). We are currently using northern blotting analyses to test whether each MS2-tagged sRNA is stably expressed in *S. venezuelae*.

Once we have established which MS2-tagged sRNAs are stable *in vivo*, we will quickly be able to isolate any interacting protein by affinity purification with the MS2 phage coat protein (Batey and Kieft, 2007). Proteins that bind multiple sRNAs may represent global RNA chaperones and will be prioritized for further characterization (Chapter 5). In the absence of an RNA chaperone, these affinity purification experiments will be important for elucidating sRNA-specific binding proteins and potentially even mRNA interacting partners (Lalaouna *et al.*, 2015) (Chapter 5). Either way, these studies will be important for understanding the regulatory role of these sRNAs in *S. venezuelae*. Applying this technique to other sRNAs will provide general insight into how sRNA-based regulation is mediated in *Streptomyces* and perhaps in related bacteria lacking an Hfq homolog [*e.g. Mycobacterium tuberculosis* (Sun *et al.*, 2002)].

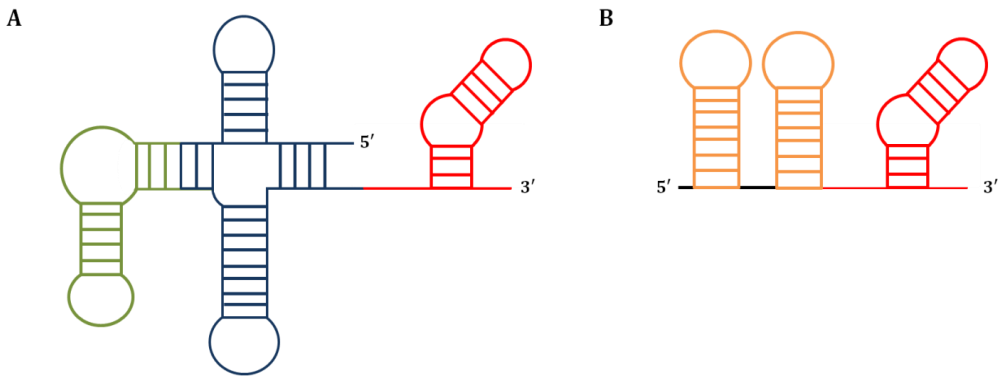

Figure 4.1 sRNA aptamer tags. Schematic representation of the (A) tRSA (Iioka *et al.*, 2011) and (B) MS2 aptamer (Said *et al.*, 2009), which are specific for streptavidin and the phage MS2 coat protein, respectively. Structures representing the streptavidin aptamer (green), the tRNA scaffold (blue) and the MS2 binding hairpins (orange) are indicated. Bait sRNAs (red) were fused to the 3' end of both aptamers.

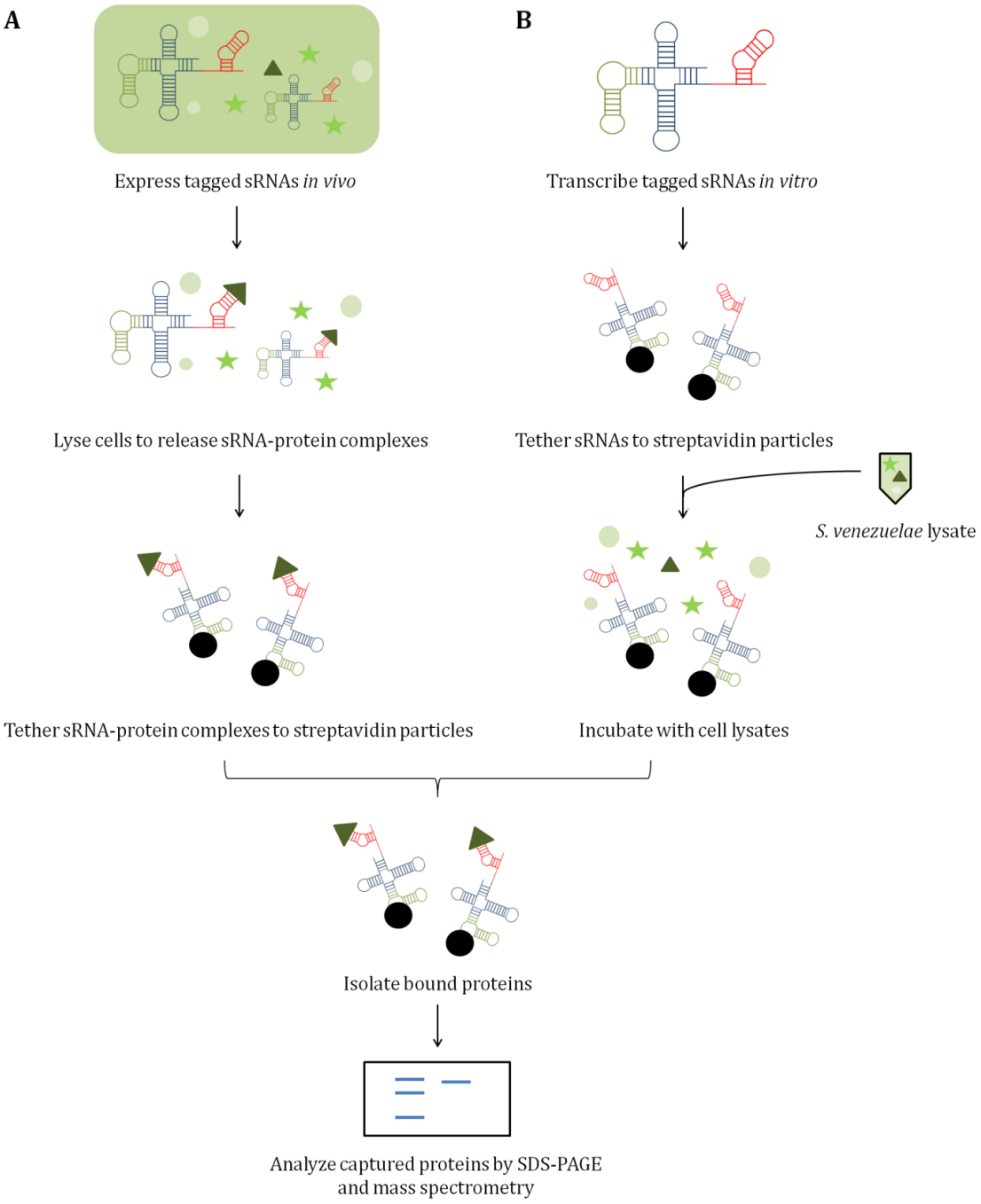

Figure 4.2. RNA pull-down assay. Schematic of the RNA affinity purification protocol used to isolate proteins bound to tRSA-tagged sRNAs expressed (A) *in vivo* or (B) *in vitro* transcribed. The streptavidin aptamer (green) is used to purify the attached sRNA of interest (red) and any interacting proteins. Bound proteins (triangles) are then isolated and analyzed using SDS-PAGE and mass spectrometry. Adapted from (Iioka *et al.*, 2011).

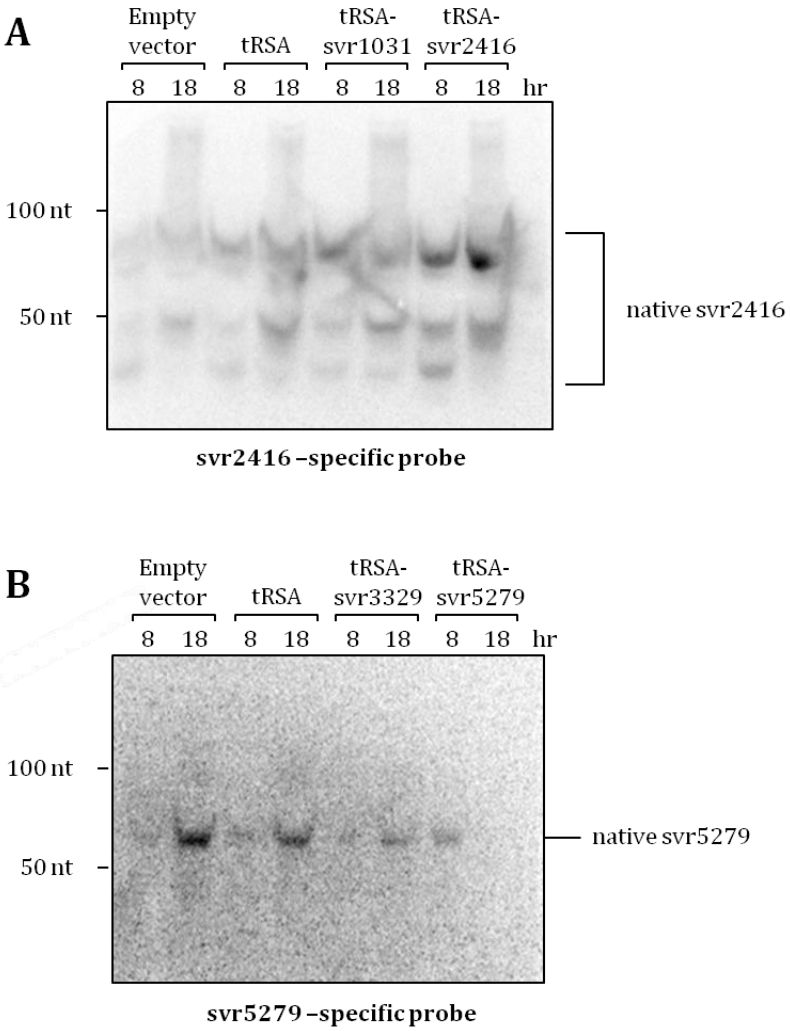

Figure 4.3 Detection of *in vivo* expressed tRSA-tagged sRNAs. Northern blotting analysis of specific tagged sRNAs (tRSA-svr2416 and tRSA-svr5279) expressed in *S. venezuelae.* Total RNA was extracted from *S. venezuelae* carrying the various tagged-sRNA constructs at two distinct time points, as indicated above the northern blots. RNA was probed with either a (A) svr2416-specific or (B) svr5279-specific oligonucleotide. As a control, RNA isolated from the tRSA and the empty vector strain was also probed with both oligonucleotides. We could detect the native sRNA (svr2416 or svr5279) in all strains tested; however, this was the only transcript observed in strains carrying the tagged-sRNA constructs (lanes 7 and 8 of both blots).

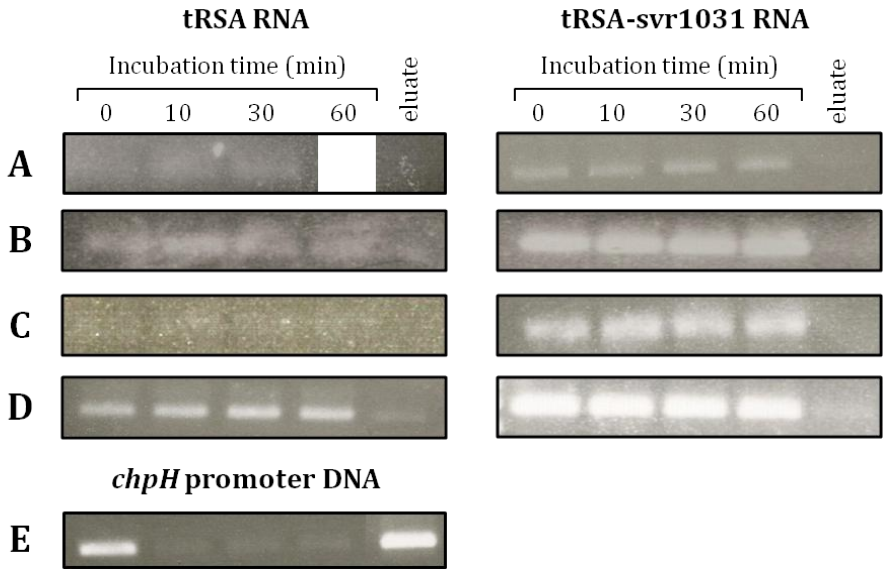

Figure 4.4 *In vitro* transcription and re-folding of tRSA-tagged sRNAs. tRSA and tRSA-svr1031 RNA were synthesized by *in vitro* transcription and then re-folded using protocols described by: (A) (Iioka *et al.*, 2011), (B) (Leppek and Stoecklin, 2013), (C) (Tahiri-Alaoui *et al.*, 2002) and (D) (Hnilicova *et al.*, 2014). Once re-folded, RNA was incubated with streptavidin particles for at least 30 minutes. Particle supernatant was removed after 10, 30 and 60 min of incubation, as indicated, in order to assess the quantity of unbound RNA. Following incubation, beads were boiled in fresh buffer and bound RNA was collected in the eluate. Only RNA re-folded in (D) bound to the streptavidin beads and was detected in the eluate. *chpH* promoter DNA (E) was also included as a control for streptavidin integrity and binding.

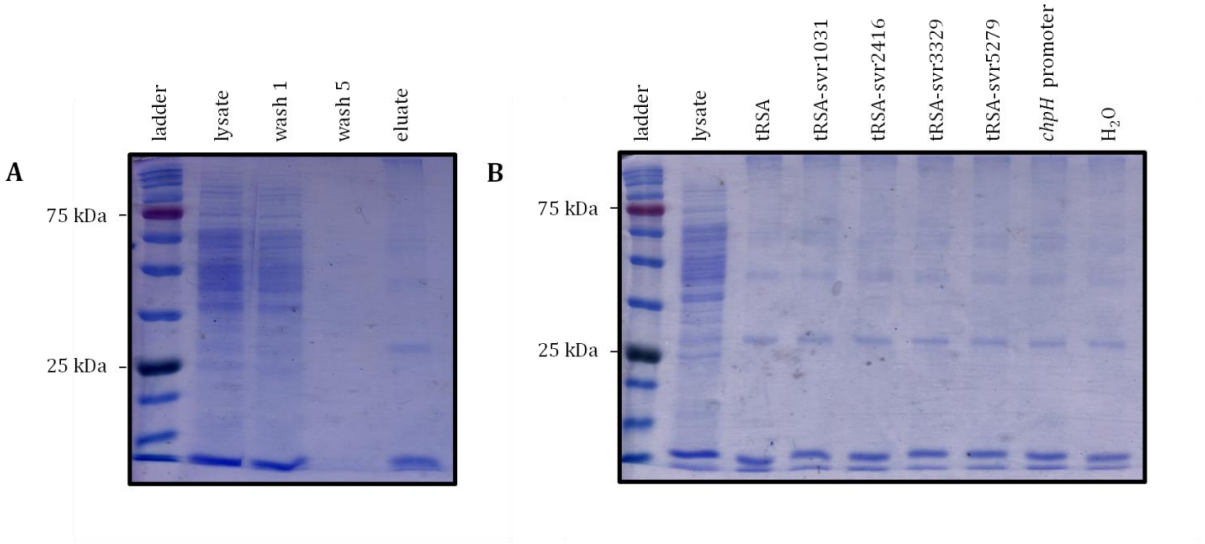

Figure 4.5 Affinity purification of *in vitro* transcribed tRSA-tagged sRNAs. (A) Synthesized tRSA RNA was immobilized onto streptavidin magnetic particles before being incubated with *S. venezuelae* lysates pre-blocked with egg white avidin (lane 2, both gels). Following incubation, particles were washed with fresh buffer to remove unbound proteins. Protein samples were collected from the particle supernatant after the first (lane 3) and last wash (lane 4). Streptavidin particles were then boiled in loading buffer and bound proteins were collected in the eluate (lane 5). (B) The affinity purification was also carried out with each tRSA-tagged sRNA, *chpH* promoter DNA and beads lacking any bound nucleic acid (H_2O; lane 9). Wash fractions were consistent with (A) and elution fractions are shown in lanes 4-8. There were no unique bands detected in any elution sample. For reference, the BLUeye prestained protein ladder (GeneDirex) was included (lane 1, both gels).

Chapter 5: Summary and future directions

5.1 Summary of research

The *Streptomyces* life cycle comprises several different morphological and metabolic states (Flärdh and Buttner, 2009), all of which are subject to multiple levels of regulation (Moody *et al.*, 2013). To further characterize the transcriptional and post-transcriptional networks required to coordinate these complex processes, we investigated the two broad RNA-based regulatory systems by (1) defining transcription terminators and (2) characterizing interacting partners (*e.g.* RNAs and proteins) of diverse sRNAs.

Driven by the fact that transcription termination is poorly understood in the streptomycetes, we used complementary bioinformatic and experimental approaches to characterize intrinsic terminators. Initially, we developed a novel *in silico* prediction tool that identified hundreds of biologically relevant termination sites in three divergent *Streptomyces* species. Through comparisons with terminators predicted by other algorithms [*e.g.* WebGeSTer (Mitra *et al.*, 2011)], we have shown that these termination sites primarily correspond to non-canonical structures. Aside from a preference for a longer hairpin stem, these terminators do not appear to have any other distinguishing characteristics. Following the computational analyses, we prioritized three diverse structures for characterization *in vivo* using a fluorescent reporter construct adapted from (Chen *et al.*, 2013); however, difficulties implementing this reporter system in *Streptomyces* have delayed *in vivo* quantification and comparison of terminator strength.

At the post-transcriptional level, we investigated the role of RBPs in sRNA-mediated regulation. In other bacterial species, RBPs are known to be important for mediating sRNA-mRNA interactions [*e.g.* Hfq (Vogel and Luisi, 2011)] and modulating sRNA stability [*e.g.* CsrD (Suzuki *et al.*, 2006)]. To identify sRNA-interacting proteins in *Streptomyces*, we have designed a sRNA-based affinity purification protocol similar to previous methods (Said *et al.*, 2009; Iioka *et al.*, 2011). Early work focused on optimizing a protocol that involved tagging four well conserved sRNAs (svr1031, svr2416, svr3329 and svr5279) with a highly structured streptavidin aptamer [*i.e.* the tRSA aptamer (Iioka *et al.*, 2011)]. We were unable to carry out the affinity purification with *in vivo* expressed or *in vitro* transcribed tRSA-tagged sRNAs, since the tagged sRNAs were not stably expressed in *S. venezuelae* and the sRNAs synthesized *in vitro* did not re-fold correctly. In an attempt to increase sRNA stability, we have recently replaced the tRSA aptamer with an MS2 binding aptamer, although the functionality of these tagged sRNAs has yet to be determined *in vivo* or *in vitro*.

5.2 Future directions

5.2.1 Understanding transcription termination in the streptomycetes

Our bioinformatic analyses have successfully laid the groundwork for understanding transcription in *Streptomyces* bacteria. We have identified numerous potential intrinsic terminators *in silico*; however, we have yet to assess their performance *in vivo*. As a result, our short term goal is to characterize the small subset of diverse terminators prioritized

previously. Since the fluorescent reporter construct was not effective in *S. venezuelae* or *S. coelicolor*, we have proposed a new *gusA*-based system to measure termination efficiency (Figure 3.6). While these assays would be able to confirm termination *in vivo,* they would not be able to provide an accurate measure of terminator strength since this system lacks an internal reference (*i.e.* a reporter upstream of the terminator). Consequently, complementary *in vitro* transcription experiments would be necessary to quantify termination efficiency. Terminators that demonstrate effective termination in these assays would be important additions to the *Streptomyces* genetic toolbox, where they could be exploited for synthetic purposes. With very few terminators previously characterized (Pulido and Jimenez, 1987; Ingham *et al.*, 1995), these additions would allow for the creation of larger and more complex expression vectors; such constructs would be particularly useful in the streptomycetes, as there are only a few sites where exogenous DNA can be integrated into the genome (Kuhstoss *et al.*, 1991; Gregory *et al.*, 2003; Fayed *et al.*, 2014). Expanding the collection of characterized terminators would also establish a data set that could be used to train our current prediction algorithm to identify more biologically meaningful terminators.

Longer term, the goal is to gain a more comprehensive understanding of how transcription terminates in the streptomycetes. Specifically, it will be important to determine which terminator characteristics (*e.g.* stem length, loop size and tail composition) are most important for efficient termination. This could be achieved by characterizing a large set of identified terminators or systematically mutating features of a small subset of terminators [as in (Chen *et al.*, 2013)], although a combination of the two approaches would likely be the most informative. These analyses would provide us with a consensus terminator motif/structure for *Streptomyces*, which could easily be searched for in a broad range of related species (*e.g. Mycobacterium*). Incorporating the common terminator sequence features into current prediction algorithms would not only improve the accuracy of terminator identification, but also the reliability of genome annotation in general (Kingsford *et al.*, 2007). If the aforementioned *in vivo* analyses are consistent with the computation predictions, we would expect non-canonical terminators to be more prevalent than canonical structures. It would be interesting to test whether these terminators are functional in species where canonical terminators are abundant [*i.e. E. coli* (Peters *et al.*, 2009)] to determine if the mechanisms of termination are inherently different between these divergent species.

5.2.2 Identifying an RNA chaperone in *Streptomyces*

To identify RBPs, we have designed an MS2 affinity purification system, which has not yet been completely implemented in *Streptomyces.* The short term goals of this work will therefore focus on testing whether the MS2-tagged sRNAs that we had introduced into *S. venezuelae* are functional *in vivo*. After sRNA expression has been validated using northern blotting experiments, the affinity purification can be performed to identify interacting RBPs.

Since proteins that interact with multiple sRNAs could be candidate RNA chaperones, these would be prioritized for downstream analyses. Chaperone characterization would begin with the creation of a *S. venezuelae* deletion strain, where changes in morphology and secondary metabolism can be easily detected. Since the *E. coli hfq* mutant displays an increased sensitivity to environmental stresses (Tsui *et al.*, 1994), the phenotype of the chaperone deletion strain would also be monitored under different stress conditions, starting with those associated with sRNA activity in other systems [*e.g.* iron starvation (Massé and Gottesman, 2002) and cell envelope stress (Hobbs *et al.*, 2010)]. In addition to phenotypic analyses, co-immunoprecipitation experiments could be used to identify other sRNAs (via RNA sequencing) that bind to the RNA chaperone. Analogous methods have been used successfully to identify Hfq-binding sRNAs in both *E. coli* (Zhang *et al.*, 2003) and *B. subtilis* (Dambach *et al.*, 2013). In both species, these approaches allowed for the identification of several novel sRNAs (Zhang *et al.*, 2003; Dambach *et al.*, 2013), raising the possibility that these investigations could also expand the sRNA repertoire of *Streptomyces* bacteria. In particular, RNA sequencing could permit the detection of lowly expressed sRNAs or stress-induced sRNAs that were missed in previous RNA sequencing experiments (Vockenhuber *et al.*, 2011; Moody *et al.*, 2013).

The power of the affinity purification approach that we propose to use with the MS2-tagged sRNAs is that it has the potential to uncover sRNA-specific mRNAs or protein interacting partners. For example, we would expect to detect RNAPσ^{HrdB} binding to svr3329 since the same interaction has been observed in *S. coelicolor* (Mikulík *et al.*, 2014). This approach could also be combined with RNA sequencing [as in (Lalaouna *et al.*, 2015)] to identify RNAs bound to each sRNA. Following identification, all sRNA interactions would be validated with reverse affinity purification experiments which use the protein or RNA target as the bait (Lalaouna *et al.*, 2015). The identity of any confirmed targets would provide insight into the biological function of these sRNAs and would guide further investigations into the sRNA's regulatory role.

5.2.3 Interplay between transcription terminators and RBPs

Although originally viewed as distinct regulators, there have been some interesting connections made between transcription terminators and RBPs. In *E. coli*, the proximal face of Hfq binds to 3' U-rich tails of sRNA intrinsic terminators (Sauer and Weichenrieder, 2011; Sauer *et al.*, 2012). In *Streptomyces*, our prediction algorithm detected well-defined termination sites downstream of over 30 potential sRNAs (Section 3.2.2); however, due to their location they were not considered in the downstream analyses. Future studies could compare the structure associated with these terminators, as they might represent a universal feature recognized by a RNA chaperone. It would also be worthwhile including those sRNAs with well-defined terminators in the MS2 affinity purification experiments.

Additionally, our *in silico* analyses revealed that expression profiles associated with mRNA terminators were consistent with the accumulation of a short transcript that could act as an sRNA (Section 3.2.2). The fact that mRNA 3' regions (Zhang *et al.*, 2003; Sittka *et al.*,

2008) and sRNAs derived from mRNA 3' UTRs (Chao *et al.*, 2012) co-immunoprecipitate with Hfq in other species (*e.g. Salmonella* and *E. coli*), suggests that these *Streptomyces* transcripts could be targets of an RNA chaperone. Including several strong mRNA terminators in the affinity purification assays may therefore also be a fruitful approach for identifying a streptomycete chaperone that is the functional equivalent of Hfq.

5.3 Conclusions

Overall, this work has laid the foundation for identifying and characterizing intrinsic terminators and RBPs in *Streptomyces*. Future studies will provide further insights into transcription termination and sRNA-mediated regulation in these bacteria and their relatives (*e.g. Mycobacterium*). This knowledge will be used to expand the limited collection of synthetic biology tools available for the streptomycetes, and our ability to manipulate the activity of RNA-based regulators. Ultimately, our goal is to develop a cohesive understanding of the transcriptional and post-transcriptional systems required to regulate multicellular development and metabolic diversity in these bacteria, and exploit this understanding for medical and agricultural purposes.

References

Adhya, S., and Gottesman, M. (1978) Control of transcription termination. *Annu Rev Biochem* **47**: 967–996.

Agris, P.F. (1996) The importance of being modified: roles of modified nucleosides and Mg^{2+} in RNA structure and function. *Prog Nucleic Acid Res Mol Biol* **53**: 79–129.

Alifano, P., Rivellini, F., Limauro, D., Bruni, C.B., and Carlomagno, M.S. (1991) A consensus motif common to all Rho-dependent prokaryotic transcription terminators. *Cell* **64**: 553–563.

Angert, E.R. (2005) Alternatives to binary fission in bacteria. *Nat Rev Microbiol* **3**: 214–224.

Änkö, M.L., and Neugebauer, K.M. (2012) RNA-protein interactions *in vivo*: global gets specific. *Trends Biochem Sci* **37**: 255–262.

Arraiano, C.M., Andrade, J.M., Domingues, S., Guinote, I.B., Malecki, M., Matos, R.G., *et al.* (2010) The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol Rev* **34**: 883–923.

Arrebola, E., Carrión, V.J., Cazorla, F.M., Pérez-García, A., Murillo, J., and Vicente, A. de (2012) Characterisation of the *mgo* operon in *Pseudomonas syringae* pv. *syringae* UMAF0158 that is required for mangotoxin production. *BMC Microbiol* **12**: 10.

Arthur, D.C., Ghetu, A.F., Gubbins, M.J., Edwards, R.A., Frost, L.S., and Glover, J.N.M. (2003) FinO is an RNA chaperone that facilitates sense-antisense RNA interactions. *EMBO J* **22**: 6346–6355.

Assche, E. Van, Puyvelde, S. Van, Vanderleyden, J., and Steenackers, H.P. (2015) RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Front Microbiol* **6**: 1–16.

Ayer, S.W., McInnes, A.G., Thibault, P., Walter, J.A., Doull, J.L., Parnell, T., and Vining, L.C. (1991) Jadomycin, a novel 8*H*-Benz[*b*]oxazolo[3,2-*f*]phenanthridine antibiotic from *Streptomyces venezuelae* ISP5230. *Tetrahedron Lett* **32**: 6301–6304.

Babitzke, P. (2004) Regulation of transcription attenuation and translation initiation by allosteric control of an RNA-binding protein: The *Bacillus subtilis* TRAP protein. *Curr Opin Microbiol* **7**: 132–139.

Bandyra, K.J., Said, N., Pfeiffer, V., Górna, M.W., Vogel, J., and Luisi, B.F. (2012) The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Mol Cell* **47**: 943–953.

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

Batey, R.T., and Kieft, J.S. (2007) Improved native affinity purification of RNA. *RNA* **13**: 1384–1389.

Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141–147.

Bérdy, J. (2005) Bioactive microbial metabolites. *J Antibiot* **58**: 1–26.

Bibb, M.J., Domonkos, Á., Chandra, G., and Buttner, M.J. (2012) Expression of the chaplin and rodlin hydrophobic sheath proteins in *Streptomyces venezuelae* is controlled by σ^{BldN} and a cognate anti-sigma factor, RsbN. *Mol Microbiol* **84**: 1033–1049.

Blencowe, B.J., Sproat, B.S., Ryder, U., Barabino, S., and Lamond, A.I. (1989) Antisense probing of the human U4/U6 snRNP with biotinylated 2'-OMe RNA oligonucleotides. *Cell* **59**: 531–539.

Boudvillain, M., Figueroa-Bossi, N., and Bossi, L. (2013) Terminator still moving forward: Expanding roles for Rho factor. *Curr Opin Microbiol* **16**: 118–124.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248–254.

Bralley, P., Aseem, M., and Jones, G.H. (2014) SCO5745, a bifunctional RNase J ortholog, affects antibiotic production in *Streptomyces coelicolor*. *J Bacteriol* **196**: 1197–1205.

Brantl, S. (2007) Regulatory mechanisms employed by *cis*-encoded antisense RNAs. *Curr Opin Microbiol* **10**: 102–109.

Bremer, H., and Yuan, D. (1968) RNA chain growth-rate in *Escherichia coli*. *J Mol Biol* **38**: 163–180.

Browning, D.F., and Busby, S.J. (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**: 57–65.

Cardinale, C.J., Washburn, R.S., Tadigotla, V.R., Brown, L.M., Gottesman, M.E., and Nudler, E. (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* **320**: 935–938.

Carey, J., Cameron, V., Haseth, P.L. de, and Uhlenbeck, O.C. (1983) Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. *Biochemistry* **22**: 2601–2610.

Cavanagh, A.T., and Wassarman, K.M. (2014) 6S RNA, a global regulator of transcription in *Escherichia coli*, *Bacillus subtilis*, and beyond. *Annu Rev Microbiol* **68**: 45–60.

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

Chen, C.Y., Galluppi, G.R., and Richardson, J.P. (1986) Transcription termination at lambda tR1 is mediated by interaction of Rho with specific single-stranded domains near the 3' end of *cro* mRNA. *Cell* **46**: 1023–1028.

Chen, Y.-J., Liu, P., Nielsen, A.A.K., Brophy, J.A.N., Clancy, K., Peterson, T., and Voigt, C.A. (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Methods* **10**: 659–664.

Chomcyzynski, P., and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**: 156–159.

Christiansen, J.K., Larsen, M.H., Ingmer, H., Søgaard-Andersen, L., and Kallipolitis, B.H. (2004) The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence. *J Bacteriol* **186**: 3355–3362.

Ciampi, M.S. (2006) Rho-dependent terminators and transcription termination. *Microbiology* **152**: 2515–2528.

Czyz, A., Mooney, R.A., Iaconi, A., and Landick, R. (2014) Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *MBio* **5**: e00931–14.

D'Alia, D., Nieselt, K., Steigele, S., Müller, J., Verburg, I., and Takano, E. (2010) Noncoding RNA of glutamine synthetase I modulates antibiotic production in *Streptomyces coelicolor* A3(2). *J Bacteriol* **192**: 1160–1164.

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

Dangerfield, J.A., Windbichler, N., Salmons, B., Günzburg, W.H., and Schröder, R. (2006) Enhancement of the StreptoTag method for isolation of endogenously expressed proteins with complex RNA binding targets. *Electrophoresis* **27**: 1874–1877.

Davies, B.W., and Walker, G.C. (2008) A highly conserved protein of unknown function is required by *Sinorhizobium meliloti* for symbiosis and environmental stress protection. *J Bacteriol* **190**: 1118–1123.

Desnoyers, G., and Massé, E. (2012) Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes Dev* **26**: 726–739.

Durand, S., Tomasini, A., Braun, F., Condon, C., and Romby, P. (2015) sRNA and mRNA turnover in Gram-positive bacteria. *FEMS Microbiol Rev* 1–15.

Elliot, M.A., Buttner, M.J., and Nodwell, J.R. (2008) Multicellular Development in *Streptomyces*. In Whitworth, D.D. editor. Myxobacteria Multicellularity Differentiation. ASM press, Washington, D.C.

Epshtein, V., Cardinale, C.J., Ruckenstein, A.E., Borukhov, S., and Nudler, E. (2007) An allosteric path to transcription termination. *Mol Cell* **28**: 991–1001.

Epshtein, V., Dutta, D., Wade, J., and Nudler, E. (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463**: 245–249.

Fayed, B., Younger, E., Taylor, G., and Smith, M.C.M. (2014) A novel *Streptomyces* spp. integration vector derived from the *S. venezuelae* phage, SV1. *BMC Biotechnol* **14**: 51.

Flärdh, K. (2003) Essential role of DivIVA in polar growth and morphogenesis in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **49**: 1523–1536.

Flärdh, K., and Buttner, M.J. (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol* **7**: 36–49.

Franze de Fernandez, M.T., Eoyang, L., and August, J.T. (1968) Factor fraction required for the synthesis of bacteriophage Qβ-RNA. *Nature* **219**: 588–590.

Fritsch, T.E., Siqueira, F.M., and Schrank, I.S. (2015) Intrinsic terminators in *Mycoplasma hyopneumoniae* transcription. *BMC Genomics* **16**: 273.

Gardner, P.P., Barquist, L., Bateman, A., Nawrocki, E.P., and Weinberg, Z. (2011) RNIE: Genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res* **39**: 5845–5852.

Gautheret, D., and Lambert, A. (2001) Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J Mol Biol* **313**: 1003–1011.

Gentz, R., Langner, A., Chang, A.C., Cohen, S.N., and Bujard, H. (1981) Cloning and analysis of strong promoters is made possible by the downstream placement of a RNA termination signal. *Proc Natl Acad Sci USA* **78**: 4936–4940.

Glazebrook, M. a, Doull, J.L., Stuttard, C., and Vining, L.C. (1990) Sporulation of *Streptomyces venezuelae* in submerged cultures. *J Gen Microbiol* **136**: 581–588.

Goodfellow, M., and Williams, S.T. (1983) Ecology of actinomycetes. *Annu Rev Microbiol* **37**: 189–216.

Gottesman, S., and Storz, G. (2011) Bacterial small RNA regulators: Versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* **3**: 1–16.

Gregory, M.A., Till, R., and Smith, M.C.M. (2003) Integration site for *Streptomyces* phage ΦBT1 and development of site-specific integrating vectors. *J Bacteriol* **185**: 5320–5323.

Gusarov, I., and Nudler, E. (1999) The mechanism of intrinsic transcription termination. *Mol Cell* **3**: 495–504.

Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. (2003) PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA* **100**: 1541–1546.

Gutiérrez, P., Li, Y., Osborne, M.J., Gutie, P., Pomerantseva, E., Liu, Q., and Gehring, K. (2005) Solution structure of the carbon storage regulator protein CsrA from *Escherichia coli*. *J Bacteriol* **187**: 3496–3501.

Haiser, H.J., Karginov, F. V., Hannon, G.J., and Elliot, M.A. (2008) Developmentally regulated cleavage of tRNAs in the bacterium *Streptomyces coelicolor*. *Nucleic Acids Res* **36**: 732–741.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**: 557–580.

Harshey, R.M., and Ramakrishnan, T. (1977) Rate of ribonucleic acid chain growth in *Mycobacterium tuberculosis* H37Rv. *J Bacteriol* **129**: 616–622.

Hart, C.M., and Roberts, J.W. (1991) Rho-dependent transcription termination. Characterization of the requirement for cytidine in the nascent transcript. *J Biol Chem* **266**: 24140–24148.

Hartmann, R.K., Bindereif, A., Schon, A., and Westhof, E. (2005) Handbook of RNA biochemistry. Vol. 2. Germany: Wiley–VCH, Weinheim.

Hartmuth, K., Vornlocher, H.-P., and Luhrmann, R. (2004) Tobramycin affinity tag purification of spliceosomes. *Methods Mol Biol* **257**: 44–64.

Harz, C.O. (1877) Actinomyces bovis, ein neuer Schimmel in den Geweben des Rindes. *Jahresbericht der Kaiserlinchen Cent Munchen* 125–140.

Henkin, T.M., and Yanofsky, C. (2002) Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *BioEssays* **24**: 700–707.

Hindra, Moody, M.J., Jones, S.E., and Elliot, M. A. (2014) Complex intra-operonic dynamics mediated by a small RNA in *Streptomyces coelicolor*. *PLoS One* **9**: e85856.

Hippel, P.H. von (1998) An integrated model of the transcription complex in elongation, termination, and editing. *Science* **281**: 660–665.

Hnilicova, J., Matejckova, J.J., Sikova, M., Pospisil, J., Halada, P., Panek, J., and Krasny, L. (2014) Ms1, a novel sRNA interacting with the RNA polymerase core in mycobacteria. *Nucleic Acids Res* **42**: 11763–11776.

Hobbs, E.C., Astarita, J.L., and Storz, G. (2010) Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: Analysis of a bar-coded mutant collection. *J Bacteriol* **192**: 59–67.

Hobbs, G., Frazer, C.M., Gardner, D.C.J., Cullum, J.A., and Oliver, S.G. (1989) Dispersed growth of *Streptomyces* in liquid culture. *Appl Microbiol Biotechnol* **31**: 272–277.

Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E., and Groisman, E.A. (2012) Riboswitch control of Rho-dependent transcription termination. *Proc Natl Acad Sci USA* **109**: 5376–5381.

Hopwood, D.A. (1957) Genetic recombination in *Streptomyces coelicolor*. *J Gen Microbiol* **16**: ii–iii.

Hopwood, D.A. (1999) Forty years of genetics with *Streptomyces*: From *in vivo* through *in vitro* to *in silico*. *Microbiology* **145**: 2183–2202.

Huang, H. (2007) Design and characterization of artifical transcriptional terminators. Master's thesis, Massachesetts Institute of Technology.

Huang, J., Shi, J., Molle, V., Sohlberg, B., Weaver, D., Bibb, M.J., *et al.* (2005) Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Mol Microbiol* **58**: 1276–1287.

Iioka, H., Loiselle, D., Haystead, T.A., and Macara, I.G. (2011) Efficient detection of RNA-protein interactions using tethered RNAs. *Nucleic Acids Res* **39**: e53.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., *et al.* (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**: 526–531.

Ingham, C.J., Hunter, I.S., and Smith, M.C. (1995) Rho-independent terminators without 3' poly-U tails from the early region of actinophage øC31. *Nucleic Acids Res* **23**: 370–376.

Ingham, C.J., Hunter, I.S., and Smith, M.C.M. (1996) Isolation and sequencing of the *rho* gene from *Streptomyces lividans* ZX7 and characterization of the RNA-dependent NTPase activity of the overexpressed protein. *J Biol Chem* **271**: 21803–21807.

Jin, D.J., Burgess, R.R., Richardson, J.P., and Gross, C.A. (1992) Termination efficiency at Rho-dependent terminators depends on kinetic coupling between RNA polymerase and Rho. *Proc Natl Acad Sci USA* **89**: 1453–1457.

Johnson, C.M., Chen, Y., Lee, H., Ke, A., Weaver, K.E., and Dunny, G.M. (2014) Identification of a conserved branched RNA structure that functions as a factor-independent terminator. *Proc Natl Acad Sci USA* **111**: 3573–3578.

Jones, G.H., and Bibb, M.J. (1996) Guanosine pentaphosphate synthetase from *Streptomyces antibioticus* is also a polynucleotide phosphorylase. *J Bacteriol* **178**: 4281–4288.

Jones, S.E., Leong, V., Ortega, J., and Elliot, M.A. (2014) Development, antibiotic production, and ribosome assembly in *Streptomyces venezuelae* are impacted by RNase J and RNase III deletion. *J Bacteriol* **196**: 4253–4267.

Jousselin, A., Metzinger, L., and Felden, B. (2009) On the facultative requirement of the bacterial RNA chaperone, Hfq. *Trends Microbiol* **17**: 399–405.

Khmelinskii, A., Keller, P.J., Bartosik, A., Meurer, M., Barry, J.D., Mardin, B.R., *et al.* (2012) Tandem fluorescent protein timers for *in vivo* analysis of protein dynamics. *Nat Biotechnol* **30**: 708–714.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000) Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich, United Kingdom.

Kim, H.M., Shin, J.H., Cho, Y.B., and Roe, J.H. (2014) Inverse regulation of Fe- and Ni-containing SOD genes by a Fur family regulator Nur through small RNA processed from 3'UTR of the *sodF* mRNA. *Nucleic Acids Res* **42**: 2003–2014.

Kingsford, C.L., Ayanbule, K., and Salzberg, S.L. (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* **8**: R22.

Komarova, A. V, Tchufistova, L.S., Boni, I. V, and Dreyfus, M. (2005) AU-rich sequences within 5′ untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *J Bacteriol* **187**: 1344–1349.

Komissarova, N., Becker, J., Solter, S., Kireeva, M., and Kashlev, M. (2002) Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Mol Cell* **10**: 1151–1162.

Kondo, M., Gallerani, R., and Weissmann, C. (1970) Subunit structure of Qβ replicase. *Nature* **228**: 525–527.

Kuhstoss, S., Richardson, M.A., and Rao, R.N. (1991) Plasmid cloning vectors that integrate site-specifically in Streptomyces spp. *Gene* **97**: 143–146.

Labes, G., Bibb, M., and Wohlleben, W. (1997) Isolation and characterization of a strong promoter element from the *Streptomyces ghanaensis* phage 119 using the gentamicin resistance gene (*aacC1*) of Tn*1696* as reporter. *Microbiology* **143**: 1503–1512.

Lalaouna, D., Carrier, M.-C., Semsey, S., Brouard, J.-S., Wang, J., Wade, J.T., and Massé, E. (2015) A 3′ external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. *Mol Cell* **58**: 393–405.

Larson, M.H., Greenleaf, W.J., Landick, R., and Block, S.M. (2008) Applied force reveals mechanistic and energetic details of transcription termination. *Cell* **132**: 971–982.

Lau, L.F., Roberts, J.W., and Ray, W. (1983) RNA polymerase pausing and transcript release at the λRl terminator. *J Biol Chem* **25**: 9391–9397.

Lee, D.N., Phung, L., Stewart, J., and Landick, R. (1990) Transcription pausing by *Escherichia coli* RNA polymerase is modulated by downstream DNA sequences. *J Biol Chem* **265**: 15145–15153.

Lee, K., and Cohen, S.N. (2003) A *Streptomyces coelicolor* functional orthologue of *Escherichia coli* RNase E shows shuffling of catalytic and PNPase-binding domains. *Mol Microbiol* **48**: 349–360.

Leela, J.K., Syeda, A.H., Anupama, K., and Gowrishankar, J. (2013) Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*. *Proc Natl Acad Sci USA* **110**: 258–263.

Leppek, K., and Stoecklin, G. (2013) An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res* **42**: 1–15.

Lingner, J., and Cech, T.R. (1996) Purification of telomerase from Euplotes aediculatus: requirement of a primer 3' overhang. *Proc Natl Acad Sci USA* **93**: 10712–10717.

Liu, M.Y., Yang, H., and Romeo, T. (1995) The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability. *J Bacteriol* **177**: 2663–2672.

Liu, N., Niu, G., Xie, Z., Chen, Z., Itzek, A., Kreth, J., *et al.* (2015) The *Streptococcus mutans irvA* gene encodes a *trans*-acting riboregulatory mRNA. *Mol Cell* **57**: 179–190.

Liu, Y., Wu, N., Dong, J., Gao, Y., Zhang, X., Mu, C., *et al.* (2010) Hfq is a global regulator that controls the pathogenicity of *Staphylococcus aureus*. *PLoS One* **5**: e13069.

Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., Case, D.A., and Sampath, R. (2001) RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res* **29**: 4724–4735.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and NacNeil, T. (1992) Analysis of *Streptomyces avermitilis* genes required for acermectin biosynthesis utilizing a novel integration vector. *Genes (Basel)* **111**: 61–68.

Mandal, M., and Breaker, R.R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* **11**: 29–35.

Massé, E., and Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci USA* **99**: 4620–4625.

McNeil, M.M., and Brown, J.M. (1994) The medically important aerobic actinomycetes: Epidemiology and microbiology. *Clin Microbiol Rev* **7**: 357–417.

Mehta, A., and Driscoll, D.M. (1998) A sequence-specific RNA-binding protein complements Apobec-1 to edit apolipoprotein B mRNA. *Mol Cell Biol* **18**: 4426–4432.

Menouni, R., Champ, S., Espinosa, L., Boudvillain, M., and Ansaldi, M. (2013) Transcription termination controls prophage maintenance in *Escherichia coli* genomes. *Proc Natl Acad Sci USA* **110**: 14414–14419.

Merzlyak, E.M., Goedhart, J., Shcherbo, D., Bulina, M.E., Shcheglov, A.S., Fradkov, A.F., *et al.* (2007) Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Methods* **4**: 555–557.

Mikulík, K., Bobek, J., Zídková, J., and Felsberg, J. (2014) 6S RNA modulates growth and antibiotic production in *Streptomyces coelicolor*. *Appl Microbiol Biotechnol* **98**: 7185–7197.

Miller, J.H. (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Mitra, A., Angamuthu, K., Jayashree, H.V., and Nagaraja, V. (2009) Occurrence, divergence and evolution of intrinsic terminators across Eubacteria. *Genomics* **94**: 110–116.

Mitra, A., Kesarwani, A.K., Pal, D., and Nagaraja, V. (2011) WebGeSTer DB - A transcription terminator database. *Nucleic Acids Res* **39**: 129–135.

Møller, T., Franch, T., Højrup, P., Keene, D.R., Bächinger, H.P., Brennan, R.G., and Valentin-Hansen, P. (2002) Hfq: A bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell* **9**: 23–30.

Moody, M.J., Young, R.A., Jones, S.E., and Elliot, M.A. (2013) Comparative analysis of non-coding RNAs in the antibiotic-producing *Streptomyces* bacteria. *BMC Genomics* **14**: 558.

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

Murakami, T., Holt, T.G., and Thompson, C.J. (1989) Thiostrepton-induced gene expression in *Streptomyces lividans*. *J Bacteriol* **171**: 1459–1466.

Myronovskyi, M., Welle, E., Fedorenko, V., and Luzhetskyy, A. (2011) β-Glucuronidase as a sensitive and versatile reporter in Actinomycetes. *Appl Environ Microbiol* **77**: 5370–5383.

Naville, M., and Gautheret, D. (2009) Transcription attenuation in bacteria: theme and variations. *Br Funct Genomics Proteomics* **8**: 482–492.

Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A., and Gautheret, D. (2011) ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA Biol* **8**: 11–13.

Nett, M., Ikeda, H., and Moore, B.S. (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Reports* **26**: 1362–1384.

Nicholson, A.W. (2014) Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdiscip Rev RNA* **5**: 31–48.

Niemann, G.S., Brown, R.N., Mushamiri, I.T., Nguyen, N.T., Taiwo, R., Stufkens, A., *et al.* (2013) RNA type III secretion signals that require Hfq. *J Bacteriol* **195**: 2119–2125.

Nudler, E., and Gottesman, M.E. (2002) Transcription termination and anti-termination in *E. coli*. *Genes to Cells* **7**: 755–768.

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., *et al.* (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol* **190**: 4050–4060.

Opdyke, J.A., Kang, J., and Storz, G. (2004) GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J Bacteriol* **186**: 6698–6705.

Paget, M.S.B., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. (1999) Evidence that the extracytoplasmic function sigma factor σ^E is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol* **181**: 204–211.

Pall, G.S., and Hamilton, A.J. (2008) Improved northern blot method for enhanced detection of small RNA. *Nat Protoc* **3**: 1077–1084.

Pandey, S.P., Minesinger, B.K., Kumar, J., and Walker, G.C. (2011) A highly conserved protein of unknown function in *Sinorhizobium meliloti* affects sRNA regulation similar to Hfq. *Nucleic Acids Res* **39**: 4691–4708.

Pandey, S.P., Winkler, J.A., Li, H., Camacho, D.M., Collins, J.J., and Walker, G.C. (2014) Central role for RNase YbeY in Hfq-dependent and Hfq-independent small-RNA regulation in bacteria. *BMC Genomics* **15**: 121.

Pánek, J., Bobek, J., Mikulík, K., Basler, M., and Vohradský, J. (2008) Biocomputational prediction of small non-coding RNAs in *Streptomyces*. *BMC Genomics* **9**: 217.

Papenfort, K., Förstner, K.U., Cong, J.-P., Sharma, C.M., and Bassler, B.L. (2015) Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *Proc Natl Acad Sci USA* **112**: E766–E775.

Park, J.-S., and Roberts, J.W. (2006) Role of DNA bubble rewinding in enzymatic transcription termination. *Proc Natl Acad Sci USA* **103**: 4870–4875.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**: 79–88.

Pernodet, J., Boccard, F., Alegre, M.-T., Gagnat, J., and Guerineau, M. (1989) Organization and nucleotide sequence analysis of ribosomal RNA gene cluster from *Streptomyces ambofaciens*. *Gene* **79**: 33–46.

Peters, J.M., Mooney, R.A., Grass, J.A., Jessen, E.D., Tran, F., and Landick, R. (2012) Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev* **26**: 2621–2633.

Peters, J.M., Mooney, R.A., Kuan, P.F., Rowland, J.L., Keles, S., and Landick, R. (2009) Rho directs widespread termination of intragenic and stable RNA transcription. *Proc Natl Acad Sci USA* **106**: 15406–15411.

Peters, J.M., Vangeloff, A.D., and Landick, R. (2011) Bacterial transcription terminators: The RNA 3'-end chronicles. *J Mol Biol* **412**: 793–813.

Prévost, K., Salvail, H., Desnoyers, G., Jacques, J.F., Phaneuf, É., and Massé, E. (2007) The small RNA RyhB activates the translation of *shiA* mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Mol Microbiol* **64**: 1260–1273.

Pulido, D., and Jimenez, A. (1987) Optimization of gene expression in *Streptomyces lividans* by a transcription terminator. *Nucleic Acids Res* **15**: 4227–4240.

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

Raghavan, R., Sloan, D.B., and Ochman, H. (2012) Pervasive transcription is widespread but rarely conserved in Enteric bacteria. *mBio* **3**: e00156–12.

Ramos, P.I.P., Picão, R.C., Almeida, L.G.P. De, Lima, N.C.B., Girardello, R., Vivan, A.C.P., *et al.* (2014) Comparative analysis of the complete genome of KPC-2-producing *Klebsiella pneumoniae* Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms. *BMC Genomics* **15**: 54.

Reinhart, A.A., Powell, D.A., Nguyen, A.T., O'Neill, M., Djapgne, L., Wilks, A., *et al.* (2015) The *prrF*-encoded small regulatory RNAs are required for iron homeostasis and virulence of *Pseudomonas aeruginosa*. *Infect Immun* **83**: 863–875.

Richardson, J.P. (2002) Rho-dependent termination and ATPases in transcript termination. *Biochim Biophys Acta* **1577**: 251–260.

Richardson, J.P., Grimley, C., and Lowery, C. (1975) Transcription termination factor rho activity is altered in *Escherichia coli* with *suA* gene mutations. *Proc Natl Acad Sci USA* **72**: 1725–1728.

Richardson, L. V., and Richardson, J.P. (1996) Rho-dependent termination of transcription is governed primarily by the upstream Rho utilization (rut) sequences of a terminator. *J Biol Chem* **271**: 21597–21603.

Roberts, J.W. (1969) Termination factor for RNA synthesis. *Nature* **224**: 1168–1174.

Robertson, G.T., and Roop, R.M. (1999) The *Brucella abortus* host factor I (HF-I) protein contributes to stress resistance during stationary phase and is a major determinant of virulence in mice. *Mol Microbiol* **34**: 690–700.

Romeo, T., Gong, M., Liu, M.Y., and Brun-Zinkernagel, A.M. (1993) Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that afffects glycogen biosynthesis, gluconeogenesis, cell size and surface properties. *J Bacteriol* **175**: 4744–4755.

Ross, A.F., Oleynikov, Y., Kislauskis, E.H., Taneja, K.L., and Singer, R.H. (1997) Characterization of a β-actin mRNA zipcode-binding protein. *Mol Cell Biol* **17**: 2158–2165.

Rouault, T.A., Hentze, M.W., Haile, D.J., Harford, J.B., and Klausner, R.D. (1989) The iron-responsive element binding protein: a method for the affinity purification of a regulatory RNA-binding protein. *Proc Natl Acad Sci USA* **86**: 5768–5772.

Rudolph, M.M., Vockenhuber, M.P., and Suess, B. (2013) Synthetic riboswitches for the conditional control of gene expression in *Streptomyces coelicolor*. *Microbiology* **159**: 1416–1422.

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

Salim, N.N., and Feig, A.L. (2010) An upstream Hfq binding site in the *fhlA* mRNA leader region facilitates the OxyS-*fhlA* interaction. *PLoS One* **5**: e13028.

Sambrook, J., and Russell, D.W. (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbour, N.Y.

Santangelo, T.J., and Artsimovitch, I. (2011) Termination and antitermination: RNA polymerase runs a stop sign. *Nat Rev Microbiol* **9**: 319–329.

Santangelo, T.J., and Roberts, J.W. (2004) Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Mol Cell* **14**: 117–126.

Sauer, E. (2013) Structure and RNA-binding properties of the bacterial LSm protein Hfq. *RNA Biol* **10**: 610–618.

Sauer, E., Schmidt, S., and Weichenrieder, O. (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc Natl Acad Sci USA* **109**: 9396–9401.

Sauer, E., and Weichenrieder, O. (2011) Structural basis for RNA 3'-end recognition by Hfq. *Proc Natl Acad Sci USA* **108**: 13065–13070.

Schmitt-John, T., and Engels, J.W. (1992) Promoter constructions for efficient secretion expression in *Streptomyces lividans*. *Appl Microbiol Biotechnol* **36**: 493–498.

Scholtissek, S., and Grosse, F. (1987) A cloning cartridge of λ to terminator. *Nucleic Acids Res* **15**: 3185.

Seghezzi, N., Amar, P., Koebmann, B., Jensen, P.R., and Virolle, M.J. (2011) The construction of a library of synthetic promoters revealed some specific features of strong *Streptomyces* promoters. *Appl Microbiol Biotechnol* **90**: 615–623.

Sello, J.K., and Buttner, M.J. (2008) The gene encoding RNase III in *Streptomyces coelicolor* is transcribed during exponential phase and is required for antibiotic production and for proper sporulation. *J Bacteriol* **190**: 4079–4083.

Sermonti, G., and Spada-Sermonti, I. (1955) Genetic recombination in *Streptomyces*. *Nature* **176**: 121.

Shen, V., and Bremer, H. (1977) Rate of ribosomal ribonucleic acid chain eleongation in *Escherichia coli* B/r during chloramphenicol treatment. *J Bacteriol* **130**: 1109–1116.

Siegl, T., Tokovenko, B., Myronovskyi, M., and Luzhetskyy, A. (2013) Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. *Metab Eng* **19**: 98–106.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., *et al.* (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* **4**: e1000163.

Skordalakes, E., and Berger, J.M. (2003) Structure of the Rho transcription terminator: Mechanism of mRNA recognition and helicase loading. *Cell* **114**: 135–146.

Skordalakes, E., and Berger, J.M. (2006) Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor. *Cell* **127**: 553–564.

Smaldone, G.T., Antelmann, H., Gaballa, A., and Helmann, J.D. (2012) The FsrA sRNA and FbpB protein mediate the iron-dependent induction of the Bacillus subtilis LutABC iron-sulfur-containing oxidases. *J Bacteriol* **194**: 2586–2593.

Sobrero, P., and Valverde, C. (2012) The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit Rev Microbiol* **38**: 276–299.

Soper, T.J., Doxzen, K., and Woodson, S.A. (2011) Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. *RNA* **17**: 1544–1550.

Soper, T.J., and Woodson, S.A. (2008) The *rpoS* mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA* **14**: 1907–1917.

Srisawat, C., and Engelke, D.R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA* **7**: 632–641.

Stackebrandt, E., and Woese, C.R. (1981) Towards a phylogeny of the actinomycetes and related organisms. *Curr Microbiol* **5**: 197–202.

St-Onge, R.J., Haiser, H.J., Yousef, M.R., Sherwood, E., Tschowri, N., Al-Bassam, M., and Elliot, M.A. (2015) Nucleotide second messenger-mediated regulation of a muralytic enzyme in *Streptomyces*. *Mol Microbiol* **96**: 779–795.

Storz, G., Vogel, J., and Wassarman, K.M. (2011) Regulation by small RNAs in bacteria: Expanding frontiers. *Mol Cell* **43**: 880–891.

Sun, J., Kelemen, G.H., Fernández-Abalos, J.M., and Bibb, M.J. (1999) Green fluorescent protein as a reporter for spatial and temporal gene expression in *Streptomyces coelicolor* A3(2). *Microbiology* **145**: 2221–2227.

Sun, X., Zhulin, I., and Wartell, R.M. (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res* **30**: 3662–3671.

Suzuki, K., Babitzke, P., Kushner, S.R., and Romeo, T. (2006) Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. *Genes Dev* **20**: 2605–2617.

Swiercz, J.P., and Elliot, M.A. (2012) Streptomyces Sporulation. In: Abel-Santos, E, editor. Bacterial Spores: Current Research and Applications. Caister Academic Press.

Swiercz, J.P., Hindra, Bobek, J., Haiser, H.J., Berardo, C. Di, Tjaden, B., and Elliot, M.A. (2008) Small non-coding RNAs in *Streptomyces coelicolor*. *Nucleic Acids Res* **36**: 7240–7251.

Tahiri-Alaoui, A., Frigotto, L., Manville, N., Ibrahim, J., Romby, P., and James, W. (2002) High affinity nucleic acid aptamers for streptavidin incorporated into bi-specific capture ligands. *Nucleic Acids Res* **30**: e45.

Takano, E., White, J., Thompson, C.J., and Bibb, M.J. (1995) Construction of thiostrepton-inducible, high-copy-number expression vectors for use in *Streptomyces* spp. *Gene* **166**: 133–137.

Taverniti, V., Forti, F., Ghisotti, D., and Putzer, H. (2011) *Mycobacterium smegmatis* RNase J is a 5'-3' exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Mol Microbiol* **82**: 1260–1276.

Thomason, M.K., Fontaine, F., Lay, N. De, and Storz, G. (2012) A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol Microbiol* **84**: 17–35.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178–192.

Trotochaud, A.E., and Wassarman, K.M. (2004) 6S RNA function enhances long-term cell survival. *J Bacteriol* **186**: 4978–4985.

Tsui, H.C.T., Leung, H.C.E., and Winkler, M.E. (1994) Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12. *Mol Microbiol* **13**: 35–49.

Tu, K.C., and Bassler, B.L. (2007) Multiple small RNAs act additively to integrate sensory information and control quorum sensing in *Vibrio harveyi*. *Genes Dev* **21**: 221–233.

Unniraman, S., Prakash, R., and Nagaraja, V. (2001) Alternate paradigm for intrinsic transcription termination in Eubacteria. *J Biol Chem* **276**: 41850–41855.

Varani, G. (1995) Exceptionally stable nucleic acid hairpins. *Annu Rev Biophys Biomol Struct* **24**: 379–404.

Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., and Artsimovitch, I. (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**: 157–162.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., and Sinderen, D. van (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71**: 495–548.

Vercruysse, M., Köhrer, C., Davies, B.W., Arnold, M.F.F., Mekalanos, J.J., RajBhandary, U.L., and Walker, G.C. (2014) The highly conserved bacterial RNase YbeY is essential in *Vibrio cholerae* , playing a critical role in virulence, stress regulation, and RNA processing. *PLoS Pathog* **10**: e1004175.

Viegas, S.C., Pfeiffer, V., Sittka, A., Silva, I.J., Vogel, J., and Arraiano, C.M. (2007) Characterization of the role of ribonucleases in *Salmonella* small RNA decay. *Nucleic Acids Res* **35**: 7651–7664.

Vockenhuber, M.-P., Sharma, C.M., Statt, M.G., Schmidt, D., Xu, Z., Dietrich, S., *et al.* (2011) Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*. *RNA Biol* **8**: 468–477.

Vockenhuber, M.P., and Suess, B. (2012) *Streptomyces coelicolor* sRNA scr5239 inhibits agarase expression by direct base pairing to the *dagA* coding region. *Microbiology* **158**: 424–435.

Vogel, J., and Luisi, B.F. (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* **9**: 578–589.

Waksman, S.A. (1961) The Actinomycetes Vol. II Classification, Identification and Descriptions of Genera and Species. Baltimore: The Williams & Wilkins Company.

Walker, S.C., Scott, F.H., Srisawat, C., and Engelke, D.R. (2008) RNA affinity tags for the rapid purification and investigation of RNAs and RNA-protein complexes. *Methods Mol Biol* **488**: 23–40.

Wang, M.C., Chien, H.F., Tsai, Y.L., Liu, M.C., and Liaw, S.J. (2014) The RNA chaperone Hfq is involved in stress tolerance and virulence in uropathogenic *Proteus mirabilis*. *PLoS One* **9**: e85626.

Washburn, R.S., and Gottesman, M.E. (2011) Transcription termination maintains chromosome integrity. *Proc Natl Acad Sci USA* **108**: 792–797.

Washburn, R.S., Marra, A., Bryant, A.P., Rosenberg, M., and Gentry, D.R. (2001) Rho is not essential for viability or virulence in *Staphylococcus aureus*. *Antimicrob Agents Chemother* **45**: 1099–1103.

Waters, L.S., and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell* **136**: 615–628.

Wietz, M., Duncan, K., Patin, N. V., and Jensen, P.R. (2013) Antagonistic interactions mediated by marine bacteria: The role of small molecules. *J Chem Ecol* **39**: 879–891.

Wilson, K.S., and Hippel, P.H. (1994) Stability of *Escherichia coli* transcription complexes near an intrinsic terminator. *J Mol Biol* **244**: 36–51.

Windbichler, N., and Schroeder, R. (2006) Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer. *Nat Protoc* **1**: 637–641.

Yang, K., Han, L., He, J., Wang, L., and Vining, L.C. (2001) A repressor-response regulator gene pair controlling jadomycin B production in *Streptomyces venezuelae* ISP5230. *Gene* **279**: 165–173.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103–119.

Yarnell, W.S., and Roberts, J.W. (1999) Mechanism of intrinsic transcription termination and antitermination. *Science* **284**: 611–615.

Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G., and Gottesman, S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol* **50**: 1111–1124.

Zhang, X., Zhu, Q., Tian, T., Zhao, C., Zang, J., Xue, T., and Sun, B. (2015) Identification of RNAIII-binding proteins in *Staphylococcus aureus* using tethered RNAs and streptavidin aptamers based pull-down assay. *BMC Microbiol* **15**: 102.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

Appendix A. Identification of transcription termination sites

```
#!/usr/bin/perl

#program to identify potential potential terminators (regions with
drastic decreases in transcript levels).

#open coverage data from S. venezualae
open FILE, "./coverage_data/GTE3.6.p.b2.bothstrands.coverage.bed";

#open coverage data file for S. coelicolor
#open FILE, "./coverage_data/GTE1.4.p.b2.bothstrands.coverage.bed";

#for coverage data from S. avermitilis
#open FILE, "./coverage_data/GTE2.5.p.b2.bothstrands.coverage.bed";

#open and clears output files, which will store potential terminators
#for each strand: POS - positive strand, NEG - negative strand.

open TERMINATORPOS,
">./refine/output/sven_terminator_identification_pos.csv";
print TERMINATORPOS "terminators_positive_strand\n";
close TERMINATORPOS;
open TERMINATORPOS,
">>./refine/output/sven_terminator_identification_pos.csv";

open TERMINATORNEG,
">./refine/output/sven_terminator_identification_neg.csv";
print TERMINATORNEG "terminators_negative_strand\n";
close TERMINATORNEG;
open
TERMINATORNEG,">>./refine/output/sven_terminator_identification_neg.csv"
;

#define all variables and set to 0.

$genome_position = 0;
$long_read_pos = 0;
$long_read_neg = 0;

#define 20-element arrays that will store read depth at 20 adjacent
#genome positions. Used for calculating moving averages.

@longreadpos_levels = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0);
@longreadneg_levels = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0);

#define arrays to store genome positions when a
#terminator is found. Used to avoid detecting elements in close
proximity.
```

```
@neg_promoter = ();
@neg_terminator = ();



#compiles the two arrays into a 2-dimensional array. '\' - stores
pointers to the other arrays.

@read_levels = (\@longreadpos_levels, \@longreadneg_levels);



while (<FILE>) {

#matches each line and defines first 3 columns as variables. Column(1):
#genome position; Column (2) + strand, long read lib; Column (3) –
strand, #long read lib;

     $_ =~ m/(.+?)\s(.+?)\s(.+?)\s(.+?)\s(.+?)$/;

      $genome_position = $1;
      $long_read_pos = $2;
      $long_read_neg = $3;

      $genome_position_minnine = $genome_position - 9;
      $genome_position_minten = $genome_position - 10;

#adds the read depth (+ strand , long read lib) as the first value in
#the longreadpos_levels array and removes the last value of the array,
#maintaining the array length at 20.

      unshift @longreadpos_levels, $long_read_pos;
      pop @longreadpos_levels;

#repeats this process for the @longreadneg_levels array. The first
element is #the read depth at the most recent
#genome position; last element is the read depth at genome position -
20.

      unshift @longreadneg_levels, $long_read_neg;
      pop @longreadneg_levels;

#iterate through both arrays of @read_levels array. $i refers to the
array. #For each array, the average of the first 10 elements is
calculated and #compared to the average of the last 10 elements.

for ($i = 0; $i < 2; $i = $i + 1) {

#resets the $first_sum and $second_sum to 0.
```

```
      $first_sum = 0;
      $second_sum = 0;

#then iterate through the different elements of both arrays in
@read_levels
#$j is the jth element of the array.

      for ($j = 0; $j < 10; $j = $j + 1) {


#calculates the sum of the first 10 elements of the $ith array of
#@read_levels.

            $first_sum = $first_sum + $read_levels[$i][$j];

#calculates the sum of the last five elements of the $ith array of
#@read_levels.

            $second_sum = $second_sum + $read_levels[$i][$j+10];

          }

#calculates the average of both sums.

      $first_average = $first_sum/10;
      $second_average = $second_sum/10;

#compares two averages; tests: first average 2X greater than second
average, and first average is between 20 and 4600.
#for S. coelicolor, first average is between 50 and 8200
#for S. avermitilis, first average is between 25 and 4300

      if ($first_average > 2*$second_average && $first_average > 20 &&
      $first_average < 4600) {

#if $i = 1, transcript levels from negative strand -> potential
terminator.

      if ($i == 1) {

#check that terminators aren't identified at neighbouring nucleotides.


      unless ($genome_position - $neg_terminator[-1] < 20) {

      print TERMINATORNEG "negative_terminator,
$genome_position_minnine, $first_average, $second_average\n";

       push @neg_terminator, $genome_position;
```

```
      }

      }

    }

#compares two averages; tests second average 2X greater than first,
#and second average greater than 20, but less than 4600.
#for S. coelicolor, second average is between 50 and 8200
#for S. avermitilis, second average is between 25 and 4300

if ($second_average > 2*$first_average && $second_average > 20
          && $second_average < 4600) {

#if $i = 0, transcript levels from positive strand -> potential
terminator.

      if ($i == 0) {

#check that terminators aren't identified at neighbouring nucleotides.


      unless ($genome_position - $pos_terminator[-1] < 20) {

      print TERMINATORPOS "positive_terminator, $genome_position_minten,
$first_average, $second_average\n";

      push @pos_terminator, $genome_position;

          }

}

}

}

}

close FILE;
close TERMINATORPOS;
close TERMINATORNEG;
```

Appendix B. Characterizing the location of identified terminators

```
#!/usr/bin/perl

#program to determine location of terminators (i.e. within or antisence
to #an annotated gene)

#open file containing all identified terminators in S. venezuelae.

open TERMINATORS, "./refine/output/sven_terminators.csv";

#for S. coelicolor terminators
#open TERMINATORS, "./refine/output/scoel_terminators.csv";
#for S. avermitilis terminators
#open TERMINATORS, "./refine/output/saver_terminators.csv";


#open output file for S. venezuelae
#for S. coelicolor, use
">./refine/output/scoel_terminators_locations.csv"
#for S. avermitilis, use
"./refine/output/saver_terminators_locations.csv"

open OUTFILE, ">./refine/output/sven_terminators_locations.csv";
print OUTFILE "";
close OUTFILE;
open OUTFILE, ">>./refine/output/sven_terminators_locations.csv";

while (<TERMINATORS>) {

#read through terminators file and match each line to pattern. Defines
#4 columns as variables. Column(1): element identifier;
#Column (2) genome position; Column (3) first average; Column (4)
#second average.

$_ =~ m/(.+?),(.+?),(.+?),(.+?)$/;

      $element = $1;
      $genome_position = $2;
      $first_average = $3;
      $second_average = $4;

#open file containing the position and strand of all annotated genes in
S. #venezuelae

#for S. coelicolor, use "./genbank_files/sco_gene_annotation.csv"
#for S. avermitilis, use "./genbank_files/saver_gene_annotation.csv"

open GENE, "./genbank_files/sven_gene_annotation.csv";

while (<GENE>) {

#read through gene annotation file and match each line to pattern.
```

```
#Defines 3 columns as variable. Column (1): gene; Column (2): start;
#Column (3): stop.

$_ =~ m/(.+?),(.+?),(.+?),(.+?)$/;

    $gene = $1;
    $gene_start = $2;
    $gene_stop = $3;
    $strand = $4;

    if ($genome_position > $gene_start && $genome_position < $gene_stop)
{

      print OUTFILE
"$element,$genome_position,$first_average,$second_average,$gene,$gene_st
art,$gene_stop,";

      if (($element =~ /negative/ && $strand =~ /\-/) || ($element =~
/positive/ && $strand =~ /\+/)) {

      print OUTFILE "intragenic\n";

             }

      else {

      print OUTFILE "antisense\n";

      }
}

}
close GENE;

}

close TERMINATORS;
close OUTFILE;
```

Appendix C. Calculating stem composition of selected WebGeSTer terminators

```
#!/usr/bin/perl

#program to look at determine webgester terminator characteristics:
sequence composition of hairpin stem.

#specify input file as argument 1. The input file should contain the
#WebGeSTer output in space delimited format.

open INFILE, "$ARGV[0]";

#specify the output file as argument 2.
open OUTFILE, ">$ARGV[1]";
print OUTFILE "";
close OUTFILE;
open OUTFILE, ">>$ARGV[1]";

#for each line, pull out terminator number, stem length and terminator
#sequence.

while (<INFILE>) {

      $_ =~
      m/\/No=(.+?),\/LP=(.+?),\/US=(.+?),\/B=(.+?),\/DS=(.+?),\/T=(.+?),
\/USL=(.+?),\/DSL=(.+?),\/SL=(.+?),\/BL=(.+?),\/Mm=(.+?),\/Gp=(.+?),\/DG
=(.+?)$/;


      $terminator_number = $1;
      $right_stem_sequence = $3;
      $left_stem_sequence = $5;

#concatenate stem sequence and remove any spaces or brackets.

$stem_sequence = $right_stem_sequence . $left_stem_sequence;

      $stem_sequence =~ s/\)//g;
      $stem_sequence =~ s/\s//g;
      $stem_sequence =~ s/\(//g;

#reset counters for sequence analysis

      $seq = "";
      @sequence = ();

      $i = 0;
      $A = 0;
      $G = 0;
      $U = 0;
      $C = 0;
      $per_A = 0;
      $per_G = 0;
```

```
      $per_U = 0;
      $per_C = 0;
      $per_AU = 0;
      $per_CG = 0;

      $stem_length = length($stem_sequence);

#separate the stem sequence into an array.

      @sequence = split '', $stem_sequence;

for ($i = 0; $i < $stem_length; $i = $i + 1) {

#for each element in @sequence array (ie. each nucleotide in each
sequence) #determine nucleotide identity. Calculate number of each
nucleotide.


                if ($sequence[$i] =~ /a/) {
                  $A = $A + 1;
                }

                if ($sequence[$i] =~ /g/) {
                  $G = $G + 1;
                }

                if ($sequence[$i] =~ /c/) {
                  $C = $C + 1;
                }

                if ($sequence[$i] =~ /u/) {
                  $U = $U + 1;
                }

}

#calculate %content of each nucleotide.

      $per_A = ($A/$i)*100;
      $per_G = ($G/$i)*100;
      $per_U = ($U/$i)*100;
      $per_C = ($C/$i)*100;
        $per_AU = (($A + $U)/$i)*100;
      $per_CG = (($C + $G)/$i)*100;

      print OUTFILE
      "$terminator_number,$per_A,$per_C,$per_G,$per_U,$per_AU,$per_CG\n"
;

}


close INFILE;
close OUTFILE;
```

Appendix D. Calculating tail composition of selected WebGeSTer terminators

```
#!/usr/bin/perl

#program to look at determine webgester terminator characteristics:
sequence composition of hairpin tail.

#specify input file as argument 1. The input file should contain the
#WebGeSTer output in space delimited format.

open INFILE, "$ARGV[0]";

open OUTFILE, ">$ARGV[1]";
print OUTFILE "";
close OUTFILE;
open OUTFILE, ">>$ARGV[1]";

#for each line, pull out terminator number and tail sequence.

while (<INFILE>) {

      $_ =~
      m/\/No=(.+?),\/LP=(.+?),\/US=(.+?),\/B=(.+?),\/DS=(.+?),\/T=(.+?),
\/USL=(.+?),\/DSL=(.+?),\/SL=(.+?),\/BL=(.+?),\/Mm=(.+?),\/Gp=(.+?),\/DG
=(.+?)$/;


      $terminator_number = $1;
      $tail_sequence = $6;

#any spaces or brackets from tail sequence

      $tail_sequence =~ s/\)//g;
      $tail_sequence =~ s/\s//g;
      $tail_sequence =~ s/\(//g;

#reset counters for sequence analysis

      $seq = "";
      @sequence = ();

      $i = 0;
      $A = 0;
      $G = 0;
      $U = 0;
      $C = 0;
      $per_A = 0;
      $per_G = 0;
      $per_U = 0;
      $per_C = 0;
      $per_AU = 0;
      $per_CG = 0;
```

```
      $tail_length = length($tail_sequence);

#separate the tail sequence into an array.


      @sequence = split '', $tail_sequence;

for ($i = 0; $i < $tail_length; $i = $i + 1) {

#for each element in @sequence array (ie. each nucleotide in each
sequence) #determine nucleotide identity. Calculate number of each
nucleotide.


                if ($sequence[$i] =~ /a/) {
                  $A = $A + 1;
                }

                if ($sequence[$i] =~ /g/) {
                  $G = $G + 1;
                }

                if ($sequence[$i] =~ /c/) {
                  $C = $C + 1;
                }

                if ($sequence[$i] =~ /u/) {
                  $U = $U + 1;
                }

}

#calculate %content of each nucleotide.

      $per_A = ($A/$i)*100;
      $per_G = ($G/$i)*100;
      $per_U = ($U/$i)*100;
      $per_C = ($C/$i)*100;
        $per_AU = (($A + $U)/$i)*100;
      $per_CG = (($C + $G)/$i)*100;

      print OUTFILE
      "$terminator_number,$per_A,$per_C,$per_G,$per_U,$per_AU,$per_CG\n"
;



}


close INFILE;
close OUTFILE;
```