Specific and global networks of gene regulation in *Streptomyces coelicolor*

Ayad Hazim Hasan

Submitted in accordance with the requirement for the degree of Doctor of Philosophy

The University of Leeds, Faculty of Biological Sciences

School of Molecular and Cellular Biology

November 2015

Intellectual Property and Publication Statements

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

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgments

It is a pleasure to thank those who made this thesis possible. First of all, I would like to express my sincere appreciation for my honorific supervisor Dr. Kenneth McDowall for his continuous support, advice, patience, generosity and understanding. He always guided me in the right direction; without his suggestions and encouragement, this thesis would never have been written. Simply, I cannot imagine a better supervisor.

My sincere thanks go to the sponsor of this project (KRG-HCDP Scholarship program). I also extend my sincere thanks to the Faculty of Biological Sciences and the Graduate School hardship program for their support during final year of my study.

Special thanks to my colleagues Justin, Louise, David, Tom, Fayez, Jaskiran and Ayat for their help and wishes for the successful completion of the project. Their scientific chats and support inspirited me to solve many problems. I would also like to thank my lab mates, everyone who has been a member of Alex O'Neill's group (especially my shisha mate Zeyad) and Peter Stockley's group during my study.

I would like to acknowledge the following individuals for their assist; Dr. Ryan Seipke for his time and valuable feedback on a preliminary version of this thesis, Dr. Joan Boyes, Dr Bin Hong, Professor Colin Smith, Dr. Iain Manfield and Mr Chris Jones. I am very thankful to Dr. Martin Huscroft and Ph.D. student Sara Narramore for their help with actinorhodin fractionation and giving me the opportunity to use their facilities in Chemistry G61 lab.

I owe a lot to my spouse and partner Shelan and my little princes Awshar for their endless love, encouragement, sacrifice and motivation. I am in greatly in debt to my parents, brothers and sisters for their love and support throughout my life. Without their support, I would not be at this stage.

Above all, I thank Almighty God for giving me the strength, patience and wisdom to overcome every obstacle and difficulty I encountered.

Finally, I would like to show my gratitude to all those who supported me in any respect during the completion of the project.

Abstract

Streptomycetes produce a plethora of secondary metabolites, including antibiotics, and undergo a complex developmental cycle. The GC-rich species, *Streptomyces coelicolor* is used by many laboratories as a model for studying gene regulation, morphological development, cellular physiology and microbial signalling. A genome-wide view of many factors that control *S. coelicolor* gene expression at the level of transcription initiation and beyond was obtained successfully from a combination of RNA-sequencing approaches. For instance, sites of transcription initiation, vegetative promoters, leaderless mRNAs, sites involved in the processing and degradation of rRNA, tRNA and mRNA, and small RNAs, including those that may be involved in attenuation-like switching mechanisms, were successfully detected. Many of the small RNAs identified in this study are novel. Overall, our approaches show the ability to identify new layers of transcriptional complexity associated with several key regulators of secondary metabolism and morphological development in *S. coelicolor*. Here we were able to show that AtrA activates the transcription of *ssgR*, the gene product of which is in turn required for the transcription of *ssgA*, the best-studied SALP, which has a crucial role in septation and the morphology aerial hyphae. AtrA also binds to the promoter region of *leuA2*, which encodes α-isopropylmalate that directly utilise acetyl-CoA. Interestingly, crude extract from M1146 strain (*Δact*, *Δred*, *Δcpk*, and *Δcda*) was found to inhibit the DNA-binding activity of AtrA; however, the specificity of the small molecule(s) interaction with AtrA should be investigated. Addition of the 3 x FLAG tag™ to the N-terminus of AtrA does not hinder its ability to substitute functionally for untagged AtrA in *S. coelicolor* and can be used for the mapping of AtrA binding sites by ChIP-sequencing. Taken together the above suggest that AtrA has a direct role in morphological development and coordinating the utilisation of acetyl-CoA for primary and secondary metabolism.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACT	actinorhodin
Amp	ampicillin
APS	ammonium persulfate
bp	base pair
BSA	bovine serum albumin
CDA	calcium-dependent antibiotic
cDNA	complementary deoxyribonucleic acid
°C	degree Celsius
CFU	colony forming unit
ChIP	chromatin immunoprecipitation
CPK	cryptic polyketide antibiotic
CSRs	cluster-situated regulators
CT	threshold cycle
DACA	DNA affinity-capture assays
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
dRNA-seq	differential RNA-sequencing
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
EtOAc	Ethyl acetate
FAM	fluorescein amidite
FRT-seq	flowcell reverse transcription sequencing
g	gram(s)
g	gravity
gRNA	global RNA-sequencing
h	hour(s)
HPLC	high pressure liquid chromatography
IMAC	immobilized metal-affinity chromatography
IPTG	isopropyl-βs-D-thiogalactopyranoside
kbp	kilo base pair

kDa	kilodalton
kg	kilogram (10^3 g)
LB	Luria Bertani medium
m / z	M stands for mass and Z stands for charge number of ions
mg	milligram (10^{-3} g)
µg	microgram (10^{-6} g)
min	minute(s)
mL	millilitre(s) (10^{-3} L)
µL	microlitre(s) (10^{-6} L)
mm	millimetre
mM	millimolar
µM	micromolar
MSF	Mannitol soya flour medium
NCBI	National Centre for biotechnology Information
ng	nanogram (10^{-9} g)
nM	nanomolar
nt	nucleotide
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pg	pictogram(s)
psi	pounds per square inch
PWM	Position weight matrix
QPCR	quantitative polymerase chain reaction
RBS	ribosome binding site
RED	undecylprodigiosin antibiotic
R_f	Retention Value
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RT	reverse transcription
Rt	Retention time
s	second(s)

S	Svedberg unit
SDS	sodium dodecyl sulphate
SELEX	selective enrichment of ligands by exponential enrichment
SMs	secondary metabolites
sRNA	small ribonucleic acid
TAP	Tobacco acid pyrophosphatase
TBE	Tris-borate-EDTA
TE	Tris-EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
TES	N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid
TEX	Terminator™ 5’ phosphate-dependent exonuclease
TGED	Tris-glycine-EDTA
TGEK	Tris-glycine-EDTA-potassium chloride
tmRNA	transfer-messenger ribonucleic acid
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
TSA	tryptone soya agar medium
TSB	tryptone soya broth medium
tsr	thiostrepton
TSS	transcriptional start site
UCSC	University of California, Santa Cruz
UTR	untranslated region
UV	ultra violet
V	volts
v/v	volume (ml) per volume (ml)
w/v	weight (g) per volume (ml)

Chapter 1

1 Introduction

1.1 *Streptomyces*

Streptomycetes are thought to have first evolved about 450 million years ago, eventually giving rise to many *Streptomyces* species, of which 576 have been described to date (Embley & Stackebrandt, 1994, Labeda, 2011). *Streptomyces* are Gram-positive, aerobic, soil-dwelling and high GC content (72-73 %) bacteria that belong to the family Streptomycetaceae within the Actinobacteria phylum (Waksman & Henrici, 1943, Ventura *et al.*, 2007). *Streptomyces coelicolor* A3(2) is one of the best-studied species of *Streptomyces*, and undergoes a complex lifecycle (Hopwood, 1999). Like other members of its genus, *S. coeliocolor* grows from a spore, which germinates in the presence of suitable nutrients and favourable growth conditions, and grows to form a branching vegetative mycelium. When nutrients are limited, the mycelium produces aerial hyphae that undergo septation to produce spores (Nodwell *et al.*, 1999, Flardh & Buttner, 2009, Bibb, 2005). This morphological differentiation coincides with the production of secondary metabolites (Figure 1.1). The programming of *Streptomyces* morphological differentiation can be broken down into two distict phases, which are aerial hyphae formation and spore maturation, and are regulated by *bld* genes (for bald) and *whi* genes (for white), respectively (Chater, 1993).

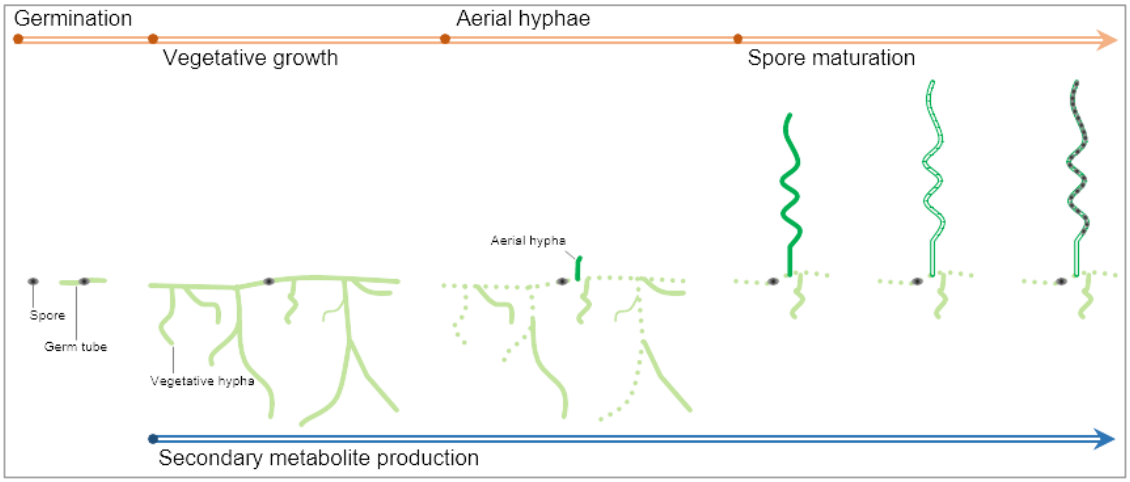

Figure 1.1: Schematic view of *S. coelicolor* life cycle. Substrate mycelium grows down into the medium by tip extension after spore germination. Depletion of nutrients leads to lysis of vegetative mycelium (dot lines), which in turn stimulate erection of aerial hyphae. When extension of these aerial hyphae is complete, the apical cells septate to form mature spores.

1.1.1 Aerial hyphyae formation regulatory networks

The signals that regulate aerial hyphae formation and the underlying mechanism have been studied extensively in *S. coelicolor*. Most of these signals are encoded by *bld* genes, mutations in which can block the production of the fuzzy colony morphololgy of the wild-type (Claessen *et al.*, 2006, Willey *et al.*, 2006, Willey *et al.*, 1993). The expression of *chp* and *ram* genes play important roles in hyphal erection and are dependent on Bld activities. In *S. coelicolor*, *ram* encodes and exports a small surface-active peptide called SapB (Spore associated protein B) (Kodani *et al.*, 2004). Mature SapB is a peptide of 21 amino acids encoded by *ramS*, which is a member of *ram* cluster that includes *ramA*, *ramB*, *ramC*. The *ramABC* genes encode components of an ABC transporter involved in SapB export. The *ram* cluster is positively regulated by another gene member, *ramR* (Willey *et al.*, 2006, Kodani *et al.*, 2004). The *chp* genes encodes chaplins (coelicolor hydrophobic aerial proteins), which are surface active proteins that share a conserved hydrophobic domain called the chaplin domain and a Sec secretion signal at their N-terminus (Elliot *et al.*, 2003, Claessen *et al.*, 2003). Collectively, SapB and the chaplins reduce surface tension between the aqueous milieu

of the vegetative mycelium and the air, thereby allowing aerial hyphae to emerge from the vegetative mass and grow into air (Figure 1.2) (Elliot *et al.*, 2003, Claessen *et al.*, 2003).

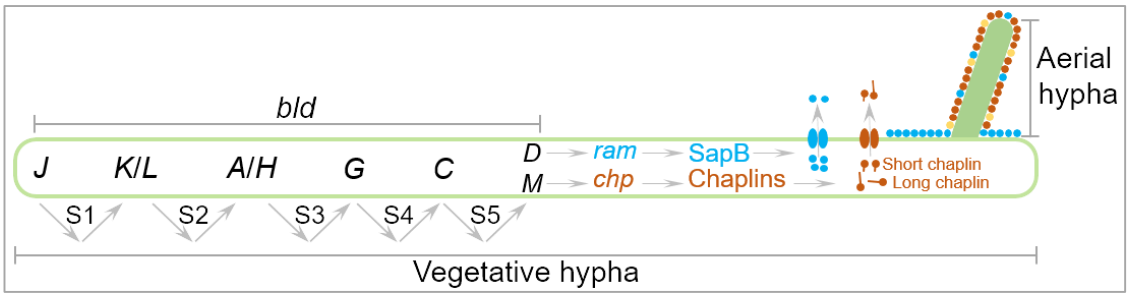

Figure 1.2: General view of extracellular signals of the *bld* cascade that govern aerial hyphae formation in *S. coelicolor.* 'S' with numbers that directed by arrow indicate the sequence of extracellular signals produced by *bld* genes, which stimulate *ram* and *chp* production. Short chaplins (ChpD-H, small brown circle with short tail) have single chaplin domain, which is a region present in each member of chaplin family, while long chaplins (ChpA-C, small brown circle with short tail) have two chaplin domain (Elliot *et al.*, 2003). SapB and rodlin are represented as small blue and yellow circles, respectively. Rodlins are not required for aerial hyphae formation, but they are necessary for organising the chaplains into rodlet layer in the surface of aerial hyphae (Claessen *et al.*, 2004). Chaplins and SapBs are exported from the cell by the Sec (secretory) system (Flardh & Buttner, 2009) and the ATP-binding cassette transpoter encoded by *ramAB* (Kodani *et al.*, 2004), respectively.

1.1.2 Sopre maturation regulatory networks

Sporulation in *S. coelicolor* is affected by many genes including the *whi* genes (A, B, D, E, G, H and I), *ftsZ*, *sigF, ssgA* and SLAP genes (Flardh & Buttner, 2009). Mutants of *whi* can be classified into two groups, early (*whiA*, *B*, *G*, *H* and *I*) and late (*whiD*, *E*) genes depending on the effect of mutation on septation and spore maturation. These genes are required for septation and transcription of the late sporulation genes, i.e. *ftsZ*, *sigF*, *whiD* and the *whiE* cluster, which are necessary for spore maturation and pigmentation (see Figure 1.3) (Chater, 2006, Flardh *et al.*, 1999). The *whiG* gene encodes the sporulation-specific σ^{WhiG} factor that directly activates the transcription of *whiH* and *whiI* (Ainsa *et al.*, 1999, Ryding *et al.*, 1998). The FixJ subfamily member WhiH responds to a change in small organic acids concentration during aerial hyphal growth,

while The GntR family member WhiI induces DNA condensation during sporulation process. The genes *whiI* and *whiH* appear to act in an autoregulatory manner, and *whiH* is negatively controlled by *whiI* (Ainsa *et al.*, 1999, Ryding *et al.*, 1998).

On the other hand, *whiA* and *whiB* genes are transcribed by RNA polymerase associated with the *whiG* sigma factor (Ainsa *et al.*, 2000, Soliveri *et al.*, 1992). WhiB upregulates expression of *whiA* and acts as an autorepressor, while WhiA represses the transcription of *whiB* and is an autoactivator that positively regulates its own transcription (Jakimowicz *et al.*, 2006). WhiA induces the expression of *ftsZ*p2 promoter, which encodes a tubulin-like protein that assembles into a cytokinetic ring (Z ring) to defines the site of cell division for sporulation septation (Mistry *et al.*, 2008, Schwedock *et al.*, 1997, Flardh *et al.*, 2000). Another target for WhiA regulation is *parAB* promoter, which encode ParA and ParB that required for chromosome segregation and sporulation (Kaiser & Stoddard, 2011). Expression of the *whiE* cluster depends on *sigF* (Kelemen *et al.*, 1996), which is required for synthesis of the grey polyketide spore pigment (Davis & Chater, 1990). The only interaction between *bld* and *whi* genes occurs via BldD, which represses sigma WhiG and *whiB* prior to sporulation (Jakimowicz *et al.*, 2006, Elliot *et al.*, 2001).

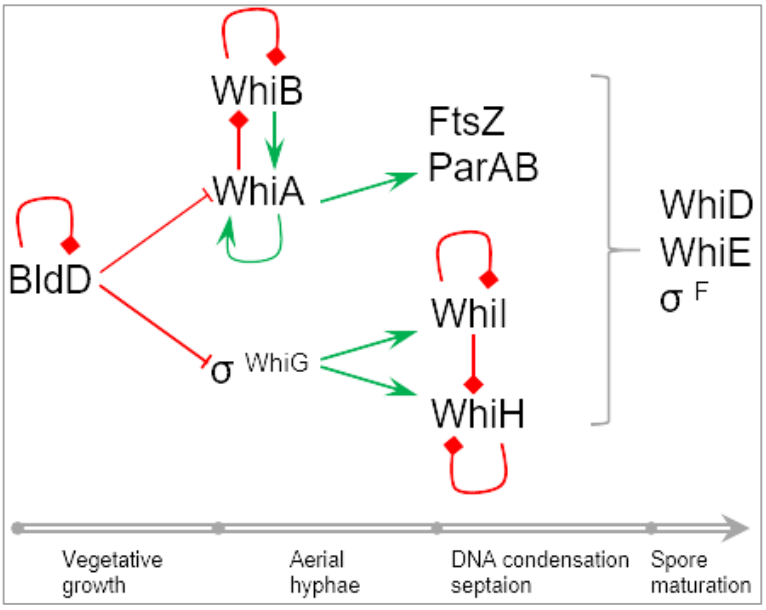

Figure 1.3: Schematic summary of the regulatory network of sporulation in *S. coelicolor*. Stealth arrows indicate positive effect, diamond arrows indicate negative effect. Information from (McCormick & Flardh, 2012, Kaiser & Stoddard, 2011, Chater & Chandra, 2006).

1.2 Antibiotic production in *Streptomyces*

Mycelial bacteria of the genus *Streptomyces* and their relatives are renowned as the premier source of the majority of antibiotics used clinically, in addition to many other therapeutics such as antihelmintics and anticancer agents (Baltz, 2008, Caffrey *et al.*, 2008, Graziani, 2009, Olano *et al.*, 2009, Shiomi & Omura, 2004). Furthermore, such substances are only a proportion of the vast array of secondary metabolites (SMs) produced by this group (Challis, 2008, Davies, 2011). Most, if not all, *Streptomyces* are capable of producing numerous SMs (Corre & Challis, 2009, Gross, 2009, Nett *et al.*, 2009) and the timing of the production of SMs and the quantities generated are exquisitely sensitive to growth and environmental conditions (van Wezel & McDowall, 2011). SMs are typically, although not invariably, produced when the period of most rapid growth gives way to slower growth and eventually the production of aerial hyphae and spores.

SM production has been associated with the depletion of sources of nitrogen, carbon or phosphate, and with the accumulation of small signalling molecules such as the stringent factor (p)ppGpp (Bibb, 2005, Chatterji & Ojha, 2001) and a family of γ-butyrolactone microbial hormones (Hsiao *et al.*, 2009, Horinouchi & Beppu, 2007, Takano, 2006). Where studied, it appears largely to reflect changes in the level of transcription of activators that are found with the corresponding biosynthetic genes as part of a cluster (Bibb, 2005, van Wezel & McDowall, 2011). It has also been proposed that certain cluster-situated regulators (CSRs) are controlled at the post-transcriptional level by the availability of the BldA leucyl-tRNA, which recognizes the rare UUA codon found in the genes of these regulators and some genes required for morphological development (Chater & Chandra, 2008, Chater, 2006, Chater, 2001). Examples of genes containing the TTA codon in *S. coelicolor* include *act*II-ORF4, *redZ*, *adpA* (*bldH*) and *actA* (Fernandez-Moreno *et al.*, 1991, White & Bibb, 1997, Nguyen *et al.*, 2003).

As a means of establishing the regulatory pathways that control antibiotic production by the *Streptomyces* genus, the model species *S. coelicolor* and its close relative *S. lividans* have been screened numerous times to identify mutants (> 50) that regulate the production of antibiotics (Bibb, 2005, van Wezel & McDowall, 2011). Indeed the importance of *S. coelicolor* as a model species stems largely from two of its main SMs being pigmented; actinorhodin (ACT) is blue, while undecylprodigiosin (RED) is red. This pigmentation makes the detection of mutants that affect secondary metabolism relatively easy. Despite these endeavours, only one regulatory pathway has been described for *S. coelicolor* that extends without gaps from a signalling molecule all the way through to the transcriptional activation of biosynthetic genes. Most of the currently identified regulators of secondary metabolism can be viewed as isolated pieces of a complex jigsaw, with few links to each other and the signals they transduce. In addition to ACT and RED, *S. coelicolor* produces undecylprodigiosin, calcium dependent antibiotics, cryptic polyketide and methylenomycins.

1.2.1 Undecylprodigiosin (RED)

Undecylprodigiosin, also known as Red, is a member of prodiginine family, which comprises red-pigmented tripyrrole antibiotics possessing anti-cancer, immunosuppressant and antimicrobial properties (Feitelson & Hopwood, 1983, White & Bibb, 1997, Williamson *et al.*, 2006). This highly nonpolar tripyrrole compound is chemically similar to prodigiosin produced by *Serratia marcescens*, and has ability to inhibit the proliferation of human T and B lymphocytes without damaging the cells (Williamson *et al.*, 2006, Songia *et al.*, 1997). The orphan response regulator RedZ activates transcription of *redD*, the final regulator of the RED biosynthetic genes cluster (White & Bibb, 1997, van Wezel & McDowall, 2011). The control of *redZ* shares some features with that of the *act*II-ORF4 gene, with the promoter in both cases subject to repression by AbsA2 and DasR (Uguru *et al.*, 2005), which suggests that the *redZ* promoter may respond to the level of GlcNAc and regulates production of antibiotic (Rigali *et al.*, 2008). The binding of RedZ to the *redD* promoter is inhibited by undecylprodigiosin *in vitro*, suggesting the existence of a feedback loop for modulating Red biosynthesis (Wang *et al.*, 2009).

1.2.2 Calcium Dependent Antibiotics (CDAs)

CDA is an acidic lipopeptide belonging to a family of anionic undecapeptides that includes daptomycin produced by *Streptomyces roseosporus*, and friulimicins produced by *Actinoplanes fruliensis* (Debono *et al.*, 1988, Vertesy *et al.*, 2000). CDAs are effective in the presence of calcium ions against a wide range of Gram-positive bacteria. They disrupt cell membrane function with effects similar to those of the cationic antimicrobial peptides, such as daptomycin (Kempter *et al.*, 1997, Lakey *et al.*, 1983, Jung *et al.*, 2004). The *cda* gene cluster (82 kb) contains forty genes, of which twenty involved in the biosynthesis of seven CDAs with different functional group and the same peptide sequence (Hojati *et al.*, 2002). The CDA cluster is regulated by three genes, which are *absA1*, *absA2* and *cadaR*. Tailoring enzymes *asnO*, *hasp* and *trpO* modify each variant after been released from *cdaPS3* and cyclised (Hojati *et al.*, 2002).

1.2.3 Cryptic Polyketide (CPK)

Both yellow-cryptic polyketide (yCPK) and polyketide-derived antibiotic (CPK) are final products of *cpk* cluster (Liu *et al.*, 2013). The extracellular molecules, γ-butyrolactone, SCB1 and SCBs have been shown to regulate the *cpk* cluster (Takano *et al.*, 2005a). The *cpkO* gene, a pathway-specific transcriptional activator essential for CPK biosynthesis, is regulated by ScbR, which also binds to operator sites in the bidirectional *scbA/scbR* promoter to activate SCB biosynthetic gene *scbA* (Takano *et al.*, 2005a, Wietzorrek & Bibb, 1997). SCBs binds to ScbR and prevent ScbR from interaction with *scbA*/*scbR* promoter (Gottelt *et al.*, 2010). As a result of ScbA and ScbR production, a heterodimer of the two proteins form, which bind to a different site upstream of *scbA* to activate its transcription. Subsequently, the produced SCBs bind to all free ScbR and activate *cpkO* transcription (Gottelt *et al.*, 2010). It is worth mentioning that *cpkO* is directly repressed by phosphate response regulator (RR) PhoP and AfsQ1 (Wang *et al.*, 2013, Allenby *et al.*, 2012).

1.2.4 Methylenomycins (Mm)

The linear plasmid SCP1 in *S. coelicolor* A3(2) harbours the methylenomycin gene cluster (*mmy* cluster), which includes 21 genes that involve in regulation (*mmfR, L, H, P-mmyR* and *mmyB*), self resistance/export (*mmr*) and biosynthesis of methylenomycin (Kirby & Hopwood, 1977, Bentley *et al.*, 2004, O'Rourke *et al.*, 2009). Methylenomycin A is a cyclopentanone antibiotic that has an activity against Gram-positive bacteria and some Gram-negative strains (Haneishi *et al.*, 1974). Methylenomycin C (desepoxy-4,5-didehydromethylenomycin A) has been shown to be the Mm precursor, while methylenomycin B demonstrated to result from the spontaneous decarboxylation of Mm C under acidic conditions (Challis & Chater, 2001, Corre & Challis, 2005).

1.2.5 Actinorhodin biosynthesis overview in *S. coelicolor*

ACT is an aromatic polyketide antibiotic that is red at acidic pH (below 8.5) and blue at basic pH (above 8.5) (Wright & Hopwood, 1976). It is found intracellularly, but a lactone form called γ-actinorhodin is excreted (Bystrykh *et al.*, 1996). The biosynthesis of ACT is well characterised, but not fully understood (Figure 1.4) (Carreras & Khosla, 1998, Craney *et al.*, 2013). It starts with the condensation of one acetyl-coenzyme A unit and seven malonyl-CoA extender units (Fernandez-Moreno *et al.*, 1992) by the minimal polyketide synthase (PKS) (Carreras & Khosla, 1998, Craney *et al.*, 2013), which is composed of a ketosynthase (KS) with two subunits (α and β) and an acyl carrier protein (ACP) (Matharu *et al.*, 1998). The genes corresponding to the minimal PKS are located within the *act*I region of the cluster (Craney *et al.*, 2013). The octaketide is reduced at the C-9 position by a ketoreducase (KR) encoded by a single gene within the *act*III region. This reduction, which converts a keto to a hydroxyl group, primer the octaketide for the first of three ring cyclisations, which is catalysed by an aromatase (ARO) that produces a hydroxyl group at C-11 as well as removing the keto group at C-7 (McDaniel *et al.*, 1994, Javidpour *et al.*, 2013).

The second cyclisation is catalysed by a cyclase (CYC) that removes the keto at C-5. The ARO and CYC enzymes are encoded by single genes within the *act*VII and *act*IV regions, respectively. The resulting bicyclic intermediate now released from the ACP is then reduced at C-3 by a dehydrogenase, the third ring is produced and the hemiketal dehydrated to form (S)-DNPA (Fernandez-Moreno *et al.*, 1994). At one point it was thought that the third ring closure is catalysed by the product of *act*VI-ORF3, but a more recent study implicates the produce of *act*VI-ORF1 (RED1) (Taguchi *et al.*, 2000). The product of *act*VI-ORF1 has also been implicated in the C-3 reduction and the dehydration step (Taguchi *et al.,* 2000). Clearly much of the enzymology of the steps immediately prior to (S)-DNPA remains to be resolved. The product of *act*VI-ORF2 with the assistance of the *act*VI-ORF4 product, both of which resemble reductases, are thought following ring closure to reduce the double bond between C-14 and C-15 of

(S)-DNPA to produce DDHK (5-deoxy-dihydrokalafungin) (Fernandez-Moreno *et al.,* 1994, Taguchi *et al.,* 2000).

DDHK is oxidised at C-6 to form dihydrokalafungin (DHK) by the product of *act*VA-ORF5, a flavin-dependent monooxygenase (FMO), which forms a two subunit system with the product of *act*VB, a flavin: NADH oxidoreductase (Taguchi *et al.*, 2011, Okamoto *et al.*, 2009). It should be noted that earlier studies had tentatively assigned other gene products to this role. The tricyclic intermediate is dimerised at C-10 in the next step by the product of *act*VA-ORF4 and then further oxidised at C-8 by the two subunit; ActVA-ORF5/ActVB system or alternatively oxidised and then dimerised (Okamoto *et al.,* 2009, Taguchi *et al.*, 2013, Taguchi *et al.*, 2012). The product of either of these final routes is ACT (Figure 1.4). Actinorhodin is then exported by the products of *act*II-ORF2/3 (now called *act*AB) operon, which is regulated by the product of *act*II-ORF1 (*actR*) (Tahlan *et al.*, 2007). Both ACT and its 3-ring intermediates, e.g. (S)-DNPA, relieve repression by ActR (Xu *et al.*, 2012). The cluster-situated activator of the transcription of *act* biosynthesis is the product of *act*II-ORF4 (Rudd & Hopwood, 1979, Hallam *et al.*, 1988, Fujii *et al.*, 1996, Gramajo *et al.*, 1993). Several genes within the *act* cluster remain to be assigned clear functions. These included *act*VA-ORF1, *act*VA-ORF2 and *act*VA-ORF3. ActVA-ORF1 resembles ActII-ORF2, which is involved in ACT export (Bystrykh *et al.,* 1996). However, disruption of *act*II-ORF2 did not affect the ability to export ACT (Fernandez-Moreno *et al.*, 1991), which suggests that ActVA-ORF1 engages in ACT export either with *act*II-ORF2 assistance or as a separate channel (Caballero *et al.*, 1991).

As outlined above, the gene products of the *act* cluster produce one molecule of ACT from fourteen molecules of malonyl-CoA and two molecules of acetyl-CoA. The supply of these carboxylic acid precursors through the catabolism of carbon substrates and their utilisation by competing pathways has an impact on the yield of ACT (Ryu *et al.*, 2006). Undecylprodigiosin, the other pigmented antibiotic produced by *S. coelicolor*, needs six molecules of malonyl-CoA to produce a 2-undecylpyrrole intermediate (Mo *et al.*, 2008). All members of the polyketide family of antibiotics, which includes

tetracyclines, anthracyclines, angucyclines, macrolides, polyenes, ansamycins, polyenes, tetranoates and isochromanequinones, are synthesised from small carboxylic acids in a manner which has much in common with the biosynthesis of fatty acids (Robinson, 1991, Zotchev, 2012). The similarities between polyketide and fatty acid biosynthesis, in particular the modular nature of the synthases that produce the carbon backbone has been the subject of several excellent reviews and will not be described herein (Staunton & Weissman, 2001, Donadio *et al.*, 2007, Hertweck *et al.*, 2007, Kwan & Schulz, 2011). Small carboxylic acids are also required for the production of peptide β-lactams antibiotics. However, small carboxylic acids are not required for the production of all antibiotics; aminoglycosides being an example.

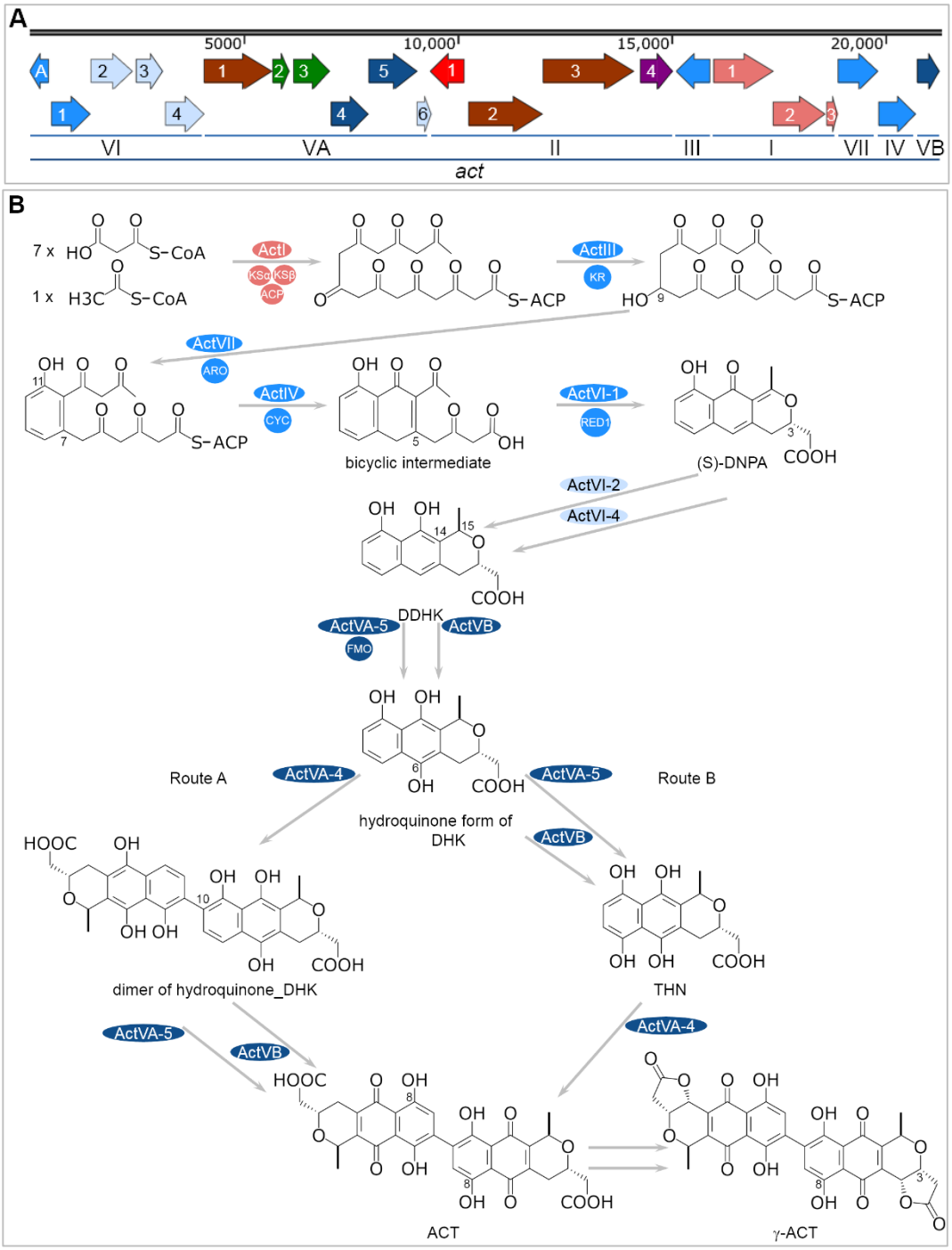

Figure 1.4: Actinorhodin biosynthesis overview. (A) 21283 bp ACT cluster is colour coded. *act*II-ORF4 activator gene is in purple and *actR* repressor in red. *act*I genes are in pink. Tailoring genes are coloured depending on their role in forming ACT (light, medium and dark blue). Putative resistance genes are in brown. Uncharacterised genes are in green. **(B)** ACT biosynthetic pathway, which adapted from (Taguchi *et al.,* 2013). KSα, beta-ketoacyl synthase alpha subunit; KSβ, beta-ketoacyl synthase beta subunit; ACP, acyl carrier protein; KS, ketosynthase; KR, ketoreductase; ARO, aromatase; CYC, cyclase.

The route from glucose, as an example, through substrate catabolism to the production of acetyl- and malonyl-CoA is shown in Figure 1.5. The route goes via the Embden-Meyerhof, Entner-Doudoroff, and pentose phosphate pathways and is subject to regulation by key enzymes in these individual pathways. Moreover, it has been shown that manipulation of central carbohydrate metabolism can increase the production of ACT in *S. coelicolor*. For example, overproduction of acetyl-CoA carboxylase (ACCase) increased the efficiency of the conversion of glucose to ACT (Ryu *et al.*, 2006). Culture conditions, in particular media formulations and available oxygen can alter flux through the pathways of central metabolism and consequently the level of precursors available for secondary metabolism (Olano *et al.*, 2008).

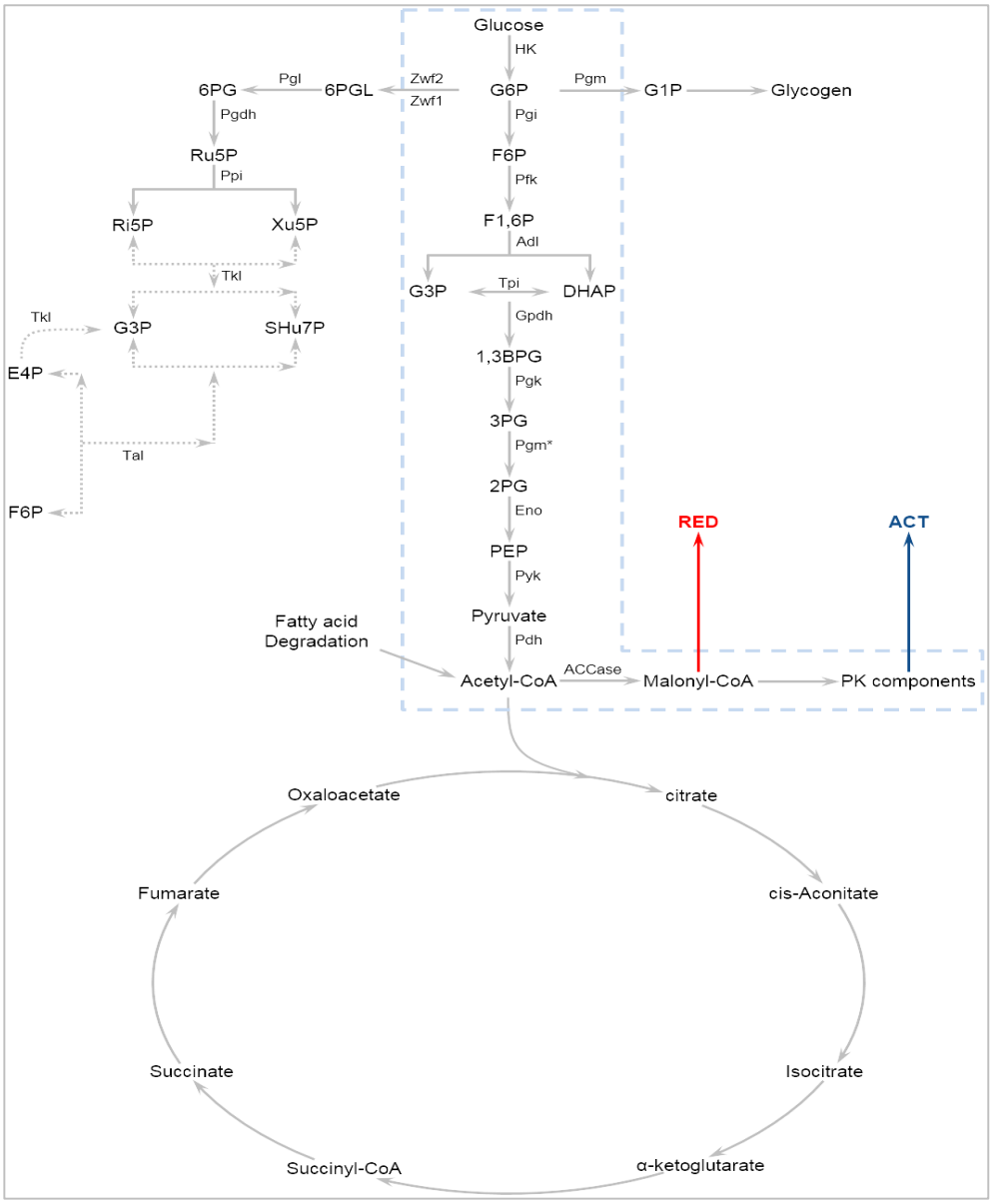

Figure 1.5: The sequence of enzymatic reactions in the glucose metabolism for Act production in *S. coelicolor.* The abbreviations used are as follows: G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; F1,6DP, fructose-1,6-diphosphate; GAP, glyceraldehyde-3-phosphate; 1,3 BPG, 1,3-diphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; 6PGL, 6-phosphoglucolactone; 6PG, 6-phosphogluconate; Ru5P, ribulose-5-phosphate; Ri5P, ribose-5-phosphate; Xu5P, xylulose-5-phosphate; SHu7P, sedoheptulose-7-phosphate; E4P, erythrose-4-phosphate; HK, hexokinase; Pfk, phosphofructokinase; Adl, aldolase; Tpi, triose phosphate isomerase; GpdH, glyceraldehyde-3-phosphate dehydrogenase; Pgk, phosphoglycerate kinase; Pgm*, phosphoglycerate mutase; Eno, enolase; Pyk, Pyruvate kinase; Pdh, pyruvate dehydrogenase; ACCase, acetyl-CoA carboxylase; Zwf, glucose-6-phosphate dehydrogenase; Pgm, phosphoglucomutase; Pgl, phosphoglucolactonase; PgdH, phosphogluconate dehydrogenase; Ppi,

phosphopentose isomerase; Tkl, transketolase; Tal, transaldolase. Taken from (Ryu *et al.,* 2006).

1.3 ActII-ORF4, its regulon and regulation

The nomenclature of regions within the *act* cluster stems from the classification of mutants blocked in ACT biosynthesis on the basis of co-synthesis experiments. Mutants in the same class were unable to produce ACT using intermediates produced by other mutants of the class, while mutants of different classes could produce ACT using intermediates produced by mutants of another class with the exception of class II, which were unable to produce ACT using intermediates that accumulated in any of the other classes (Rudd & Hopwood, 1979). It was presumed, correctly it transpires, that these mutants were deficient in a positive-regulatory factor. This notion gained further credence with the finding that a fragment of the *act* cluster when inserted into a multi-copy-number plasmid was not only able to restore, but increased ACT production by Class II mutants (Hopwood, 1986). Mutational analysis of the region of the *act* cluster that complements Class II mutants, guided by prior sequencing, identified *act*II-ORF4 as the gene encoding the positive regulator of ACT biosynthesis (Fernandez-Moreno *et al.*, 1991). An analysis of transcript levels revealed the mRNA of *act*II-ORF4 increased as growth slowed, as did the mRNA of *act*III and *act*VI-ORF1, but their increase was delayed relative to *act*II-ORF4 (Gramajo *et al.,* 1993). The *act*II-ORF4 gene contains the rare UUA codons recognised by BldA leucyl-tRNA (Fernandez-Moreno *et al.*, 1991).

The product of *act*II-ORF4 belongs to a family of *Streptomyces* antibiotic regulatory proteins (SARPs), members of which are homologous to each other and share similarities with members of a larger family based on OmpR (Wietzorrek & Bibb, 1997), which is a well-characterised transcription factor that is part of a two-component system (TCS) that regulates the levels of the outer membrane porin F (OmpF) (Rampersaud *et al.*, 1994, Batchelor *et al.*, 2005). Many mutations that affect the production of ACT have been shown to affect the level of *act*II-ORF4 mRNA (Aceti & Champness, 1998). The simplest interpretation of these results is that the effects of

these mutations on ACT production are a consequence of direct or indirect effects on the transcription of *act*II-ORF4. Biochemical characterisation of ActII-ORF4 revealed that it binds to the intergenic regions of *act*VI-ORFA/ORF1 and *act*III/*act*I, more specifically to two or three direct repeats that overlap the -35 regions of the corresponding promoters (Arias *et al.*, 1999). These repeats and their location are characteristic of the DNA-binding sites of members of the OmpR family (Wietzorrek & Bibb, 1997). Subsequent bioinformatic analysis (McDowall, unpubl. data) of a type described below revealed similar direct repeats upstream of *act*VA-ORF3 and *accA2* (SCO4921). The latter of which encodes one of two subunits of an acyl-CoA carboxylase (converts acetyl- to malonyl-CoA) (Rodriguez *et al.*, 2001), that suggest an intriguing link to primary metabolism.

The finding that mRNA levels of *act*II-ORF4 are growth dependent and seem to correspond with the production of ACT raised the possibility that the promoter of *act*II-ORF4 was regulated. A biochemical screen of a *S. coelicolor* extract identified an activity that could bind the promoter region of *act*II-ORF4 *in vitro* (Uguru *et al.*, 2005). Through purification of the activity it was identified as being conferred by AtrA, a member of TetR family of transcriptional regulators (TFRs) (Uguru *et al.,* 2005). TFRs are one-component systems that dominate signal transduction in bacteria, archaea and streptomycetes (Cuthbertson & Nodwell, 2013, Ahn *et al.*, 2012, Ramos *et al.*, 2005). In *S. coelicolor* 153 TFRs have been found, which makes 15.8% of the regulatory proteins in this bacterium (Cuthbertson & Nodwell, 2013). Homologues of AtrA are found in all streptomycetes (Ahn *et al.*, 2012). Of these, *S. griseus*, which binds to the *strR* promoter and moderate streptomycin production (Hirano *et al.*, 2008), *Streptomyces roseosporus*, which binds to *dptE* promoter to positively regulate daptomycin biosynthetic cluster (Mao *et al.*, 2015), *Streptomyces avermitilis*, which negatively regulate avermectin biosynthesis (Chen *et al.*, 2008), and *S. globisporus*, which binds to *sgcR1* and activates lidamycin biosynthesis. It is worth highlighting that the latter binding activity can be inhibited by heptaene, a biosynthetic intermediate of lidamycin in *S. globisporus* and actinorhodin from *S. coelicolor* (Li *et al.*, 2015). Disruption of the corresponding *atrA* gene (SCO4118), which is not associated with any

SM gene cluster (Uguru *et al.,* 2005), reduced the production of ACT, but had no detectable effect on the production of RED, indicating that *atrA* has specificity with regard to the biosynthetic genes it influences (Uguru *et al.,* 2005). Within the streptomycetes, *atrA* (SCO4118) is associated with a divergent gene (SCO4119) encoding NADH dehydrogenase (Ahn *et al.,* 2012), a member of the oxidoreductases. The role of this enzyme in maintaining the redox balance is maybe what links it to ACT. ACT is an oxidant, it is redox active, that when production or added exogenously induces the SoxR regulon, which in *S. coelicolor* encodes enzymes that provide protection against redox-sensitive molecules (Shin *et al.*, 2011).

There are two binding sites of AtrA in the *act*II-ORF4 promoter region, centred at −162 and +86 bp relative to the transcriptional start site (Gramajo *et al.,* 1993). This places them within the coding regions of *act*II-ORF3 and *act*II-ORF4, respectively. Binding to the -162 site has been confirmed *in vivo*, although the identity of the transcription factor was not confirmed (McArthur & Bibb, 2008). Since the identification of AtrA as a regulator of ACT production, several other transcription factors have been proposed to regulate the promoter of *act*II-ORF4. In most case, it was been shown that mutation of the corresponding gene affects the production of ACT and binding to the promoter of *act*II-ORF4 was tested directly *in vitro*. A limitation of such studies is that the relative specificity and affinity of interactions, *i.e.* likely functional relevance, is difficult to gauge. Evidence for binding *in vivo* such as that obtained using chromatin immunoprecipitation (ChIP) -based approaches (Rico *et al.*, 2014, McKenzie & Nodwell, 2007) is lacking in most cases, and in all cases the effects of disrupting the binding sites of transcription factors on promoter activity has not been determined to our knowledge.

The first transcription factor to join AtrA as a regulator of *act*II-ORF4 was AbsA2, the response regulator of the AbsA two-component system, which negatively controls antibiotic production in *S. coelicolor* (McKenzie & Nodwell, 2007, Aceti & Champness, 1998, Anderson *et al.*, 2001). In addition to showing that AbsA2 binds *in vitro* to the promoter region of *act*II-ORF4 (plus the promoter regions of *cdaR* and *redZ*, which

encode CSRs of the calcium-dependent antibiotic and RED, respectively), evidence for binding *in vivo* was provided by showing the promoter region of *act*II-ORF4 could be amplified in an extract precipitated using anti-AbsA2 antibodies (McKenzie & Nodwell, 2007). However, an extract from an *ΔabsA* strain was not included as a negative control. The signal to which the AbsA locus responds and consequently the wider physiological of AbsA-mediates regulation is not known.

AtrA was next joined by DasR, a member of the GntR family that controls the uptake and metabolism of N-acetylglucosamine (GlcNAc) and the degradation of chitin to N-acetylglucosamine (Rigali *et al.*, 2008, Rigali *et al.*, 2006). The DasR regulon is induced by glucosamine-6-phosphate (GlcN-6-P) (Rigali *et al.,* 2006), an internal intermediate of GlcNAc, allowing mycelia to make maximum use of GlcNAc when it is available. Analysis of the promoters of genes within the DasR regulon identified a conserved sequence called the DasR responsive element (*dre*), which was subsequently found in the promoter region of *act*II-ORF4. Binding of DasR to the promoter region of *act*II-ORF4 was shown *in vitro* and subsequently confirmed *in vivo* by ChIP in combination with sequencing (Rigali *et al.,* 2008, Swiatek-Polatynska *et al.*, 2015). The *dre* in the promoter region of *act*II-ORF4 overlaps the -35 box of the promoter and as is thought to mediate repression of *act*II-ORF4 in the absence of induction by GlcN-6-P. It is thought that de-repression occurs under conditions of famine via GlcNAc derived from the degradation of peptidoglycan as part of wider recycling of the vegetative material to fuel the production of aerial hyphae and the production of spores (Rigali *et al.*, 2008, Manteca *et al.*, 2006).

The analysis of proteins in crude extracts of *S. coelicolor* that could be "captured" *in vitro* using immobilised fragments of promoter regions identified AdpA and the products of SCO5405 and SCO1480 as candidate regulators of *act*II-ORF4 after confirming binding using electrophoretic mobility shift assays and purified components (Park *et al.*, 2009). AdpA was first described in *S. griseus* as part of a cascade that facilitates morphological development and antibiotic production in response to the accumulation of a microbial hormone called A factor (2-isocapryloyl-3-R-

hydroxymethyl-γ-butyrolactone) (Khokhlov *et al.*, 1973, Horinouchi, 2002, Horinouchi, 2007). AdpA was named A factor-dependent protein as its production requires induction by the binding of A factor to ArpA (A-factor receptor protein A), which would otherwise continue to repress transcription from the *adpA* promoter (Onaka & Horinouchi, 1997, Ohnishi *et al.*, 2005). AdpA in *S. griseus* has been shown to directly control the expression of hundreds of genes, acting as a repressor or activator of transcription (Onaka *et al.*, 1995, Ohnishi *et al.*, 2004, Higo *et al.*, 2012). Antibiotic and morphological development in *S. coelicolor*, like *S. griseus*, is dependent on *adpA* (also called *bldH*), but unlike *S. griseus*, is not under the control of an A factor-like γ-butyrolactone (Takano *et al.*, 2003, Nguyen *et al.,* 2003, Chater & Horinouchi, 2003). The binding of DNA by *S. coelicolor* AdpA has been studied and the consensus binding site found to be identical to that of *S. griseus* AdpA (Yamazaki *et al.*, 2004, Lee *et al.*, 2013). Unfortunately, the signal(s) that is relayed by AdpA in *S. coelicolor* is not known. The products of SCO5405 (*absC*) and SCO1480 encode a member of the MarR family and a novel nucleoid-associated protein specific to Actinobacteria, respectively. The signal relayed by SCO5405 (*absC*) and the role of SCO1480 beyond being associated with chromosomal DNA are not known. ROK7B7 (SCO6008) was also captured, but binding was not confirmed, at least clearly, by EMSA. ROK7B7 is a member of the ROK family and regulates the utilisation of xylose, carbon catabolite repression, and secondary metabolism in *S. coelicolor* (Swiatek *et al.*, 2013, Park *et al.,* 2009).

In a separate study, DraR, the response regulator of another TCS (SCO3062/3063), was shown to bind the promoter region of *act*II-ORF4 *in vitro* and the site of binding was mapped by footprinting. Deletion of this TCS causes a reduction in production of ACT and an increase in the production of RED and the yellow-pigmented type I polyketide (γCPK) on minimal medium with glutamate. Under the same growth conditions, the production of ACT, RED and CDA was shown to require the AfsQ TCS (Yu *et al.*, 2012, Shu *et al.*, 2009). Moreover, the response regulator (AfsQ1) was shown to bind promoter regions of *act*II-ORF4, *red*Z and *cdaR in vitro* and the sites of binding mapped by footprinting (Wang *et al.*, 2013). RedZ and CdaR are CSRs of RED and CDA production, respectively. The RED cluster contains another CSR called RedD

(Malpartida *et al.*, 1990). The response regulator of the AbrC TCS has also been shown to bind to the promoter region of *act*II-ORF4. The AbrC system is atypical in that it has two histidine kinases. Deletion of the gene encoding the RR (*abrC3*; SCO4596) reduces ACT and RED production and delays morphological development. The binding of AbrC3 to the promoter region of *act*II-ORF4 was shown in this case by ChIP-chip *in vivo* as well as EMSA *in vitro* (Rico *et al.*, 2014). Again the signal transduced the AbrC system is unknown.

Crp, the cyclic AMP receptor protein, which is perhaps best known for mediating carbon catabolite repression of the *lac* operon in *E. coli* (Gorke & Stulke, 2008) and controls diverse cellular processes in many bacteria (Korner *et al.*, 2003), is a key regulator of secondary metabolism as well as spore germination and colony development in *S. coelicolor* (Piette *et al.*, 2005, Gao *et al.*, 2012). Carbon catabolite repression is an important control system in carbon breakdown, allows bacteria to metabolic and uptake preferred (rapidly metabolisable) substrate in the presence of different carbon sources (Deutscher, 2008). Using ChIP-chip and transcriptome assays, it has been shown that Crp binds within the promoter region of *act*II-ORF4 (plus those of other CSRs) and appears to activates transcription (Gao *et al.,* 2012). In addition to direct regulation, the expression of *act*II-ORF4 is also completely dependent on *relA*, which is required for the switching on of the stringent response (Sun *et al.*, 2001, Kang *et al.*, 1998). The stringent response enables bacteria to survive sustained periods of nutrient deprivation by enhancing the transcription of numerous genes required to survive stress, while lessening transcription of genes, such as those encoding the stable RNAs, whose products are required in significantly reduced amounts during periods of slowed growth. *S. coelicolor relA* deletion mutants are unable to produce detectable amounts of ACT and RED and normal levels of the mRNA of *act*II-ORF4 and *redD* suggesting a direct role of the stringent response in activating transcription from the promoters of these genes. The relative positions of the sites of binding for the transcription factors described above are shown in (Figure 1.6).

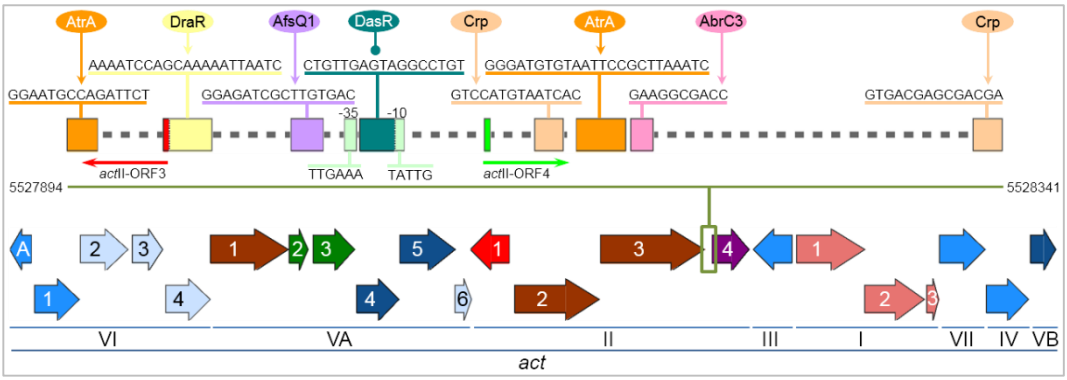

Figure 1.6: Binding of global regulatory proteins to the *act*II-ORF4 promoter. The pathway-specific activator ActII-4 integrates many of the global regulators. Transcriptional activation in the network is illustrated by line ending with an arrow, whereas transcriptional repression indicated by spherical end. The binding sites for all regulators are experimental proved. The central line (olive green) represents the chromosomal position of magnified region (olive green box) including noncoding intergenic region between *act*II-ORF3 and *act*II-ORF4. AtrA binds at position (- 200/ - 186) and (+ 45/ + 68) from *act*II-ORF4 start codon (Uguru *et al.,* 2005), DasR binds to the region between -35 and - 10 at position (- 60/ - 44) (Rigali *et al.,* 2008), DraR binds at (- 151/ - 131) (Yu *et al.,* 2012), AfsQ1 bind at (- 93/ - 78) (Wang *et al.,* 2013), AbrC3 bind at (+ 72/ + 81) (Rico *et al.,* 2014) and Crp bind at (+ 25/ +38) and Crp (+ 235/ + 248) (Gao *et al.,* 2012).

1.4 The regulon of *atrA* - unpublished work leading to this project

DNA Footprinting of *S. coelicolor* AtrA identified two binding sites within the *act*II-ORF4 promoter and one within the *S. griseus strR* promoter (Uguru *et al.,* 2005). To obtain better knowledge of the range of sequences that could be recognised by AtrA, double-stranded DNA ligands (called aptamers) were selected (Bunka, Yepes and Cobley, unpubl. results) using the SELEX procedure (Tuerk & Gold, 1990, Gold, 1995). Sequencing of 20 clones revealed 7 unique sequences of which 5 were similar to those that had been footprinted previously in the promoter region of *act*II-ORF4 (Figure 1.7).

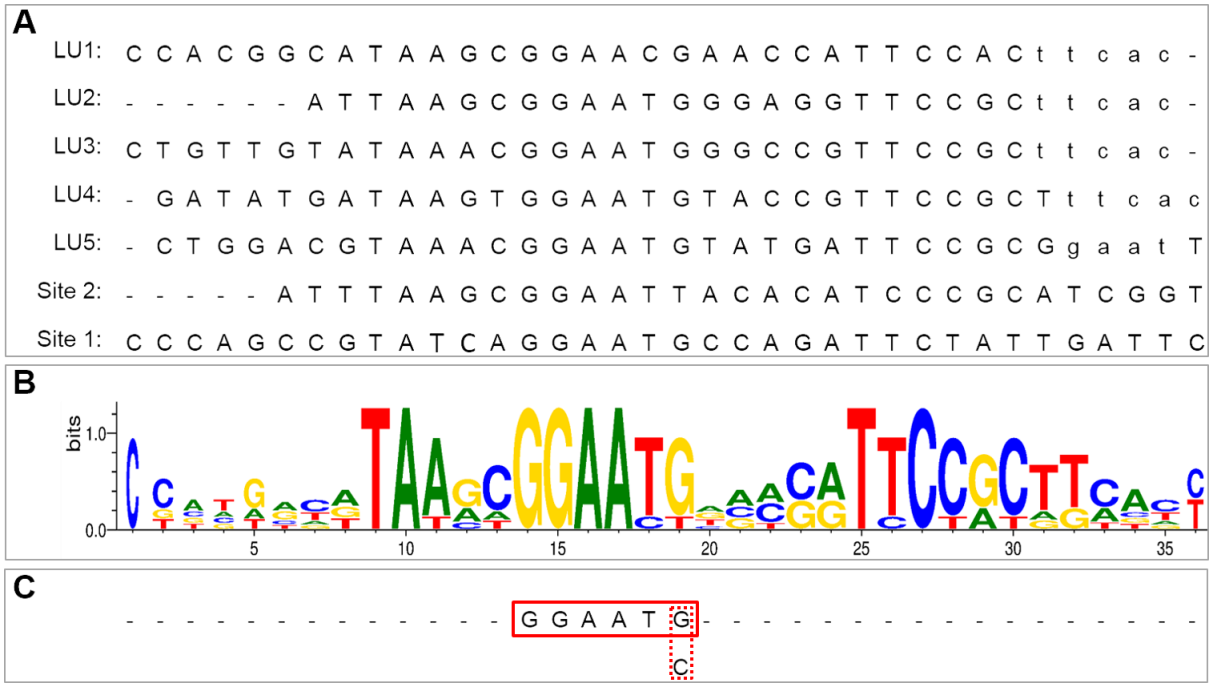

Figure 1.7: Recognition of AtrA binding sequence using aptamers. Panel A shows the 5 sequence alignment of AtrA-bound aptamers. The reverse complement (RC) of aptamer LU5 and Site 2 for the *act*II-ORF4 promoter (Uguru *et al.,* 2005) are shown to maximise sequences matches when aligned with the other sequences. Lower case indicates sequences found in constant (primer) regions of the SELEX template. **Panel B** represents a sequence WebLogo derived from AtrA aptamers. **Panel C** shows the original consensus sequence (Uguru *et al.,* 2005) that coincides with the sequence logo in panel B.

Alignment of the sequences identified a 25 nt, palindromic motif between positions 9 and 33 inclusively. A palindromic sequence is expected for a member of the TetR family, which uses a dimeric, head-to-head unit to bind DNA (Meier *et al.*, 1988, Cuthbertson & Nodwell, 2013, Hillen & Berens, 1994). The sequences between positions 9 and 33 motif were then used to build a position weight matrix, which in turn was used to scan the *S. coelicolor* genome for similar sequences. This was done using PREDetector (see Table S1) (Hiard *et al.*, 2007).

PCR was then used to amplify segments containing candidate binding sites for AtrA, and binding was assayed *in vitro* using EMSAs (Boomsma, unpubl. results). Sequences that were found to be bound by AtrA using the electrophoretic mobility shift assay were added to the alignment and the *in silico* prediction of binding sites repeated. It

should be noted that the most recent predictions were generated using a core 19 nt motif (positions 12 to 30 in the original alignment). Sequences beyond this core were found to be not highly conserved in genomic sequences that bind AtrA *in vitro*.

Several genes associated with acetyl-CoA metabolism are predicted to be bound directly by AtrA *in vivo*. These are SCO5424 (*ackA*), SCO5529 (*leuA2*), SCO6027 and SCO6026 (encoding acetyl-CoA acetyltransferase and the α-subunit of the fatty acid oxidation complex, respectively). In addition, it should be remembered that SCO4921 (*aacA2*) is thought to be a potential target for *act*II-ORF4, i.e. subject to indirect regulation by *atrA*. The locations of the predicted binding sites and the positions of each of the corresponding genes in acetyl-CoA metabolism are shown in Figure 4.14. A role for AtrA in the regulation of acetyl-CoA metabolism is potentially very exciting, as acetyl-CoA provides the precursors for the production of many types of antibiotic and other therapeutic agents (see above). Thus, AtrA may have a pivotal role in coordinating primary and secondary metabolism.

1.5 General and Specific aims

The broad objective of this proposal is to better understand the role of AtrA in the growth, physiology and morphological development of *S. coelicolor*, with a view to uncovering hitherto undescribed links between primary and secondary metabolism that might allow the manipulation of streptomycetes to awaken the production of cryptic SMs and aid the discovery of new potential antibiotics, which are much needed given the alarming increase in resistance amongst bacterial pathogens (Beceiro *et al.*, 2013, Beever *et al.*, 2015).

Specific aims of the project were to:

1. Develop a pipeline for the generation of maps of transcription units that detail the positions of transcription start sites; thereby providing a platform to interpret the functional significance of the binding of regulators of transcription such as AtrA to specific chromosomal locations.
2. Assay the binding of AtrA to the promoters of a selection of genes identified by scanning the genome sequence of *S. coelicolor* using the position weighted matrix for sites of AtrA binding and the effects of disrupting *atrA* on the transcription of potential targets of AtrA regulation in *S. coelicolor*.
3. Investigate the possibility that the DNA-binding activity of AtrA, like many other members of the TetR family, is regulated by a small molecule(s).
4. Tag AtrA with the 3 X FLAG™ tag and confirm that this recombinant protein is still functional and suitable for the mapping of sites of AtrA binding by ChIP-sequencing.

Chapter 2

2 Material and Methods

2.1 Bacterial strains and culture conditions

All bacterial strains used are listed in Table 2.1. *S. coelicolor* strains were cultivated at 30°C with shaking (220 rpm) in 250 mL Erlenmeyer flasks fitted with a spring baffle to aid dispersed growth of the mycelia (Kieser *et al.*, 2000) and containing 50 mL of media. *Escherichia coli* strains were cultivated at 37°C with shaking (220 rpm) in either 50 mL Falcon tube containing 10 mL media for the purpose of cloning or in 2 L flask containing 500 mL media for protein production.

Strains	Description	Source
***Streptomyces* Strains**		
M145	SCP1-, SCP2-. Contains a mutation of the *sre-I* gene.	(Kieser *et al.,* 2000)
M145 (pAU3-45)	M145 containing plasmid pAU3-45	This thesis
M145 (pAU-atrA3x)	M145 containing plasmid pAU-atrA3x, which encodes *S. coelicolor* AtrA with 3 x FLAG tag™ at the C-terminus.	This thesis
M145 (pAU-atrA)	M145 containing plasmid pAU-atrA, which encodes *S. coelicolor* AtrA	This thesis
M145 (pAU-3xatrA)	M145 containing plasmid pAU-3xatrA, which encodes *S. coelicolor* AtrA with 3 x FLAG tag™ at N-terminus.	This thesis
M145 (pSET152)	M145 containing plasmid pSET152	This thesis
M145 (pSET-atrA3x)	M145 containing plasmid pSET-atrA3x, which encodes *S. coelicolor* AtrA fused with 3 x FLAG tag™ at C-terminus.	This thesis
M145 (pKC1139)	M145 containing plasmid pKC1139	This thesis
M145 (pKCEatrA-gl-3xFlag)	M145 containing plasmid pKCEatrA-gl-3xFlag which encodes *S. globisporus* AtrA with 3 x FLAG tag™ at C-terminus.	This thesis
L645	M145 containing a disrupted *atrA* gene.	(Uguru *et al.,* 2005)
L645 (pAU3-45)	L645 containing plasmid pAU3-45	This thesis
L645 (pAU-atrA3x)	L645 containing plasmid pAU-atrA3x, which encodes *S. coelicolor* AtrA with 3 x FLAG tag™ at C-terminus.	This thesis
L645 (pAU-atrA)	L645 containing plasmid pAU-atrA, which encodes *S. coelicolor* AtrA	This thesis
L645 (pAU-3xatrA)	L645 containing plasmid pAU-3xatrA, which encodes *S. coelicolor* AtrA with 3 x FLAG tag™ at N-terminus.	This thesis
LA145 (L646)	M145 containing plasmid pL646, which constitutively expresses *S. coelicolor atrA*	(Towle, 2007)
M511	M145 containing a disrupted *act*II-ORF4 gene (Δ*act*II-ORF4)	(Floriano & Bibb, 1996)

M1146	M145 derivative in which four well described metabolite gene clusters have been deleted (*Δact*, *Δred*, *Δcpk*, *Δcda*)	Dr Ryan Seipke, (Gomez-Escribano & Bibb, 2011)
***E. coli* Strains**		
DH5α	F$^-$ Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK$^-$, mK$^{+)}$ *pho*A *sup*E44 λ$^-$ *thi*-1 *gyr*A96 *rel*A1	Laboratory stock
ET12567 (pUZ8002)	*dam13*::Tn*9(chl^R) dcm-6 hsdM hsdR recF143 zjj-201*::Tn*10 galK2 galT22 ara14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtlI glnV44*, pUZ8002 (*kan*^{R})	Dr Ryan Seipke, Laboratory stock
XL1-Blue (MRF')	Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F´ *proAB lacI^{q}ZΔM15* Tn*10* (tet^R)]	Dr Ryan Seipke, Laboratory stock
BL21-Gold (DE3)	*E. coli* B F$^-$ *ompT hsdS*(r$_B^-$ m$_B^-$) *dcm*$^+$ tet^R *gal* λ(DE3) *endA* Hte	Stratagene Corp.

Table 2.1: Description and sources of bacterial strains.

2.2 Media and chemicals

Unless otherwise stated, all media, chemicals, solutions were purchased from Fisher Scientific UK Ltd, Sigma-Aldrich Ltd, MP Biomedicals™, Alfa Aesar®, Oxoid™, or Lab M™. Media and solutions were prepared as instructed by the vendor and sterilized by autoclaving at 121°C for 20 min. Some components were sterilized by filtration through a 0.22 µm Millex®GP syringe filter unit, as indicated.

2.2.1 R5 (Kieser *et al.,* 2000)

The following were added to 1 L of dH_2O and then autoclaved: 103 g sucrose, 0.25 g K_2SO_4, 10.12 g $MgCl_2.6H_2O$, 10 g glucose, 0.1 g casamino acids, 2 mL trace elements (1 L made of 4 mg $ZNCl_2$, 200 mg $FeCl_3.6H_2O$, 10 mg $CuCl_2.4H_2O$, 10 mg $MnCl_2.4H_2O$, 10 mg $Na_2B_4O_7.10H_2O$ and 10 mg $(NH_4)_6Mo7O_{24}.4H_2O$), 5 g yeast extract, 5.73 g TES buffer, and 20 g agar (Fisher Scientific). After autoclaving, the following reagents, which had been autoclaved separately, were added: 10 mL 0.5% KH_2PO_4, 4 mL 5 M $CaCl_2.2H_2O$, 15 mL 20% L-proline, and 7 mL 1N NaOH.

2.2.2 Minimal medium (MM) (Hopwood, 1967)

The following were added to 1 L of dH_2O, the pH adjusted to between 7.0 and 7.2 and then autoclaved: 1 mg $(NH_4)_2SO_4$, 0.5 g K_2PO_4, 0.2 g $MgSO_4.{}_7H_2O$, 0.01 g $FeSO_4.7H_2$). Then 10 mL of 50% mannitol, which had been sterilized by filtration through a 0.22 µm filter unit, was added as a carbon source.

2.2.3 Mannitol soya flour medium (MSF) (Hobbs *et al.*, 1989)

The following were added to 1 L of tap water: 20 g from each of agar, mannitol and soya flour and autoclaved twice, with gentle shaking after each cycle.

2.2.4 Yeast extract-malt extract medium (YEME) (Kieser *et al.,* 2000)

The following were added to 1 L of dH_2O and then autoclaved: 3 g yeast extract, 5 g peptone, 3 g malt extract, and 340 g sucrose (34%). After autoclaving 2 mL of 2.5 M $MgCl_2.{}_6H_2O$ was added.

2.2.5 2X YT medium (Kieser *et al.,* 2000)

The following were added to 1 L of dH_2O and then autoclaved: 16 g tryptone, 10 g yeast extract, and 5 g NaCl.

2.2.6 Tryptone Soya Agar (TSA)

The following were added to 1 L of dH_2O and then autoclaved: 17 g pancreatic digest of casein, 3 g pancreatic digest of soyabean meal, 5 g sodium chloride, 2.5 g Di-basic potassium phosphate, 2.5 g glucose, and 20 g agar.

2.3 Antibiotics

All antibiotics were sterilized by filtration and used at the following final concentrations: ampicillin (AMP), 50 µg/mL; apramycin (AM), 50 µg/mL; thiostrepton (TS), 50 µg/mL; chloramphenicol (CHL), 25 µg/mL; kanamycin (KAN), 25 µg/mL; and nalidixic acid (Piette *et al.,* 2005), 25 µg/mL.

2.4 Preparation of *S. coelicolor* spore stocks

Spore stocks of *S. coelicolor* were prepared by scrapping spores from the surface of a well isolated colony on an agar plate using a sterile loop and streaking onto R5 agar to produce well-isolated colonies. After incubation for an appropriate time, the spores from a single colony were harvested using a sterile loop and resuspended in 5 mL of sterilized dH_2O. Multiple aliquots (~300 µL) of the suspension were spread over mannitol soya flour medium (MSF) plates with a sterile glass spreader and incubated until the surface of confluent growth displayed a grey colouration. The spores were harvested by adding 10 mL of sterilized distilled water to each plate and scraping the medium surface with sterile cotton swap. The combined suspensions were filtered through a plug of sterilized cotton (~5 cm^3) inserted into the barrel of 20 mL syringe. The filtered spores were centrifuged at 4700 x *g* for 10 min and the pellet was re-suspended in sterile 25% glycerol. The spore suspension was divided into aliquots and stored at -20°C.

2.5 Nucleic acid methods

2.5.1 Extraction of chromosomal DNA from *S. coelicolor* and *E. coli* for use in PCR

An overnight culture of 50 mL was poured into 50 mL Falcon tube containing a $1/8^{th}$ volume (6.25 mL) of stop solution (95% [v/v] ethanol; 5% [v/v] phenol). The culture was centrifuged at 5000 x *g* (Sigma, 11133-rotor) for 10 min and the cell pellet washed with 10 mL sterilized dH_2O and re-harvested. The pellet was suspended in 10 mL of 1 M TE buffer (pH 8) containing lysozyme (20 mg/mL). The suspension was incubated for 60 min at 37°C. SDS and NaCl were added to final concentrations of 0.5% [w/v] and 150 mM, respectively. The tube was vortexed briefly and added to a boiling-water bath for 1 min and then cooled down by placing in ice. An equal volume of buffer-saturated phenol (pH 8) was added and the tube vortexed. The cell lysate was split between multiple 2 mL Eppendorf tubes and the cell debris removed by centrifugation at 13,000 x *g* (Eppendorf centrifuge 5415 R) for 10 min at room temperature. The upper layer from each tube was transferred to a new tube and extraction further using an equal volume of phenol (pH 8): chloroform: isoamyl alcohol (25:24:1) as described for the

phenol extraction. The extraction was then repeated using chloroform: isoamyl alcohol (49:1). The resulting aqueous phase was transferred to new tubes and a 2.5 volume of absolute ethanol was added to each. NaCl was added to a final concentration of 150 mM and the mixture incubated at -20°C for 30 min. The precipitate was harvested by centrifugation at 13,000 x *g* for 30 min at 4°C. Finally, the pellet was washed with 70% [v/v] ethanol, re-suspended in nuclease-free water (Sigma) and stored at -20°C.

2.5.2 *Streptomyces* RNA extraction from broth culture

RNA was isolated from *S. coelicolor* grown in YEME, as described previously (Lin *et al.*, 2013), with minor modification. At the required stage of growth, a $1/8^{th}$ volume of stop solution was added to inhibit cell metabolism (Kime *et al.*, 2008) and the cells were harvested by centrifugation at 4700 x *g* for 10 min (Sigma, 11180-rotor). When necessary, cell pellets were stored frozen at -80°C. In order to lyse the cells, 5.0 OD_{600} units of mycelia were re-suspended in 375 µL sterilized dH_2O and transferred to Lysing Matrix B tube (MP Biomedicals) containing 375 µL of modified Kirby mix (1% [w/v] SDS, 6% [w/v] sodium-4-aminosalicylate, and 6% [v/v] phenol buffered with 50 mM Tris-HCl [pH 8.3]; (Kieser *et al.,* 2000). Lysis was achieved using a benchtop homogenizer (FastPrep®-24 MP Biomedicals, India). Mycelial samples were subjected to three cycles of 50 s bursts at 6.5 M/s with 1 min incubation in ice between each cycle. The resulting lysates were extracted with an equal volume of phenol saturated with 100 mM citrate buffer (pH 4.3), which involved mixing by inverting the tube several times. The phases were separated by centrifugation at 13,000 x *g* for 10 min at room temperature. The aqueous phase was then extracted with an equal volume of phenol (pH 4.3): chloroform: isoamyl alcohol (25:24:1), and once with an equal volume of chloroform: isoamyl alcohol (49:1). Total nucleic acid was then precipitated by adding 2.5 volumes of absolute ethanol and NaCl to 150 mM. The samples were incubated for 30 min at -20°C then centrifuged at 13,000 x *g* for 30 min at 4°C. The pellet was washed with 70% [v/v] ethanol, dissolved in nuclease-free water (Sigma) and stored at -80°C.

2.5.3 DNase I Treatment

In order to obtain RNA samples free of DNA, 100 µg of nucleic acid was incubated with 2 U of DNase RQ1 (Promega) and 100 U of RNaseOUT^{TM} (Invitrogen) in 50 µL of 1 x RQ1 DNase buffer at 37°C for 60 min. The reaction mix was extracted with phenol: chloroform: isoamyl alcohol as described above and DNA in the aqueous phase precipitated by the addition of 2.5 volumes of absolute ethanol and NaCl to a final concentration of 150 mM. The mixture was incubated at -20^{o}C for 30 min, and then the DNA was harvested by centrifugation at 13,000 x *g* for 30 min at 4°C. The pellet was washed with 70% [v/v] ethanol, air dried and re-suspended in nuclease-free water (Sigma). To check the integrity of RNA, 500 ng of a sample was run in a 1.2% [w/v] agarose gel for 40 min at 10 V cm^{-1}. The gel was imaged using a GeneGenius UV transilluminator (Syngene).

2.5.4 Quantify the Nucleic Acid Concentration

The concentration of nucleic acid was determined by NanoDrop^{TM} 1000 spectrophotometer (Thermo Fisher Scientific).

2.6 Preparation of mycelial patches

2.6.1 Solid media overlaid with cellophane membrane

Cellophane discs (A. A. Packaging Ltd.) were soaked in dH_2O and sterilized by autoclaving after placed between blotting paper. Two pairs of sterile forceps were used to lay the membranes on surface of solid media plates and exclude any bubbles. The plates were dried in a laminar flow (Faster TWO 30) for a few minutes.

2.6.2 *Streptomyces* RNA isolation from plate

Appropriate numbers of spores were spotted onto TSA or minimal media plates overlaid with cellophane and allowed to dry. Plates were incubated at 30°C for the required amount of time. Mycelium patches were scraped off the surface of the cellophane using a sterile blunt-end spatula and placed in 200 µL sterilized nuclease-free water (Sigma) containing 1/8th volume of stop solution and harvested by

centrifugation at 13,000 x *g* for 1 min. The pellets were transferred to Lysing Matrix B tube (MP Biomedicals) containing modified Kirby mix (Kieser *et al.,* 2000) and the patches smashed using sterilized glass rod. The suspensions were sonicated using a benchtop homogenizer (FastPrep®-24 MP Biomedicals, India), and then extracted with phenol (pH 4.3), phenol: chloroform: isoamyl alcohol, and chloroform: isoamyl alcohol as described previously in Section 2.5.2.

2.7 RNA sequencing methods

2.7.1 RNA extraction

S. coelicolor was grown in YEME broth and RNA isolated as described previously for *Propionibacterium acnes* (Lin *et al*., 2013). Briefly, cell pellets were resuspended in Kirby mix (Kieser *et al.,* 2000). Cells were lysed by three cycles of homogenizing for 1 min (set at 6.5 M/s) followed by cooling in an ice-water bath for 1 min. RNA was extracted using an equal volume of acidic phenol: chloroform: isoamyl alcohol (50: 50: 1), and then chloroform: isoamyl alcohol (49: 1). The RNA was precipitated, treated with DNase I, quantified, and integrity was determined as mentioned in Sections 2.5.2, 2.5.3 and 2.5.4. Samples were then enriched for mRNA using *MICROBE*xpress™-Bacteria beads, as described by the manufacturer (Ambion).

2.7.2 Differential RNA sequencing

Differential RNA-seq data was generated by Vertis Biotechnologie AG (Germany) as a service that included the construction of cDNA libraries before and after treatment with TAP, and the alignment of RNA-sequence reads to the corresponding genome positions (Lin *et al*., 2013), which were retrieved from the NCBI database (Pruitt *et al.*, 2007). For each genome position, the number of times it was the first nucleotide in sequence reads, *i.e.* associated with a 5′ end *in vivo*, was counted. This was done separately for each of the two libraries and the counts compared. It should be noted that, as described previously (Lin *et al*., 2013), the 5′-sequencing adaptor was ligated to transcripts prior to fragmentation, thereby allowing the 5′ ends of both long and short transcripts to be detected.

2.7.3 Global RNA sequencing

Global RNA-seq was performed at the Wellcome Trust Sanger Centre (Cambridge, UK) by Lira Mamanova using a published methodology (Mamanova *et al.*, 2010) and the sequences processed as described previously (Lin *et al*., 2013). After aligning the gRNA-seq reads to the genome, the number of times each genome position was present in a read irrespective of its position was counted. For both types of RNA-seq, the reference genomes for *S. coelicolor* A3(2) strain M145 was AL645882. The RNA-seq data has been deposited in the GEO archive (Barrett *et al.*, 2013) under accession number GSM1126846. The same RNA from *S. coelicolor* was also analysed using custom 105,000 x 60 mer whole genome arrays manufactured by Agilent Technologies (Lewis *et al.*, 2010) in collaboration with Prof. Colin Smith and members of his group (University of Surrey). The preparation of cDNA, labelling, and hybridization was performed as described previously (Bucca *et al.*, 2009, Bucca *et al.*, 1997). Two technical repeats of co-hybridising the RNA with labelled genomic DNA were performed. A single Log_2 RNA/gDNA value for each probe that passed the Agilent probe-quality criteria on at least one array was generated by averaging the global median normalized microarray signals. Corresponding Log_2 "probe" signals from the RNA-seq data were generated by averaging the signals of each base within the 60-bp sequence that a probe targets. RNA-seq coverage vectors for the forward and reverse strands were used to generate the data for forward-gene-targeting and reverse-gene-targeting probes respectively.

2.8 Northern blotting

RNA for northern blotting was isolated as described in Section 2.5.2. For each sample, an aliquot of 5-10 µg was mixed with an equal volume of 2 x RNA-loading dye (New England BioLabs), denatured by incubation at 90°C for 90s, chilled on ice, and analysed along with other samples by denaturing electrophoreses using a 6% [w/v] polyacrylamide sequencing-type gel (acrylamide: *bis*-acrylamide was 29: 1) containing 1 x TBE (Severn Biotech Ltd) and 7 M urea. Fractionated RNA was electro-transferred to a Hybond-N^+ membrane (Amersham) using 20 x saline-sodium citrate (SSC) buffer at

11 V for 1 h, and subsequently fixed to the membrane by UV-crosslinking. The nylon membrane was cut into separate strips for each group of lanes.

Specific *S. coelicolor* transcripts were probed using riboprobes generated by *in vitro* transcription. The primers used to construct the templates for the riboprobes are listed in Table S2. Reactions of 20 µL contained 100 nM template (produce by PCR), 100 U T7 RNA polymerase (Invitrogen), 8 pmoles α-^{32}P UTP (3000 Ci mmol^{-1} 10 mCi mL^{-1}, 250 µCi; Perkin-Elmer), 5 µM UTP, 0.5 mM rATP, rGTP and rCTP, 1 U yeast inorganic pyrophosphatase (Sigma), 80 U RNaseOUT™ in 1 x T7 RNA polymerase buffer (40 mM Tris-HCl [pH 8.0], 8 mM $MgCl_2$, 2 mM spermidine-$(HCl)_3$, 25 mM NaCl, and 5 mM DTT0. Reactions were incubated at 37°C for 3 h. Unincorporated nucleotides were removed by adding nuclease free water (Sigma) to 50 µL and passing reaction products through G-25 spin columns, as per the manufacturer's instructions (GE healthcare). Hybridizations were carried out overnight in the PerfectHyb™ Plus Hybridization buffer (Sigma-Aldrich) at 68°C by incubating each membrane strip separately with a specific riboprobe. After hybridization, 2 x 20 min washes in high stringency buffer (0.5 x SSC, 0.1% [w/v] SDS) at 68°C was performed. The hybridized blots were then exposed to a phosphoimager plate (Fuji) and read with a FLA 5000 scanner (Fuji).

2.9 PCR Methods

The volume of PCR was 50 µL. GoTaq polymerase (Promega) was used for amplifying small DNA products (80-700 bp), while FailSafe™ Premix J and FailSafe™ PCR Enzyme Mix (Epicentre Technologies, Cambio, Ltd) were used for gene amplification longer than 700 bp from *S. coelicolor*.

2.9.1 Standard PCR Method

The following conditions were used for standard PCR: the initial denaturation was 95°C for 5 min, followed by 40 cycles of denaturation at 95°C for 30 s, annealing at temperature indicated in Table S2 for 30 s and extension time at 72°C for 1 min/kbp, and a final extension at 72°C for 5 min.

2.9.2 Purification of PCR amplified DNA fragments

PCR products were analysed by gel electrophoresis. The concentration of agarose was dependent on the sizes of the expected products. A high percentage of agarose was use to analysed small fragments of DNA, and vice versa. 1X TAE buffer was used to analysis digested plasmids as the image resolution slightly better than 1X TBE which used to analysis PCR products (<2 kb). Concentration of agarose gels and running buffer are stated in the corresponding figure legends. Products of the expected size were cut out from the gel using a clean blade. The DNA was then extracted from the gel slice and purified using the QIAquick® gel extraction kit, as per the manufacturer's instructions. DNA was eluted from spin columns using 100 µL of nuclease-free water (Sigma), and then quantified using NanoDrop™ 1000 spectrophotometer. To prevent bleaching, fluorescent PCR products were stored in foil covered tubes at -20°C.

2.9.3 Reverse Transcriptase Quantitative polymerase chain reaction (RT-qPCR)

A 429 bp DNA fragment from *rpsL* gene was amplified by PCR using primers SCO4735p1 and SCO4735p2. The PCR product was purified and quantified as mentioned in Section 2.9.2. To test the reliability of qPCR, 4-fold serial dilution series from *rpsL* DNA fragment ranging from 2 ng/ µL to 1.907 X 10^{-6} ng were prepared to construct standard curve. Each dilution was analysed by PCR using a SensiMix™ SYBR No-ROX kit, as instructed by the vendor (Bioline) in combination with 0.5 µM of primers and 3.5 mM $MgCl_2$. Reactions were assembled in 0.1 mL strip tubes (Corbett) and analysed using a Corbett Rotor-Gene 6000 system. The PCR conditions were 10 min at 95°C, followed by 40 cycles of 30 s at 95°C, 20 s at 55°C and 30 s at 72°C. The CT value in each dilution were measured and plotted against the logarithm of their concentration.

Quantitative PCR analysis was carried out following reverse transcription that used SuperScript® II, as instructed by the vendor (Invitrogen), in combination with 500 ng of DNase I-treated total RNA as template and 50 ng of random hexamers (Applied Biosystems) as primer in 20 µL reactions. Mock reactions minus the reverse transcriptase were also conducted to control for residual chromosomal DNA

contamination (Uguru *et al.*, 2005). At the end of each of the reactions, 80 µL of yeast tRNA (Ambion) was added to a final concentration of 8 ng/µL, and 2 µL aliquots analysed by PCR as mentioned above. The PCR conditions were 10 min at 95^{o}C, followed by 40 cycles of 10 s at 95°C, 15 s at 60°C or 55°C and 15 s at 72°C. The products were analysed by determining melting curves and the sizes of the amplicons. The latter was done as a second step by gel electrophoresis. A threshold common to the exponential phases of all the reactions in a single run was selected manually and the corresponding number of cycles for each reaction recorded. These CT values were then expressed as the difference relative to *rpsL* (SCO4735), which was used as the internal control. In turn, the ΔCT values were used to calculate the difference in abundance using the equation $2^{\Delta CT}$. qPCR for each gene was run in duplicate and no amplification was detect for negative control (NTC) up to cycle 35.

2.10 Oligohistidine tagged protein

Fragments for cloning were generally generated by PCR (see Section 2.9.1). NdeIatrA-F primer, which adds CAT to the start codon of *atrA* generating an NdeI cloning site and an atrABamHI-R primer, which adds a BamHI cloning site after the stop codon of the recombinant gene were used to amplify *atrA*3x from pAU-atrA3x construct. The resulted DNA fragments and vectors were subjected to double digest with NdeI and BamHI. The digested vector purified using a QIAquick® PCR purification kit while the insert was purified and quantified as described in Sections 2.9.2. The molar ratio of vector: insert of 1: 4 was used to prepare 20 µL ligation reactions in 1 x T4 ligase buffer (Invitrogen) containing 1 U T4 DNA ligase (Invitrogen). The reaction was incubated at 16°C for 15 h, and then inactivated by heating at 65°C for 10 min.

Plasmid	Size	Description	Source
pUC19	2686bp	*E. coli* cloning vector, ampicillin resistance.	Laboratory stock
pSET152	5715bp	*Streptomyces* cloning vector containing the ϕC31 attachment site, confers apramycin resistance.	Dr Bin Hong, (Bierman *et al.*, 1992)
pSET-atrA3x	7041bp	pSET152-derivatived plasmid containing 891 bp coding region of *S. coelicolor atrA* fused with 3 x FLAG tag™ at C-terminus under its own promoter.	This thesis
pKC1139	6500bp	*E. coli*/*Streptomyces* shuttle vector, apramycin resistance.	Dr Bin Hong,(Bierman *et al.,* 1992)
pKCEatrA-gl-3xFlag	7800bp	pKC1139 derivative plasmid containing 792 bp coding region of *S. globisporus atrA* fused with 3 x FLAG tag™ under control of *ermE**p.	Dr Bin Hong
pET16b	5711bp	A T7 *lac* promoter expression vector, encodes polyhistidine at N-terminus of recombinant proteins, ampicillin resistance.	Laboratory stock
pRA001	6600bp	pET16b-derivatived plasmid containing 901 bp DNA fragment including coding region of *S. coelicolor atrA*.	Laboratory stock, Jon Stead
pET-atrA3x	6666bp	pET16b-derivatived plasmid containing 891 bp coding region of *S. coelicolor atrA* fused with 3 x FLAG tag™ at C-terminus.	This thesis
pET16b-atrA-gl	6498bp	pET16b-derivatived plasmid containing 792 bp coding region of *S. globisporus atrA*.	Dr Bin Hong
pET16b-ActR	6483bp	pET16b-derivatived plasmid containing 780 bp coding region of *S. coelicolor actR* (*SCO5082*).	Dr Bin Hong
pET16b-TetR	6330bp	pET16b-derivatived plasmid containing 627 bp coding region of *tetR* from *E. coli* XL1-Blue MRF'.	Dr Bin Hong
pAU3-45	6525bp	pSET152-derivatived with *tsr* gene from pAU5 cloned into the *Nhe*I site, apramycin and thiostrepton resistance.	Stock of Dr Ryan Seipke, (Bignell *et al.*, 2005)
pAU-atrA3x	7851bp	pAU3-45-derivatived plasmid containing 891 bp coding region of *S. coelicolor atrA* fused with 3 x FLAG tag™ at C-terminus under its own promoter.	This thesis
pAU-atrA	7733bp	pAU3-45-derivatived plasmid containing 894 bp coding region of *S. coelicolor atrA* under its own promoter.	This thesis
pAU-3xatrA	7893bp	pAU3-45-derivatived plasmid containing 891 bp coding region of *S. coelicolor atrA* fused with 3 x FLAG tag™ at N-terminus under its own promoter.	This thesis

Table 2.2: List of plasmids and their sources.

2.11 Preparing *E. coli* competent Cells

E. coli competent cells were prepared as described previously (Inoue *et al.*, 1990). 10 mL of LB broth medium containing antibiotic to select for retention of plasmids as appropriate was inoculated with cells from single bacterial colony. The culture was incubated overnight at 37°C with shaking (220 rpm). This starter culture was diluted 1:100 in required volume with LB broth and incubated at 37°C. When the OD_{600} reached 0.55, the culture was placed on ice for 15 min. The cells were harvested using 50 mL Falcon tube by centrifugation at 4700 x *g* for 10 min at 4°C and resuspended in 1/3 volume of ice-cold Inoue transformation buffer (55 mM $MnCl_2.4H_2O$, 15 mM $CaCl_2.2H_2O$, 250 mM KCl, 10 mM PIPES pH 6.7). The cells were harvested again at 4700 x *g* for 10 min at 4°C. The pellet was re-suspended in Inoue transformation buffer at 1/12.5 of the original volume of cells. DMSO was added to 0.6% [v/v], swirled and left on ice for 10 min. Aliquots of 150 µL of the suspension were dispensed into chilled, sterile Eppendorf tubes. The competent cells were snap-frozen by placing the tubes on dry ice. The tubes were stored at -80°C until time of use.

2.12 Heat shock transformation of cloned gene

An aliquot (one µL) of ligated reaction or 50 ng of a plasmid construct was added to 50 µL of competent cells, mixed gently, and kept on ice for 30 min. The cells were "heat shocked" for 45 s at 42°C, and then placed back on ice for 2 min. An aliquot (475 µL) of LB broth was added to the cells and incubated at 37°C for 1 h at 225 rpm. Aliquots of 50 µL, 100 µL, 250 µL, were spread onto LB agar plates containing appropriate antibiotic and incubated at 37°C overnight. Alongside two reactions were set up as above using pUC19 as positive control and water as negative control. Single colonies from transformants (routinely 10) were used to inoculate 10 mL aliquots of LB broth and incubated overnight. Plasmid DNA was isolated from overnight culture using Wizard® Plus SV Minipreps DNA according to manufacturer's instructions (Promega).

2.13 Enzymatic and sequencing confirmation

One µg of the isolated plasmid DNA was subjected to double digest with appropriate restriction enzyme, and then run on a 1% [w/v] agarose gel. The restriction enzymes are stated in the corresponding figure legends. Inserts of constructed plasmid were sequenced by Beckman Coulter Genomics using universal primers (T7P and T7TERM) for pET16b-based plasmids; Bec-F and Bec-R primers for pAU3-5 and pSET152 based plasmids; and gl-atrA-F and gl-atrA-R for pKC1139-based plasmids (see Table S2).

2.14 Protein Methods

2.14.1 Overproduction and purification of oligohistidine-tagged protein

Recombinant *S. coelicolor* AtrA, AtrA-3xFlag tagged at C-terminus, ActR, *Streptomyces globisporus* AtrA and *Escherichia coli* TetR were overproduced and purified from cultures of *E. coli* BL21 (DE3) cells containing pRA001, pET-atrA3x, pET16b-ActR, pET16b-atrA-gl, pET16b-TetR, respectively. Typically, a 10 mL overnight culture was diluted 1:100 in 400 mL of LB broth containing 50 µg/mL of ampicillin. Then the culture was incubated at 37°C with shaking (225 rpm). Expression of the chromosomally-encoded T7 RNA polymerase gene was induced by adding IPTG (Isopropyl β-D-1-thiogalactopyranoside) to a final concentration of 1 mM, when the cultures reached an OD_{600} of 0.6. Incubation was continued for a further 3 h to allow over-expression of the recombinant gene encoded in the pET16b-based construct. After 3 h, cells were harvested by centrifugation at 4000 x *g* for 30 min at 4°C (Sorvall RC-5B, SLA-3000 rotor). The pellet was washed with 20 mL of ice-cooled lysis buffer (20 mM Tris-HCl [pH 7.6], 0.5 M NaCl, 50 mM Imidazole) and centrifuged at 4000 x *g* for 30 min at 4°C. Cell pellet was re-suspended in 30 mL ice-cold lysis buffer (20 mM Tris-HCl pH 7.6; 0.5 M NaCl; 50 mM Imidazole) with appropriate amount of cOmplete Mini EDTA-free protease inhibitor cocktail tablet (1 tablet/ 10 mL) (Roche Diagnostics). Pre-cooled cell disruptor with a one head setting at 15000 psi was used to lyse the cells. Lysate was cleared by ultra-centrifugation at 28000 x *g* at 4°C for 35 min (Optima™ L-80 XP Beckman Coulter ultracentrifuge), and then passed through a 0.45 µm filter (Sartorius stedim) before loading onto column.

In order to purify the protein, 1 mL HisTrap HP column (GE Healthcare Life Sciences) was washed with 5 mL dH_2O charged using 0.1 M $NiCl_2$ and washed with 5 mL dH_2O again. The sample lysate applied to the column at flow rate of 0.7 mL/min using a peristaltic pump (Econo Pump model EP-1 Bio Rad). Purification was performed using an AKTA Explorer FPLC system. The protein was eluted from column using gradient of elution buffer (0 to 100%) (20 mM Tris-HCl [pH 7.6], 0.5 M NaCl, 1 M Imidazole) and collected 1.5 mL fractions at flow rate of 1 mL/ min. Peak fraction were pooled into dialysis tube (Spectra/Por® Membrane Dialysis Spectrum, 3500 MWCO) and dialysed twice for 3 h at 4°C against 500 mL dialysis buffer (50 mM Tris-HCl [pH 7.6] 300 mM NaCl, 1 mM DTT, 1 mM EDTA, 40% [v/v] glycerol). Sample was removed from dialysis tube and then aliquots of 50 µL were stored at -80°C until required.

2.14.2 Analysis of purified oligohistidine-tagged protein

SDS-polyacrylamide gel electrophoresis (PAGE) was performed using 15% [w/v] polyacrylamide resolving gel with 5% [w/v] stacking gel. The 15% resolving gel contained 15% [w/v] (29:1) acrylamide: *bis*-acrylamide solution; 380 mM Tris-HCl (pH 8.8), 0.1% [w/v] SDS; 0.1% [w/v] APS; 0.1% [v/v] TEMED. The 5% stacking gel contained 5% [w/v] (29:1) acrylamide: *bis*-acrylamide solution; 126 mM Tris-HCl pH (6.8), 0.1% [w/v] SDS; 0.1% [w/v] APS; 0.1% [v/v] TEMED. After adding an equal volume of 2 x sample buffer (Laemmli, Sigma-Aldrich), samples were incubated at 99°C for 5 min and then centrifuged briefly (13,000 x *g*), before being loaded using a 200 µL fine pipette tip. Samples were run alongside PageRuler Unstained Protein Ladder (Fermentas). Gel electrophoresis was performed in 1 x Tris-Glycine buffer (25 mM Tris-HCl [pH8.3], 190 mM glycine; 0.1% [w/v] SDS) for 2 h at a constant 12 W. The gel was removed from between the glass plates and stained in 20-50 mL of Coomassie Blue solution (1 tablet of PhastGEl® Blue R was dissolved in 200 mL dH_2O after 200 mL methanol and 40 mL glacial acetic acid were added) for 3 h with shaking. Excess stain was removed by rinsing the gel in tap water and the gel incubated in destain solution (50% [v/v] methanol, 10% [v/v] glacial acetic acid) for a further 2 to 12 h, destain solution was

replaced every 2 h. Protein bands were visualized by illuminating from underneath with white light using GeneGenius UV transilluminator (Syngene) and photographed.

2.14.3 Protein Quantification

Protein concentration was determined by running serial dilution of protein samples on SDS-PAGE alongside dilution of prepared Bovine serum albumin (BSA) (Sigma) standards and comparing the intensity after staining with Coomassie blue. The protein concentrations were quantified using AIDA software (v4.22.034, Raytest GambH).

2.15 Electrophoresis Mobility Shift Assays (EMSA)

Promoter regions for EMSA were amplified and purified as described in Sections 2.9.1 and 2.9.2. To visualise the promoter regions, the 5' end of each of the reverse primers (see Table S2) was labelled with 6-FAM™ (fluorescein amidite) during its synthesis (MWG Eurofins). Each 20 µL of a reaction contained 5 nM of probe, 2 µg of Salmon Sperm DNA (Invitrogen) and a 10-fold dilution series of protein in the range 1.2 nM to 5 µM. The reaction buffer was 1 x TGEK buffer (10 mM Tris-HCl [pH 8], 10% [v/v] glycerol, 0.1 mM EDTA, 50 mM KCl). Control reactions had all the compounds except protein. The binding reactions were allowed to proceed at room temperature for 20 min. Then analysed by running in a 4% [w/v] 37.5:1) acrylamide: *bis*-acrylamide gel in 1 x TGED buffer (50 mM Tris-HCl [pH 8.5], 0.4 M glycine, 2 mM EDTA). Electrophoresis was performed in Bio-Rad Protein II XI tank at 120 V for 40 min in 1 x TGED buffer with a water-cooled central core. The protein-DNA complex were visualized by using Fujifilm FLA-5000 imaging analyser system, an excitation laser wavelength of 473 nm was used.

2.16 Reintroducing *atrA* into *S. coelicolor*

2.16.1 Plasmids constructed as part of this study

Plasmid pSET-atrA3x is a pSET152 derivative which contains the coding region of *S. coelicolor atrA* fused with 3 x FLAG tag™ at C-terminus. Plasmid pAU-atrA3x is a pAU3-45 derivative which contains *atrA* coding region fused with 3 x FLAG tag™ at C-

terminus. Plasmid pAU-atrA is a pAU3-45 derivative which contains the coding region of *atrA*. Plasmid pAU-3xatrA is a pAU3-45 derivative which contains *atrA* gene fused with 3 x FLAG tag™ at N-terminus. The coding sequence of the 3 x FLAG tag™ was codon optimised for streptomycetes (ATGGACTACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGACG ACGACAAGTGA). pSET-atrA3x, pAU-atrA3x, pAU-atrA, and pAU-3xatrA were constructed by cloning 1356-bp, 1356-bp, 1270-bp, and 1392-bp DNA fragment into the BamHI and EcoRI sites, respectively (see Figure 5.1, panel A).

A 1456bp DNA fragment, including the coding region of *atrA* (891) fused with 3 x FLAG tag™ (72 bp) at C-terminus in addition to 332 bp upstream of the gene and 49 bp downstream of the gene flanking with 56 bp of pUC18 at each side, was synthesized at Life Technologies. A 1270-bp DNA fragment including *atrA* gene was amplified using atrAF2 and atrAEcoRI-R2 primers. The forward primer annealed upstream of the BamHI restriction site in pAU-atrA3x plasmid, which was used as a template. While the reverse primer annealed to the end of *atrA* gene adding a stop codon and an EcoRI site to the DNA fragment (Table S2). A 1392-bp DNA fragment including the 894-bp *atrA* coding region fused with 3 x FLAG tag™ at N-terminus was obtained as a gift from Professor Colin Smith (University of Surrey). Each recombinant has two cloning sites BamHI site at the N-terminus and EcoRI site at the C-terminus. The cloned *atrA* was under it is own promoter in all constructs. After gel electrophoresis the fragments were purified from excised gel using a QIAquick® gel extraction kit. All DNA fragments and pAU3-45 plasmid proceeded to double digested with BamHI and EcoRI using conditions recommend by the vendor of these enzymes (New England Biolabs). Ligation, transformation, enzymatic and sequencing confirmation were done as described previously under Section 2.13.

2.16.2 Conjugation

Conjugation was performed using an established method (Gust *et al.*, 2003). Competent cells of *E. coli* ET12567 (pUZ8002) were grown in LB at 37°C. Kanamycin and chloramphenicol were added to maintain selection for pUZ8002 and the *dam*

mutation, respectively. The plasmid of interest was introduced into the strain by transformation as described in Section 2.12. Transformants were selected by spreading the cells on the LB agar plates containing apramycin (50 µg/mL). A single transformed colony was used to inoculate 10 mL of LB broth containing apramycin, chloramphenicol, and kanamycin, and then incubated overnight. An aliquot (100 µL) of overnight culture was inoculated into 10 mL LB broth containing appropriate antibiotics and incubated for 8 h to an OD_{600} of 0.4. The cells were washed twice with LB broth to remove antibiotics, and re-suspended in 1 mL of LB broth. 1 x 10^6 *S. coelicolor* spores in 0.5 mL 2 x YT broth were "heat shocked" at 50°C for 10 min and place in an ice-water batch.

A mixture of 0.5 mL *E. coli* cells and 0.5 mL of *S. coelicolor* spores was pelleted and re-suspended in 50 µL residual liquid. A serial dilution of the mixture was made in sterile dH_2O from 10^{-1} to 10^{-4}. 100 µL from each dilution was plated out and spread over SFM plate containing 10 mM $MgCl_2$, and incubated at 30°C for 18 h. Transconjugants were selected by overlaying the plates with 1 ml sterilized dH_2O containing 1.25 mg apramycin (for M145 strains) or 1.25 mg thiostrepton (for L645 strains) as selective marker in addition to 0.5 mg of nalidixic acid to kill *E. coli* cells. Spore stocks were prepared from each single transconjugants colony (see Section 2.4). The control was also performed as above, but without donor bacteria.

2.16.3 Analysis of actinorhodin (ACT) production

Ten mL cultures (YEME) of the strains to be analysed were set up in 50 mL Falcon tubes with appropriate selection. All tubes were incubated for 5 d at 30°C with shaking (220 rpm). Mycelia were collected and weighted. 200 µg of mycelium in 20 µL of 25% [v/v] glycerol were spotted onto R5 plates. The well sporulated strains were analysed by spotted appropriate numbers of spores onto R5 plates. All plates were incubated for 48 h before being photographed at this point and every 24 h for the next 120 h.

2.17 ACT extraction and EMSA assay

2.17.1 Preparation of crude extract

5 X 10^5 spores of *S. coelicolor* L646 (Towle, 2007), which overproduces ATC, were spread on TSA plates and incubated at 30°C for 6 d to obtain an even lawn. The agar was collected, submerged in 2.5 x volumes of water and incubated at room temperature for 1 h with shaking (200 rpm). The liquid, which containing ACT was collected, acetic acid added to lower the pH until the ACT changed colour from blue to red. The supernatant was then extracted with an equal volume of ethyl acetate (Fisher Scientific Ltd). The ethyl acetate extract was concentrated using a rotary evaporator (Rotavapor, RE Buchi), and then dried using a parallel evaporator (EZ-2 personal evaporator, Genevac). The dried crude extract was dissolved into methanol at high concentration. Crude activity was checked as described in Section 2.15 with exception of using fixed concentration of AtrA-c (150 nM) and 5 nM of *act*II-ORF region 2 probe as substrate. To provide controls, crude extract was similarly prepared from M511 (Floriano & Bibb, 1996), which is *Δact*II-ORF4, and M1146 (Gomez-Escribano & Bibb, 2011), which is *Δact*, *Δred*, *Δcpk*, and *Δcda*.

2.17.2 Thin layer chromatography (TLC) purification of actinorhodin (ACT)

Aliquots of crude extract dissolved in methanol were spotted on a pencil line drown about a 1 cm from the bottom of TLC silica gel plate (Merck Millipore) using a capillary glass tube. The plate was moved into a TLC tank and immersed in toluene: glacial acetic acid (4:1 [v/v]) just under the line where the crude extract loaded. After the solvent front reaches around 1/3 of the plate, the plate was removed and the solvent front was marked with a pencil line. After the plate was dried the upper red band with R_f 0.5 was cut out and eluted with ethyl acetate. The EtOAc extract was dried and weighted. The resulting material was dissolved in methanol and subjected to Liquid chromatography – Mass Spectrometry (LC-MS). 1 mL (1 mg) of extracted material was injected onto Phenomenex Luna C18(2) column (50 x 2 mm) and connected to an Agilent 1200 series HPLC with diode array detector and Bruker HCT Ultra Ion Trap Mass

Spec. Distance travelled by pigment was divided by distance travelled by solvent front to calculate R_f value. ACT activity was checked as same as crude extract (Section 2.17.1)

2.17.3 Preparative High performance liquid chromatography (HPLC)

Crude extract from L646 was fractionated by injecting 3 x 500 µL (5 mg) onto a Waters XBridge Prep C18 5 µm OBD column (19 x 100 mm). This column was connected to an Agilent 1260 series HPLC with Diode array detector and an Agilent 6100 series single quad mass spec. The column was washed with 95% MeOH. Components that bound to the column were eluted by a gradient of MeOH (5 to 95%) in 0.1% [v/v] formic acid (300 mL, 20 mL / min). At the end, the gradient was held at 95% [v/v] MeOH for 7 min. ACT eluted as a peak at 90% [v/v] MeOH. All peaks and the flow through were concentrated, dried, weighed and then dissolved in 100% [v/v] MeOH to a concentration of 9.5 mg/mL.

2.17.4 Analysis of ACT

Various concentrations of each fraction were used to test their activities (see Section 2.17.1). Fraction concentration and reaction conditions are stated in the corresponding figure legends. ActR (SCO5082), which is transcription regulator of *actA* (SCO5083) and TetR, which is transcription regulator of *tetA* were used as controls. Promoter region for *actA* and *tetA* were amplified and purified as described in Section 2.9.1 and 2.9.2.

Chapter 3

3 The production and initial interpretation of a nucleotide-resolution transcription map for *Streptomyces coelicolor* showing the positions of sites of transcription initiation and RNA processing and degradation.

3.1 Introduction

As outlined in the General Introduction, numerous genes and growth conditions that influence secondary metabolite production and morphological development have been identified for *S. coelicolor* and for some the underlying changes in gene expression have been determined using microarrays and to a lesser extent proteomic approaches (for reviews, see (van Wezel & McDowall, 2011, Liu *et al.*, 2013). Despite this considerable effort, there is no coherent understanding of the pathways that control the flow of genetic information at the level of the whole cell. A limiting factor is the scarcity of experimental information on the organisation and regulation of the transcriptional units that encode proteins and RNA components of the translation and regulatory machinery. Elements such as promoters, transcription start sites and untranslated regions have not been identified experimentally on a genome-wide scale. Other key aspects of gene expression for which virtually no information was available for *S. coelicolor*, and indeed any other bacterial species, were the sites of RNA cleavage that facilitate rapid mRNA turnover (Carpousis *et al.*, 2009), a key process that ensures translation follows programs of transcription, and generate the RNA components of the translation machinery (Deutscher, 2009, Hartmann *et al.*, 2009). Currently, genome-wide knowledge of operons and *cis*-regulatory elements is largely inferred from a promoter-centred analysis of modules of co-expressed genes that were

identified from a large compendium of transcriptome data (Castro-Melchor *et al.*, 2010).

The specific objective of this chapter was to develop a pipeline for the generation of maps of transcription units that detailed the positions of transcription start sites. A major driver for developing this pipeline was to provide a platform for interpreting the functional significance of the binding of regulators of transcription to specific sites. As indicated earlier, of particular interest to this project is the function of AtrA, a TetR family member that is highly conserved in *Streptomyces* spp. and regulates the production of actinorhodin in *S. coelicolor* (Uguru *et al.,* 2005). It was also appreciated that the generation of detailed transcription maps for *S. coelicolor* would provide new insight beyond the transcription initiation and its regulation. The pipeline described here combined two RNA sequencing approaches, one that identifies and "differentiates" sites of transcription initiation and endonucleolytic cleavage (dRNA-seq) and another that provides a "global" view of transcript abundance and the boundaries of transcription (gRNA-seq) (Lin *et al*., 2013). The transcriptome of *Escherichia coli* was mapped in parallel by David Romero-Alves, another member of the laboratory. The inclusion of *E. coli* validated the approach, as many features identified previously using traditional gene-specific methods, such as northern blotting and primer extension, were identified by RNA sequencing. As detailed elsewhere (Romero *et al.*, 2014), it also provided new insights into the regulation of gene expression in *E. coli* and revealed the limitations of previous global analyses, for example, in the mapping of transcription start sites (Cho *et al.*, 2009).

3.2 Results

3.2.1 Overview of RNA-seq approaches

Sites of transcription initiation and RNA cleavage were mapped for *S. coelicolor* M145 grown in YEME broth until the mycelia became visibly pigmented. This point was chosen as it is known to coincide with exit from exponential growth and the onset of secondary metabolism (Bibb, 2005, van Wezel & McDowall, 2011). A detailed growth curve was not produced as the primary objective was to develop a methodological pipeline and not describe a specific phase of growth. M145 was chosen for this study as it has been, and continues to be, used extensively worldwide for biochemical and genetic studies (Chater *et al.*, 2002, Baba *et al.*, 2006). Total RNA was isolated from the single point in growth and enriched for mRNA by depleting 16S and 23S rRNA. To differentiate 5'-triphosphorylated ends generated by transcription from 5'-monophosphorylated ends produced by RNA processing or degradation, an aliquot of the enriched mRNA was incubated with tobacco acid pyrophosphatase (Loria *et al.*, 2008, Loria *et al.*, 2007, Breter & Rhoads, 1979), which leaves a monophosphate on 5' ends that were originally triphosphorylated. A second aliquot was incubated under the same conditions, but without TAP. 5' end fragments were then cloned and sequenced via a strategy that required the ligation of an adapter to 5'-monophosphorylated ends (Lin *et al*., 2013). A significant increase in the number of sequencing reads at a specific position following TAP treatment provided an identifier of a transcription start site (TSS). It should be noted that after the addition of the 5' adaptor the RNA was fragmented to allow good coverage of the 5' ends of large transcripts and intermediates. A third aliquot of each RNA sample was analysed using FRT-seq, an amplification-free form of strand-specific global RNA-seq (Lin *et al.,* 2013, Mamanova *et al.,* 2010). This allowed the transcripts to be mapped along their entire length. Importantly, the RNA-seq approaches used here should not be affected unduly by high GC content (Lin *et al.,* 2013, Mamanova *et al.,* 2010), a feature of streptomycete genomes. In brief, the gRNA-seq approach avoided the use of PCR, which introduces bias with GC-rich templates (McDowell *et al.*, 1998), while the dRNA-seq approach did not use Terminator™ 5' monophosphate-dependent exonuclease (TEX), which does

not degrade *Streptomyces* RNA efficiently, presumably because of the high prevalence of stable secondary structures. In the original and most widely used dRNA-seq approach to date, TSSs are identified by their resistance to treatment with TEX (Sharma *et al.*, 2010).

3.2.2 Identification of transcription start sites

To classify the 5' ends identified by dRNA-seq, reads obtained before and after TAP treatment were analysed using M-A (ratio-intensity) scatterplots. For *S. coelicolor* (Figure 3.1 A) and *E. coli* (Figure 3.1 B), two populations of values were found, as described previously for *Propionibacterium acnes* (Lin *et al*., 2013). The largest population corresponds to sites of processing and degradation and centres on a value of M close to 0, while the smaller population associated with higher M values corresponds to TSSs. The TSSs for *S. coelicolor* were on average associated with higher M values and lower A values (Figure 3.1 A), which may reflect a lower proportion of the 5' ends of primary transcripts being monophosphorylated *in vivo*. *E. coli* has an RNA pyrophosphohydrolase (Celesnik *et al.*, 2007, Deana *et al.*, 2008). The situation for *S. coelicolor* is not yet known. For both organisms, nucleotide positions with M values above what was judged to be the upper boundary of the central cone of the population corresponding to processing and degradation sites (red line, Figure 3.1 A and B) were designated positions of possible TSSs. The positioning of the upper boundary was based on knowledge of the shape of cones produced by plotting two biological replicates of samples (before TAP treatment) against each other (data not shown). Positions within 8 nt of each other were assigned to the same TSS, as it is known that many promoters initiate transcription from a cluster of nucleotide positions (Salgado *et al.*, 2006). These sites were then mapped against leading edges of transcription, which were determined independently via manual inspection of the global RNA-seq data, as described previously (Lin *et al*., 2013).

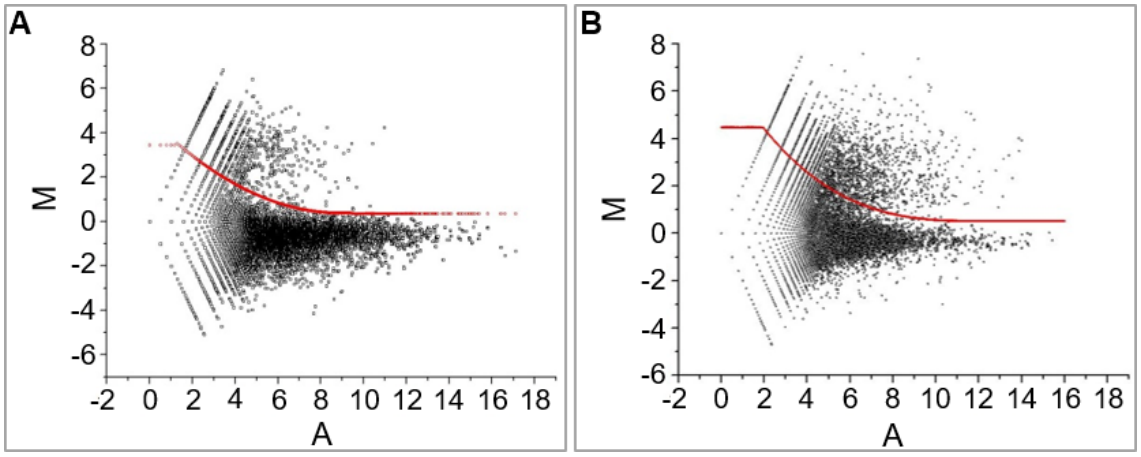

Figure 3.1: M-A scatterplots of values from the differential RNA-seq analysis. Panels A and **B** represent data for *S. coelicolor* M145 and comparator *E. coli* BW25113 (Romero *et al.*, 2014). The M values correspond to Log_2 (plus/minus) and A values to (Log_2 plus + Log_2 minus)/2, where minus and plus refer to the number of reads before and after treatment with TAP. The numbers correspond to individual genome positions, not genes. For further details, see Section 2.7.3. In each panel the red line represents the upper boundary of the population of values corresponding to sites of processing and degradation (see main text). The upper boundaries were placed manually to enclose the majority of the lower population, while taking into consideration the spread of M values scattered around 0. The boundaries were then described by the polynomial equations. These were $M = 0.054A^2 – 0.96A + 4.68$ and $M= – 0.003A^3 + 0.13A^2 – 1.57A + 7.08$ for *S. coelicolor* M145 and *E. coli* BW25113, respecively.

For *S. coelicolor,* 1147 sites associated with leading edges of transcription as well as being enriched following TAP treatment (Class I) were identified. *E. coli* BW25113 data were included to allow comparison. The number of class I sites was higher for *S. coelicolor* than *E. coli* and probably reflects the larger gene content (7.8 vs 4.3k genes). TSSs were identified for all types of functional RNA: mono- and poly-cistronic mRNAs, ribosomal RNAs, transfer RNAs and ubiquitous small RNAs (for examples, see below). As described previously for *P. acnes* (Lin *et al.,* 2013) , 217 *S. coelicolor* sites were identified that were enriched following TAP treatment, but not associated with leading edges of transcription (Class II), and 451 sites were associated with leading edges, but not enriched (Class III). In comparison, 311 TSSs form class II and 1554 TSSs from class III have been identified in *E. coli* (Romero *et al.*, 2014). Although, Class I sites were assigned with the most confidence, being based on two criteria, all three classes

contained TSSs that have been identified previously by others and recorded in RegTransBase (Cipriano *et al.*, 2013). Examples of different classes of TSS are shown in Figure 3.2. This information has been included in the annotation of these sites (Table S2). The probable origin of the different classes of TSSs, which has been described previously (Lin *et al*., 2013), is outlined in the Discussion section.

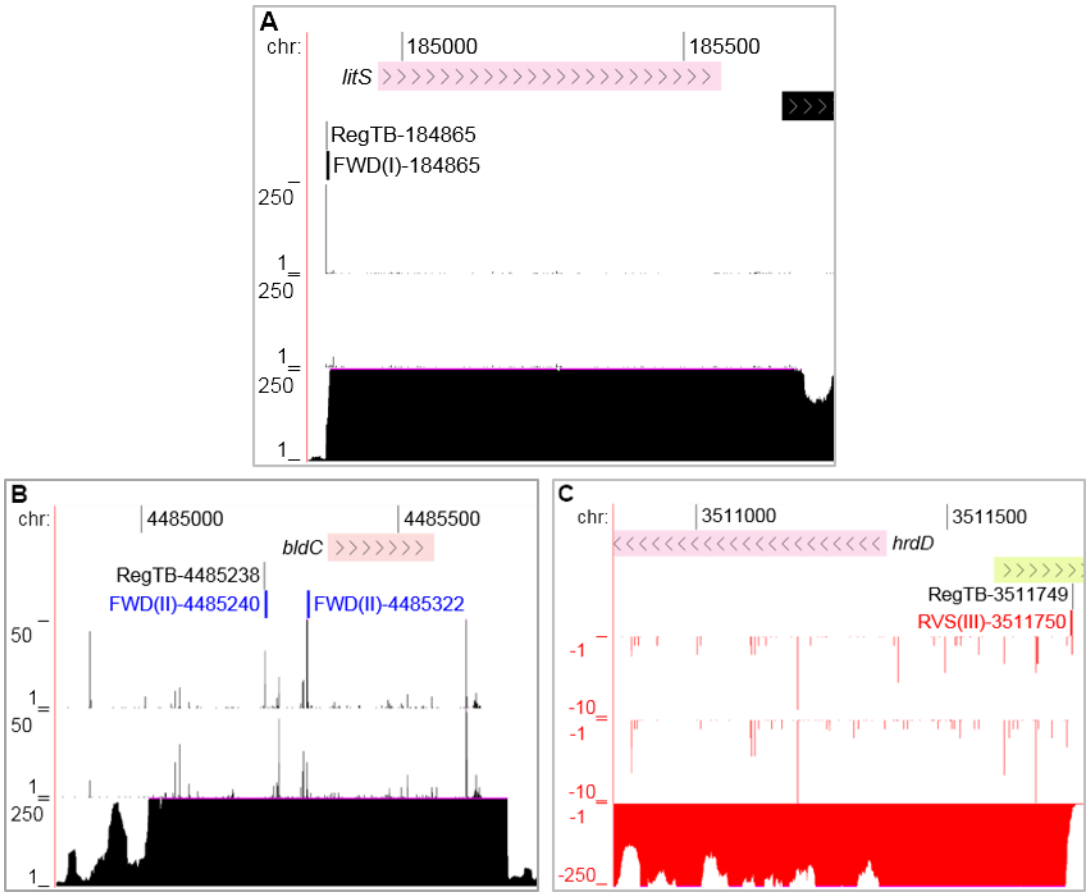

Figure 3.2: Examples of different classes of TSSs. Panels A, B and **C** correspond to previously identified TSS of *litS* (Takano *et al.*, 2005b), *bldC* (Hunt *et al.*, 2005) and *hrdD* (Paget *et al.*, 2001), respectively. The panels are modified screenshots from the UCSC Microbial Genome Browser (Schneider *et al.*, 2006, Chan *et al.*, 2012). The tracks depict from top to bottom show the genome positions, gene locations, position of RegTransBase TSSs, position of TSSs classes, differential RNA-seq reads (in the presence and absence of TAP treatment) and gRNA-seq reads. The numbers at the left of the RNA-seq tracks indicate the scale of the sequencing reads. dRNA-seq and gRNA-seq data for the forward strand are coloured black and have positive values, while the reverse strand is coloured red and has negative values. The positions of TSSs identified by the analysis of M-A scatterplots (Figure 3.1), and the number of times each position on the corresponding strand was sequenced following fragmentation of the transcriptome (gRNA-seq). TSSs are marked by short vertical lines that are labelled to

indicate the forward (FWD) and reverse (RVS) strand of DNA. The class are colour coded. Class I is in black, class II in blue and class III in red.

3.2.3 Leaderless mRNAs

By mapping TSSs onto the annotated genomes, 264 mRNAs that cannot be translated via the canonical Shine-Dalgarno (SD) interaction (Shine & Dalgarno, 1974, Shine & Dalgarno, 1975) because they either lack or have a short 5' leader (>10 nt) were identified (Table S3). These mRNA species are classified here as being leaderless. A prevalence of leaderless mRNAs (lmRNAs) in *S. coelicolor* has been reported previously (Vockenhuber *et al.*, 2011). The global RNA-seq data for some of the many *S. coelicolor* lmRNAs are shown in Figure 3.3. This includes data for a homologue of WhiH, which also has an lmRNA (Ryding *et al.*, 1998). WhiH and other members of the GntR family have roles in controlling morphological development and secondary metabolism (Hoskisson *et al.*, 2006, Persson *et al.*, 2013, Hillerich & Westpheling, 2006). The possibility that the function of WhiH and other regulators is dependent on the leaderless status of the mRNA has not been investigated. It should be noted that here and elsewhere in this report the range of the global RNA-seq reads was restricted in genome-browser views to make it easier to determine the boundaries of transcription units. This can result in a block-like appearance. For the reverse strand, the RNA-seq data was given negative values and plotted in red instead of black. TSSs are represented by vertical lines and labelled according to strand, class and genome position.

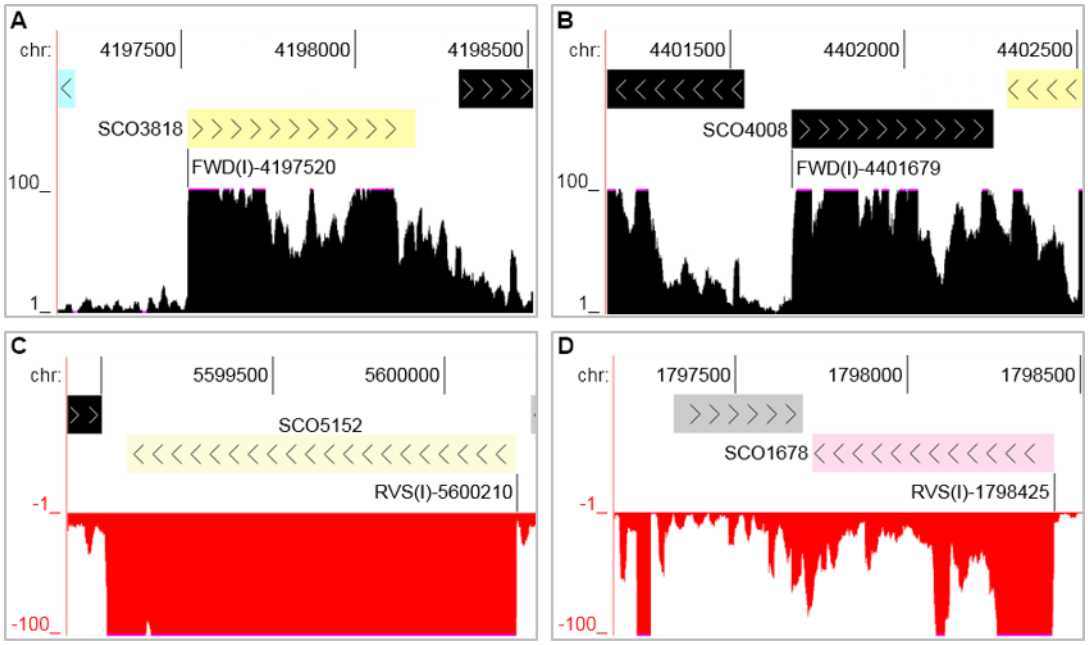

Figure 3.3: Examples of leaderless mRNAs. (A), **(B)**, **(C)**, and **(D)** correspond to SCO3818, SCO4008, SCO5152 and an unnamed paralogue (SCO1678) of *whiH*. The panels are modified screenshots from the UCSC Microbial Genome Browser (Schneider *et al.*, 2006, Chan *et al*., 2012). In each panel the tracks depict from top to bottom, the genome position, location of annotated genes, positions of TSSs and gRNA-seq data. Labelling is as in Figure 3.2.

3.2.4 Processing and maturation of the stable RNAs

Recently, it has been shown that *E. coli* lmRNAs, or at least those generated by 5' processing under conditions of stress, are translated by specialised ribosomes from which the last 43 nt of the 3' end of 16S rRNA have been removed endonucleolytically by MazF, the toxic component of a toxin-antitoxin (TA) system (Vesper *et al.*, 2011). This region of 16S rRNA contains the anti-SD sequence and the binding site of S1 (Lauber *et al.*, 2012, Shine & Dalgarno, 1974), a protein that augments the SD interaction (Sorensen *et al.*, 1998). Interestingly, while cleavage at the -43 site (numbered relative to the 3' end of mature 16S rRNA) was identified for *E. coli* by the parallel analysis (Romero *et al*., 2014). Processing within the equivalent region in *S. coelicolor* 16S rRNA was not detected; however, this was not unexpected as it has been shown for streptomycetes that specialised ribosomes capable of translating lmRNA can be produced by modification of the 16S rRNA in the presence of

kasugamycin (ksg), which disrupting h2, h26, and h27 loops. Subsequently, stable ribosomal particles (61S) lacking six proteins of the small subunit (S1, S2, S6, S12, S18, and S21) will form that selectively translate lmRNAs (Kaberdina *et al.*, 2009, Poot *et al.*, 1996).

Despite the central role of ribosomes in translation, little is known about the processing and degradation of its RNA in *S. coelicolor*. However, the study of *E. coli* and *B. subtilis* has revealed that mature ribosomal RNAs are produced via a series of nucleolytic steps involving several ribonucleases and that rRNA can be degraded in response to aberrant assembly of the ribosome or cellular stress (Deutscher, 2009). The parallel RNA-seq analysis detected most of the known endonucleolytic processing sites for *E. coli* (Romero *et al*., 2014). It was initially considered that many of the processing sites might not be detected given the transitory nature of the corresponding intermediates and the fact that the mRNA had been enriched. Analysis of the *S. coelicolor* data revealed evidence of process at the 5′ ends of mature 16S, 23S and 5S rRNA, and at positions +3 and +52 relative to the mature 3′ end of 16S rRNA (Figure 3.4, panel A). The +52 site is within a segment complementary to another at the 5′ end of 16S rRNA. Thus, this site may correspond to cleavage by RNase III (AbsB, SCO5569), which is specific for double-stranded regions and has been shown to process rRNA in many bacteria (Nicholson, 2003) including *S. coelicolor* (Price *et al.*, 1999). However, the product of the staggered cut within the complementary region at the 5′ end of 16S rRNA was not detected. This differs from *E. coli* and might reflect the closer coordination of subsequent 5′ processing steps, which could include 5′ to 3′ exonucleolytic processing by RNase J (SCO5745), an ribonuclease with dual endonucleolytic and 5′ to 3′ exonucleolytic activity (Mathy *et al.*, 2007, Bralley *et al.*, 2014) that is absent in *E. coli*, but present in *B. subtilis*. One of the functions of RNase J in *B. subtilis* is to generate the 5′ end of 16S mature rRNA (Britton *et al.*, 2007).

No obvious processing sites were detected at the 3′ end of mature 23S or 5S rRNA suggesting that, as described for *E. coli* (Gutgsell & Jain, 2012), the 3′ ends of these RNAs in *S. coelicolor* are generated primarily by 3′ exonucleolytic activity. A large

number of cleavage sites internal to the functional regions of the mature rRNAs of *S. coelicolor* were also detected. Cleavage at these internal sites is likely to be involved in the recycling of rRNA transcripts. The ribonucleases responsible for cleavage at these internal sites as well as those involved in processing the 5′ and 3′ ends can now be determined by analysing knockout mutants using fragment-specific approaches, such as those used to analyse *B. subtilis* rRNA processing (Redko & Condon, 2010). However, the incorporation of dRNA-seq into such studies is recommended as it will provide a genome-wide view on the roles of *S. coelicolor* ribonucleases beyond rRNA processing.

Like rRNAs, tRNAs are produced with 5′ and 3′ segments that have to be removed in order for the molecule to become functional. 53 of the 65 tRNAs in *S. coelicolor* encode the 3′ CCA motif to which amino acids are attached, the others must have this motif added post-transcriptionally. Studies of tRNA processing in *E. coli* have led to a model in which the mature 5′ end is generated by the ubiquitous endonuclease RNase P, and the mature 3′ end via endonucleolytic cleavage a few nucleotides downstream followed by 3′ exonucleolytic trimming to the CCA motif. The maturation of the 3′ end can be mediated by tRNase Z (RNase BN), which has dual endo/3′ to 5′ exonucleolytic activity (Dutta *et al.*, 2012, Dutta & Deutscher, 2010), or by the combined action of RNase E and 3′ to 5′ exonucleases, mainly RNases PH and T (Hartmann *et al*., 2009). In contrast to this model for *E. coli*, most of the 12 CCA-encoding tRNAs in *S. coelicolor* were found to be cut within the CCA motif between the Cs. This includes the *bldA* (Leu) tRNA (Figure 3.4, panel B), which is required for morphological development and accompanying secondary metabolism (Lawlor *et al.*, 1987). The positions of cleavages within the CCA motif have been recorded (Table S4). It is possible that the *S. coelicolor* homologue of tRNA nucleotidyltransferase (SCO3896), which presumably adds CCA to tRNAs that are not transcribed with this motif (Cudny & Deutscher, 1986), may also be capable of recognising partial CCA ends and adding only the residues that are missing. There is evidence that at least some tRNA nucleotidyltransferases, including the *E. coli* enzyme (Reuven *et al.*, 1997), have the capability of repairing CCA (Betat *et al.*, 2010). Such an activity in *S. coelicolor* would mean that cleavages within 3′ CCA triplets would

not result in terminal inactivation of the tRNA. For several of the *S. coelicolor* tRNAs encoded without the CCA motif, processing within a few nucleotides downstream of the 3' end was identified (Table S4). *S. coelicolor* has homologues of both RNase E (SCO2599) and tRNase Z (SCO2547), which could cut on the 3' side of tRNAs. These cleavages presumably allow 3' to 5' exonucleolytic trimming of the tail prior to addition of the CCA by tRNA nucleotidyltransferase.

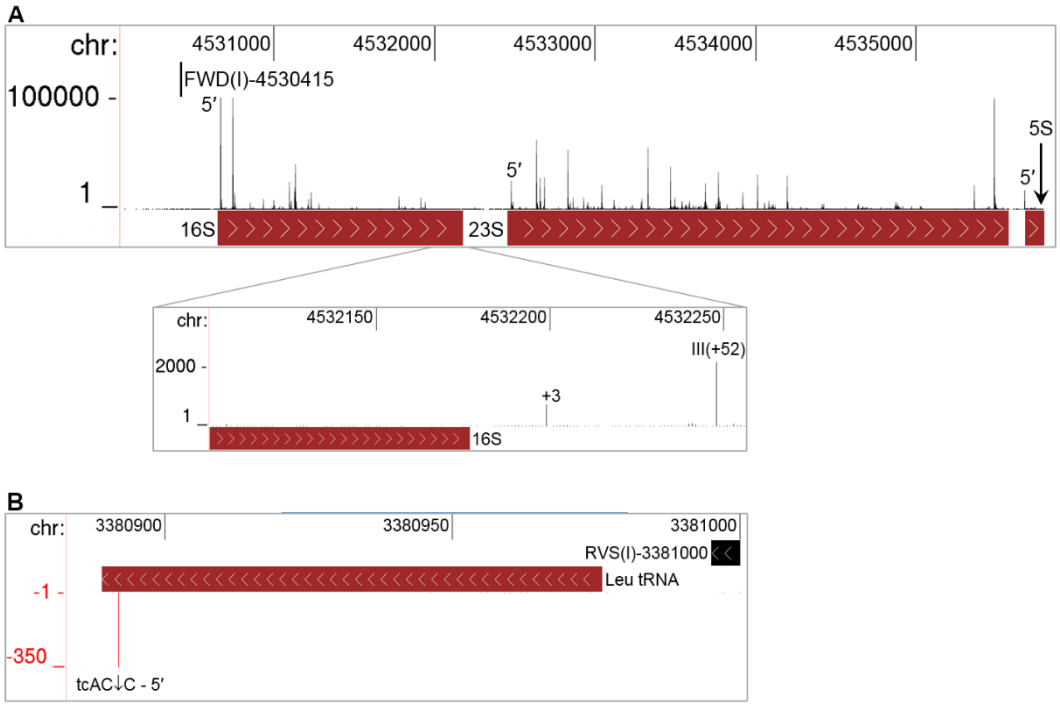

Figure 3.4: Maturation of stable RNAs. (**A**) Annotated view of the *rrnA* operon of *S. coelicolor*. The tracks from top to bottom show the genome position, differential RNA-seq reads in the absence of TAP treatment and the location of genes within the operon. The track containing the differential RNA-seq reads has been labelled to shown the position of cleavage sites referred to in the text and TSSs. As Figure 3.3, the labelling of the latter also indicates the strand and nucleotide position. (**B**) Processing within the CCA at 3' end of *S. coelicolor bldA* (Leu) tRNA. An arrow indicates the position of cleavage within the CCA motif.

3.2.5 The degradation and processing of mRNA

With regard to mRNA, the parallel analysis of *E. coli* detect many endonucleolytic sites known to be involved in both the degradation and processing of mRNA (Romero *et al*., 2014), thereby further validating the RNA-seq strategy. Numerous processing sites

were detected within *S. coelicolor* mRNA. This included sites upstream of the coding segment of *pnp* (Figure 3.5), which are cut by RNase III to initiate a mechanism that autoregulates the cellular level of PNPase activity (Gatewood *et al.*, 2011). Thus, although *S. coelicolor,* unlike *E. coli*, contains RNase J (SCO5745), a endoribonuclease with dual 5′ to 3′ exonuclease activity (Condon, 2010), the precise 5′ ends of the downstream products of endonucleolytic events can be detected. Moreover, the density of 5′ ends (number per transcribed kbp) is not significantly higher in *S. coelicolor* than in *E. coli* (data not shown) suggesting that 5′ to 3′ exonucleolytic decay does not dominate bulk mRNA degradation in the former, even though it contains a ribonuclease with 5′ to 3′ exonucleolytic activity.

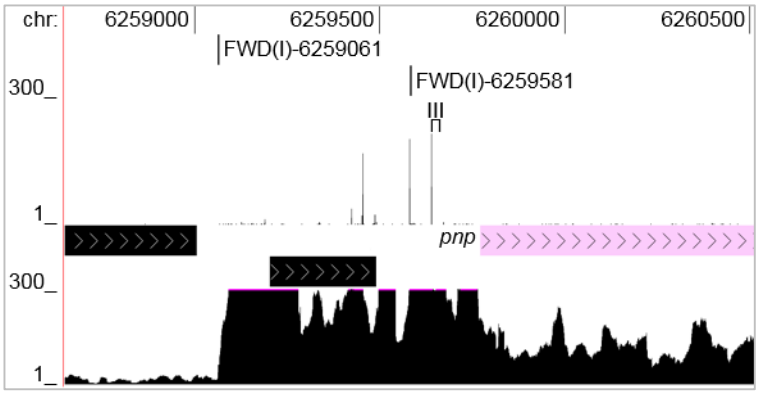

Figure 3.5: Cleavage sites within the mRNA of the *S. coelicolor pnp* (SCO5737) gene. The tracks from top to bottom show the genome positions, the positions of TSSs, forward strand differential RNA-seq reads (in the absence of TAP treatment), gene locations and gRNA-seq reads. TSSs and cleavage sites referred to specifically in the text are labelled. The RNase III site on the downstream side was detected using a lower range of reads. TSSs are labelled according to the strand, class and nucleotide position.

3.2.6 Identification of potential sRNAs

The mapping of the transcription data against the annotated genomes of *S. coelicolor* revealed a number of short transcripts of high abundance within intergenic region and did not extend into annotated gene, which from hereon will be referred to as small RNA (sRNA). A sRNA is a non protein-coding RNA molecule with a size range from fifty to five hundred nucleotides (Liang *et al.*, 2011). A proportion of these transcripts mapped to palindromic sequences, which are the signatures of intrinsic transcriptional

terminators (Peters *et al.*, 2011) and possibly only correspond to metastable decay intermediates. Short transcripts corresponding to 3'UTRs not included here in the list of potential regulatory RNAs. However, it should be noted that a number of reports indicate that some stable secondary structures in 3' UTRs do correspond to functional sRNAs, in at least *E. coli* (Gossringer & Hartmann, 2012). The remaining group of 83 small RNAs was found to contain all of the ubiquitous bacterial sRNAs (Figure 3.6): 6S RNA, tmRNA, and the RNA components of RNase P and the Signal Recognition Particle (Storz *et al.*, 2011). Moreover, subsequent analysis revealed that 51 of the 83 sRNAs had been identified by prior RNA-seq studies focussed on the small RNA component of *S. coelicolor* (Vockenhuber *et al.*, 2011, Moody *et al.*, 2013) or predicted by bio-computational approaches (Swiercz *et al.*, 2008, Panek *et al.*, 2008). This left 32 sRNAs that were previously undetected. Interestingly, 12 transcripts annotated previously as sRNAs (Vockenhuber *et al.*, 2011) were found to extend into protein-coding sequences (for examples, see Figure 3.7). The differences in annotation may simply reflect an increase in the sensitivity of detection and transcript coverage provided by the gRNA-seq approach adopted here. The comparison with the prior RNA-seq study of *S. coelicolor* has been summarised (Table S5).

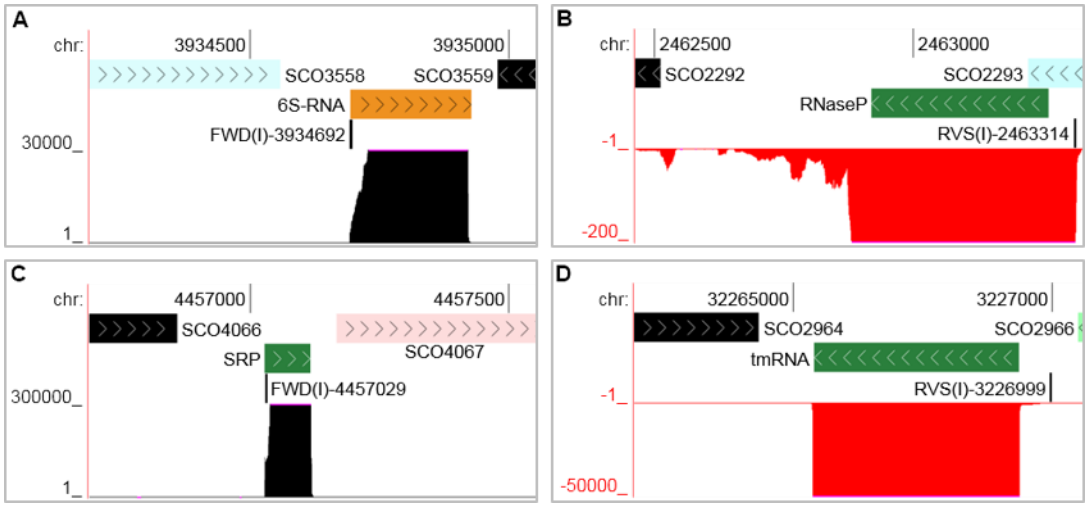

Figure 3.6: Ubiquitous sRNAs. Panels A to **D** show data for 6S RNA, the RNA components of RNase P and the signal recognition particle (SRP), and tmRNA, respectively. The tracks from top to bottom show the genome positions, gene locations, TSSs and gRNA-seq reads. The *S. coelicolor* homologue of 6S RNA that is shown is the one known to interact with the housekeeping sigma factor (Panek *et al.*,

2011). The other was poorly expressed in our sample. As in Figure 3.3, the gRNA-seq data for the forward strand is coloured black and has positive values, while the reverse strand is coloured red and has negative values.

To verify the ability of our approach to detect small RNAs for *S. coelicolor*, nine were selected randomly (indicated in Table S6) from a list of sRNAs that at the time had not been identified experimentally. These were then analysed using Northern blotting under stringent conditions that detect tmRNA and the RNA component of SRP, both of which are relatively abundant species. For three, scr2100 (Touzain *et al.*, 2008), scr2822(d+) and scr3871 (Xu *et al.,* 2012), signals were detected readily (Figure 3. 6, panels A and B). The sRNAs are labelled according to the nearest annotated gene, while the symbols in parenthesis indicate whether the RNA is downstream (d) or upstream (u) and transcribed from the same (+) or opposite strand (-). Moreover, the estimated sizes of the largest of the bands in each case corresponde well with the segment of highest abundance. For scr2822(d+) and scr3871 (Xu *et al.,* 2012), these segments did not coincide with the predicted transcription start sites indicating a role for processing in their maturation. The presence of scr2100 (Touzain *et al*., , 2008) and scr2822d(+) were confirmed by Moody *et al.* (2013). The list of 32 small RNAs that are listed here for *S. coelicolor* as being potentially novel include examples of riboswitches that appear to be actively reducing the levels of downstream transcripts and potential *cis*-encoded antisense RNAs (Figure 6, panel C).

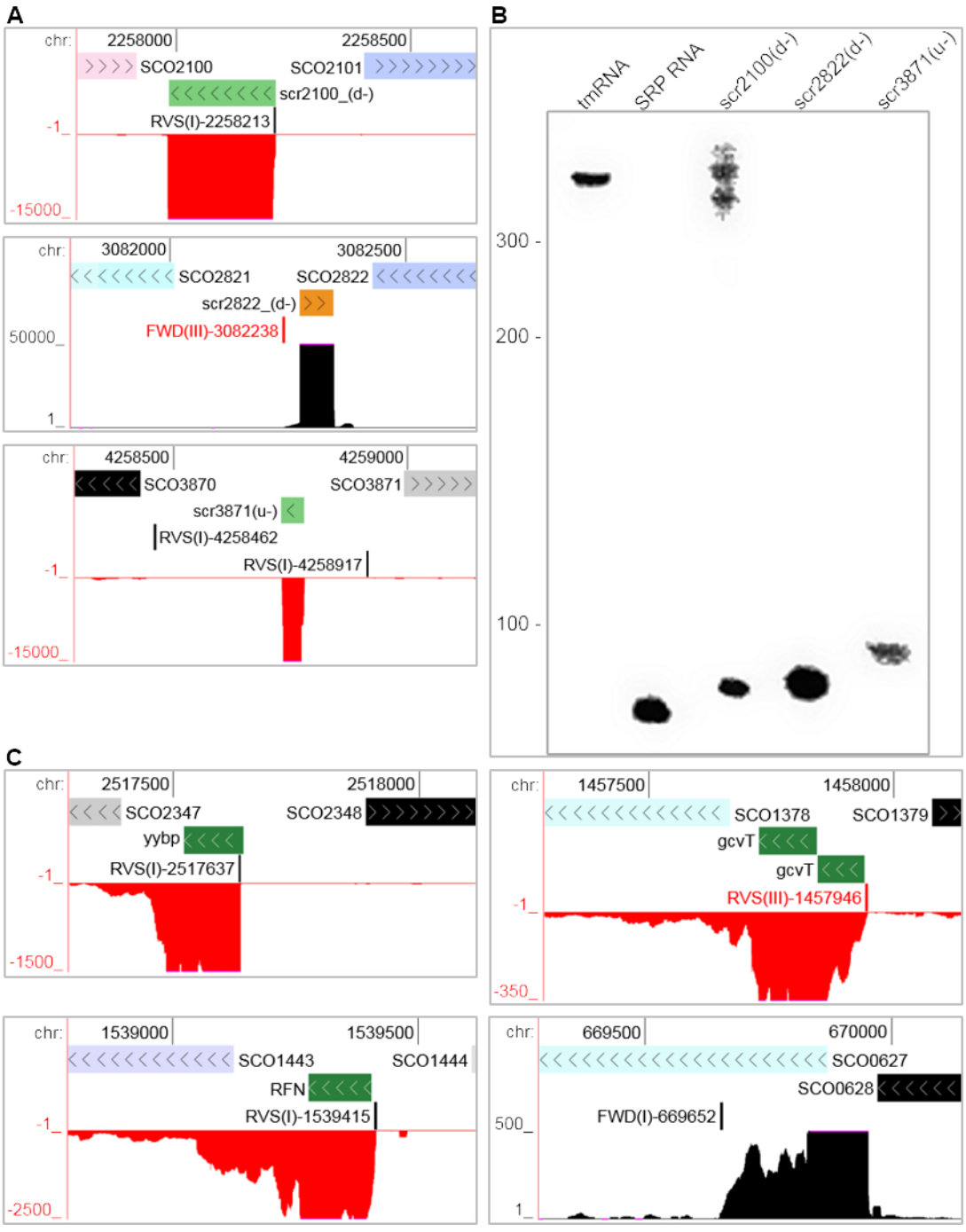

Figure 3.7: Northern blot analysis of *S. coelicolor* sRNAs. Labelling of sRNAs in parentheses indicates whether the sRNA is upstream (u) or downstream (d) of the nearest protein-coding gene and whether on the same (+) or opposite (-) strand. **Panel A** shows annotated views of sRNAs downstream of SCO2100 on the opposite strand, downstream of SCO2822 on the opposite strand and upstream of SCO3871 on the opposite strand. The tracks from top to bottom show the genome positions, gene locations, TSSs and gRNA-seq reads. **Panel B** shows northern blot analysis of sRNAs depicted in (A). The tmRNA and RNA component of the signal recognition particle were

probed to provide controls. The expected sizes of the most abundant species of these controls as judged from gRNA-seq data were ~400 and 80 nt, respectively. **Panel C** shows examples of active ribo-switching (attenuation) and a possible cis-encoded antisense RNA. The yybP element is reported to be pH responsive (Nechooshtan *et al.*, 2009) and is found in a large number of bacteria (Barrick *et al.*, 2004) including *E. coli* (Argaman *et al.*, 2001), SCO2347 encodes an integral membrane protein. The gcvT element binds the amino acid glycine (Mandal *et al.*, 2004), SCO1378 encodes glycine dehydrogenase. The RFN element (or FMN riboswitch) binds flavin mononucleotide (Serganov *et al.*, 2009), SCO1443 encodes riboflavin synthase. SCO0627, the target of the putative cis-encoded asRNA, encodes a putative ATP-utilising protein.

3.2.7 Transcription regulation and organisation

The previous RNA-seq analysis of *S. coelicolor* (Vockenhuber *et al*., 2011) identified 193 TSSs for mRNA, 98 were associated in our analysis with readily detectable transcription downstream, and of these 79 were identified as Class I sites (see Table S2). The finding that not all TSSs identified previously by RNA-seq (Vockenhuber *et al*., 2011) or indeed conventional mapping approaches, such as S1 mapping and primer extension (recorded in RegTransBase) (Cipriano *et al.,* 2013), were identified here is not surprising given the physiological and regulatory complexity of streptomycetes (van Wezel & McDowall, 2011, Chater, 2001, Strohl, 1992) and the analysis here of only one time point. There are also technical reasons that are discussed below. Nevertheless, the identification of 1147 Class I TSSs within the overall transcriptional landscape for *S. coelicolor* provides a much-improved platform for studying gene expression. For example, several obvious transcription units that started within the 5′ portion of regions annotated as being coding sequences were identified (Table S7) suggesting that the corresponding genes are actually shorter than previously thought. In support of this, it was found using the BlastP track of the UCSC browser, which displays the results from running BLASTP for all predicted proteins in the genome against those from other prokaryotic species, that many homologues were predicted to be shorter, with regard to their N-terminal ends, than their *S. coelicolor* counterpart (data not shown). Thus, as others have indicated, RNA-seq data can aid the accurate prediction of the 5′ ends of protein-coding regions (Sallet *et al.*, 2013).

In addition, our global RNA-seq data revealed many examples of operon structures that differ significantly from 'Arkin Lab' predictions (Price *et al.*, 2005). Fortunately, automated processes are being developed that allow transcriptional units identified by RNA-seq to be mapped onto genome sequences (Sallet *et al*., 2013). Clearly accurate annotation is required for gene expression and regulation to be modelled closely at the level of the whole cell (Karr *et al.*, 2012). In the context of gene expression, transcript levels in our *S. coelicolor* RNA were also analysed by Colin Smith and co-workers (Surrey) using high-density oligonucleotide-based microarrays and compared the level of hybridisation (as a $\log_2$ ratio of signal obtained for the RNA sample to signal obtained for chromosomal DNA) against the number of global RNA-seq reads obtained over the regions covered by the oligonucleotide probes. This revealed reasonable congruence (Figure 3.8) with a Pearson correlation coefficient of 0.63, supporting the view that the global RNA-seq approach adopted here, which does not require PCR, is suited to the study of bacteria with GC-rich genomes (Lin *et al*., 2013). The disparity that exists appears to be due at least in part to some regions that were not sequenced with high frequency producing significant hybridisation signals. The latter may have resulted from a limited amount of cross-hybridisation. Although the probes were experimentally validated and selected on the basis of sensitivity and selectivity (Bucca *et al*., 2009) and the microarrays have been used successfully (Lewis *et al.*, 2010, Allenby *et al.*, 2012, Swiatek *et al.*, 2013, Rico *et al.*, 2014), it is not possible to completely eliminate the effects of cross-hybridisation (Wernersson *et al.*, 2007, Mulle *et al.*, 2010).

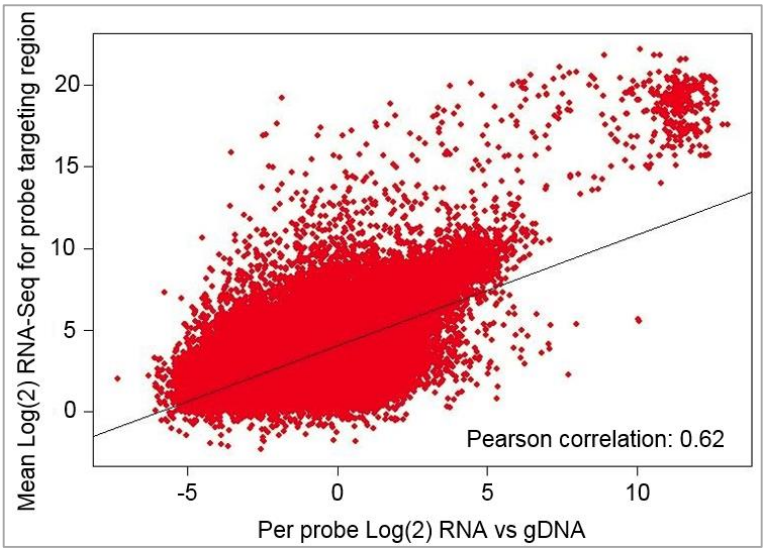

Figure 3.8: Comparison of global RNA-seq and microarray data for *S. coelicolor*. The mean RNA-seq reads for each base within the 60-bp region targeted by a microarray probe is directly compared with the microarray signal for the same probe target. Trendline calculated by the linear model fit within R (lm function) when given the formula RNA-seq_data ~ Array_data.

Consistent with the *S. coelicolor* RNA being isolated during exit from exponential growth, viewing of the transcriptome map revealed that the production of several secondary metabolites was primed. Readily detected was transcription of *act*II-ORF4, *redZ* and *redD*, and *cdaR*, the cluster-situated regulators of actinorhodin (Act), undecylprodigiosin (Red) and the calcium-dependent antibiotic (CDA), respectively (Gramajo *et al.*, 1993, Guthrie *et al.*, 1998, Narva & Feitelson, 1990, Hojati *et al.*, 2002), and *bldA*, the leucyl tRNA for the rare TTA codon (Lawlor *et al*., 1987), which is required for the effective translation of *act*II-ORF4 and *redD*, two of the aforementioned regulators. The level of transcription of *bldA* was similar to other tRNAs such as Val (CAC), Leu (TAG) and Arg (CCG) (data not shown). Transcription of the biosynthetic genes for Act, Red and CDA was barely detectable suggesting secondary metabolite production may have been awaiting the triggering of the stringent response (Strauch *et al.*, 1991). Interestingly, however, in the case of *act*II-ORF4 a long asRNA (>500 nt) that conceivably has a regulatory role was detected (Figure. 9, panel A). Moreover, a long asRNA was also detected for *scbA* (Figure. 9, panel B), which encodes the synthase of γ-butyrolactones (Kato *et al.*, 2007, Hsiao *et*

al., 2007), small signalling molecules that regulate secondary metabolism and morphological differentiation (Willey & Gaskell, 2011). A transcript antisense to *scbA* was also reported by Moody *et al.* (2013). Further unexpected features were identified for other key regulators. For example, previously unknown transcripts in the intergenic region between *afsR2* (Vogtli *et al.*, 1994) and *afsR* (Horinouchi *et al.*, 1990), both of which are regulators of secondary metabolism was detected. This transcript could have a discrete function or be the result of active riboswitching upstream of *afsR2* (Figure 9, panel C). Additional transcriptional complexity was also detected for a number of transcription factor genes including *whiB*, a redox-sensitive transcription factor required for sporulation (Davis & Chater, 1992); in addition to the two promoters previously identified for *whiB* (Soliveri *et al.*, 1992), a strong promoter farther upstream was identified (Figure. 9, panel D).

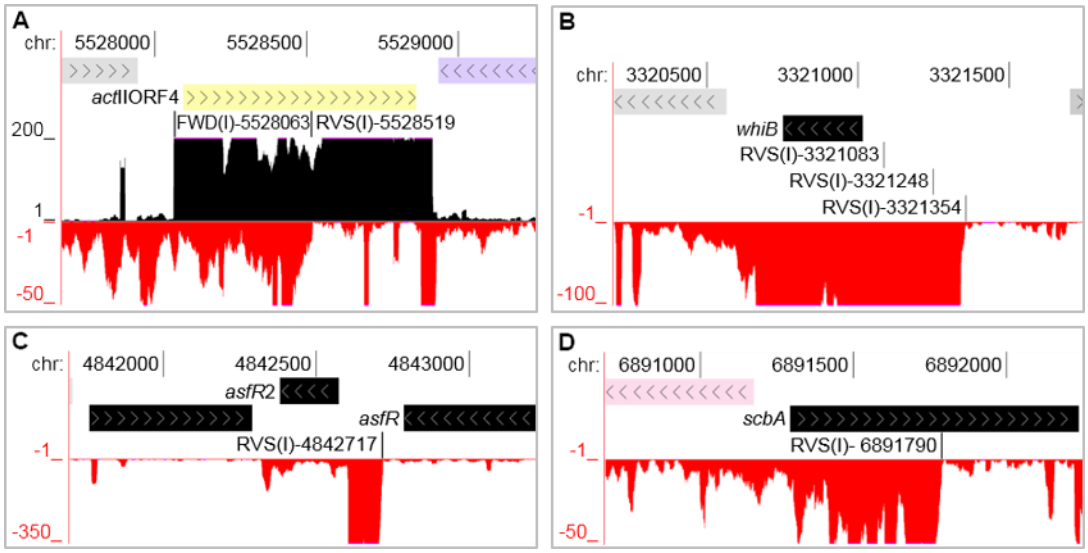

Figure 3.9: Examples of transcriptional complexity associated with key regulators of *S. coelicolor* metabolism and development. Panels A, **B**, **C** and **D** show annotated views of *act*II-ORF4, *whiB*, *afsR2*, and *scbA*, respectively. Labelling is as in Figure 3.3.

Finally, knowing the nucleotide position of a control step narrows and simplifies the search for *cis*-regulatory elements (Lin *et al*., 2013). By way of illustration, sequences immediately upstream of TSSs associated with genes of the translational machinery were searched using MEME (Figure S1) (Bailey *et al.*, 2009). This reveal conserved

hexanucleotide regions similar to the consensus sequences reported previously for the 'vegetative' promoters of streptomycetes (Strohl, 1992) and *E. coli* (Harley & Reynolds, 1987, Lisser & Margalit, 1993). Moreover, as has been reported recently for *P. acnes* (Lin *et al*., 2013), another actinomycete, the '-35' box in *S. coelicolor* appears to be on average 2 to 3 bp further upstream from the TSS than its *E. coli* counterpart (Figure 3.10). This means that a shared TTG motif, located in the 5' half of the *E. coli* box and in the 3' half of the *S. coelicolor* box, is on average in the same position relative to TSSs in both organisms. This may explain at least in part why many *Streptomyces* promoters function effectively in *E. coli* (Strohl, 1992). Alignment of the *S. coelicolor* promoters did not reveal the GC-rich discriminator region (Travers, 1980), which is located immediately downstream of the '-10' box of promoters of *E. coli* genes encoding rRNA and r-proteins and is now known to facilitate regulation by the RNA polymerase-binding factors DksA and (p)ppGpp (Haugen *et al.*, 2006) as part of the bacterial stringent response (Aldridge *et al.*, 2013).

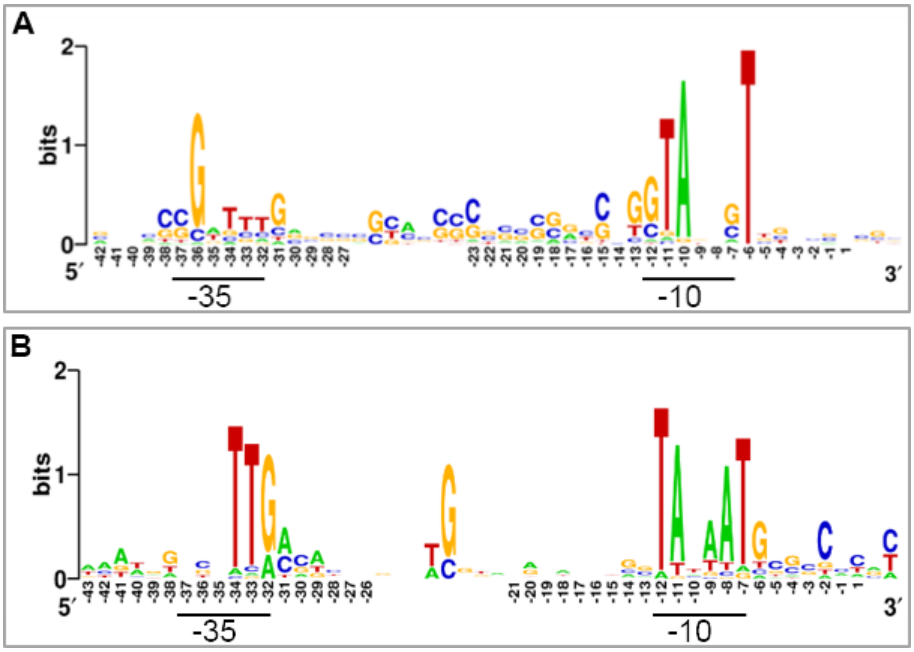

Figure 3.10: Conserved sequences in promoters associated with genes encoding the translational machinery. Panels A and **B** are Weblogo representations of the promoters of *S. coelicolor* and *E. coli*, respectively. The length of the space upstream of the -10 box was adjusted manually to maximise alignment of the −35 box as described previously (Lin *et al*., 2013). The combined height of nucleotide symbols shows the level of sequence conservation at a particular position, while the height of individual symbols within a stack of nucleotides indicates the relative frequency at that position. The blank positions in the Weblogo show sites lacking nucleotide sequence conservation. The nucleotide positions are numbered relative to the average position of TSSs to the point at which gaps were introduced to maximise the alignment. The numbering of position over the region of the -35 box is based on the average length of the spacer region.

3.3 Discussion

By combining global and differential RNA-seq methodologies that should not be unduly affected by high GC content (Lin *et al*., 2013), a genome-wide view of many factors that control *S. coelicolor* gene expression at the level of transcription initiation and beyond was obtained successfully. For example, lmRNAs (Figure 3.3), key steps in the processing and degradation of rRNA, tRNA and mRNA (Figure 3 & 4), and small RNAs, including those that may be involved in attenuation-like switching mechanisms (Figure 3.7), were detected. Many of the small RNAs identified in this study are novel (Table S6). Moreover, it is likely that more exists because the limited number of growth conditions studied to date (Vockenhuber *et al*., 2011, Moody *et al*., 2013) are unlikely to have captured the full physiological depth of *S. coelicolor* (Bentley *et al.*, 2002). Our approach also identified over one thousand TSSs (Table S2) and transcription units, encompassing all classes of RNA. The global RNA-seq data can now be mapped onto the genome sequence (Sallet *et al*., 2013) to provide experiment-based annotation that should aid whole-cell modelling of, for example, regulatory modules (Castro-Melchor *et al*., 2010). The examination of individual regulators that are central to the control of secondary metabolism and morphological development has already revealed new layers of transcriptional complexity (for examples, see Figure 3.9). Moreover, as was illustrated using promoters associated with the translational machinery (Figure 3.10), knowing the nucleotide positions of key events in gene control can aid the identification of *cis*-regulatory sequences. This type of analysis can now be extended to other steps in the control of gene expression such as RNA processing and degradation and transcriptional termination. The RNA-seq data analysed here has been deposited in the GEO archive (Barrett *et al*., 2013). The parallel analysis of *E. coli* in addition to validating the approach provided new insights into its gene regulation (for further details, see Romero *et al*., 2014), despite it being one of the best-studied model organisms (Neidhardt, 1996).

Although lmRNAs are prevalent in *S. coelicolor*, processing of 16S rRNA at the equivalent of the -43 site in *E. coli*, which generates specialised ribosomes capable of

'leaderless' translation was not detected (Figure 3.4). As mentioned earlier, this was not unexpected, as it has been shown for streptomycetes that specialised ribosomes capable of translating lmRNA can be produced by modification of the 16S rRNA (Kaberdina *et al*., 2009). It would be interesting to determine whether *S. coelicolor* produces a molecule analogous to kasugamycin, a ribosome-interacting aminoglycoside originally isolated from *S. kasugaensis* that promotes 16S rRNA modification and thereby the translation of lmRNA (Kaberdina *et al*., 2009, Schluenzen *et al.*, 2006, Schuwirth *et al.*, 2006), and what role, if any, links the functions encoded by lmRNAs in *Streptomyces* spp. The substantial difference in the prevalence of lmRNAs between *S. coelicolor* and *E. coli* may also reflect the fact that the r-protein S1, which strongly promotes SD interactions in *E. coli* (Sorensen *et al*., 1998), is truncated at its C-terminus in *S. coelicolor* (SCO1998). Regardless of the underlying molecular biology, the results of this study add to the growing body of evidence that bacteria differ substantially in the extent to which they use lmRNAs (Nakagawa *et al.*, 2010).

While this study adds to growing evidence for sRNAs in *S. coelicolor*, simply because a region is transcribed does not mean it has a function (Graur *et al.*, 2013). Evidence for background (or pervasive) transcription on a genome scale has been obtained for several bacterial species (e.g. (Lin *et al*., 2013)) including *E. coli* (Raghavan *et al.*, 2012). Therefore, an assessment of the impact on cell physiology of the plethora of small RNAs being discovered will require careful genetic analysis. Background transcription could explain at least a proportion of the TSSs in Class II, which are associated with TAP enrichment, but not an obvious step increase in transcription. However, verification of background transcription initiation will require a number of biological replicates and statistical analysis, as applied recently to *P. acnes* (Lin *et al*., 2013). TSSs associated with alternative promoters nested downstream of ones that produce a substantial increase in transcription would also have been assigned here to Class II.

The ability to detect sites involved in the initiation or mediation of rapid mRNA degradation using the differential RNA-seq approach described here (Figure 3.5) offers a much-improved platform to further understanding of this key aspect of gene

regulation. By extending the analysis to strains defective in key ribonucleases and their regulators, it should be possible to determine the impact of individual factors on a genome-wide scale, and to identify model transcripts, whose subsequent characterisation should reveal the underlying molecular and structural biology. Moreover, by adapting our differential RNA-seq approach to include the ligation of a 3' adaptor prior to fragmentation of the RNA and the subsequent addition of a 5' adaptor, it should be possible to investigate degradation by 3'-5' exonucleases. While the results of this and related studies show that treatment with TAP is able to differentiate nascent 5' ends, treatment with TEX, the 5' to 3' exonuclease specific for 5'-monophosphorylated transcripts, may offer increased discrimination of TSSs that cannot be identified by TAP enrichment *in vitro* should efficient de-pyrophosphorylation occur for some transcripts in *S. coelicolor* as in *E. coli* and *B. subtilis* (Sharma *et al.*, 2010, Irnov *et al.*, 2010, Thomason *et al.*, 2015). TEX treatment would remove the majority of the de-pyrophosphorylated species thereby allowing those with nascent 5' triphosphorylated ends to be detected. Another improvement would be to remove the PCR step from the differential RNA-seq approach. As illustrated by the global RNA-seq approach, amplification is not required during the preparation of cDNA libraries (Mamanova *et al*., 2010). The removal of PCR would remove amplification bias.

Combining Class I and III, 1598 TSSs were obtained for *S. coelicolor*. This number is slightly lower than the number of proteins that have been detected for single conditions by proteomic approaches (Manteca *et al.*, 2006, Rodriguez-Garcia *et al.*, 2007); however, it should be remembered that many transcription units in both contain multiple genes. In addition, the numbers of transcripts detected will likely increase as more conditions are analysed. The inclusion of differential RNA-seq, regardless of its form, is crucial for accurate TSS assignment. Sites of processing, including many which are well characterised and documented (*e.g.* RNase P maturation of the 5' end of tRNA), have been identified erroneously as transcriptional start sites by a previous RNA-seq analysis of *E. coli* (Cho *et al*., 2009). Finally, the addition of a phosphorylation step to our differential RNA-seq approach would allow

the identification of the cleavage sites of RNases that produce downstream products with 5'-hydroxyl group. This is likely to be particularly relevant to studies of suboptimal growth conditions under which such RNases, e.g. *E. coli* MazF (Vesper *et al*., 2011), are highly activated. The addition of a phosphorylation step would also allow the identification of positions with transcriptional start sites that are primed by nanoRNAs (Nickels & Dove, 2011, Goldman *et al.*, 2011), which a recent study indicates tend to have a 5'-hydroxyl group (Vvedenskaya *et al.*, 2012). The latter study also suggests that while nanoRNAs can alter gene expression, this class of sRNA are not absolutely required for transcription from individual promoters. Thus, the omission of a phosphorylation step in this study should not have prevented TSSs being identified.

Data access

The RNA-seq data has been deposited in the GEO archive (Barrett *et al*., 2013) under accession number GSM1126846.

Chapter 4

4 Potential targets of AtrA, the binding activity of recombinant orthologues and regulation by small molecules

4.1 Introduction

As already described (see Section 1.3), AtrA is a TetR-family transcriptional regulator that was initially discovered in the guise of an activity that binds to the promoter region of *act*II-ORF4 *in vitro* and is require for maximum production of actinorhodin production in *Streptomyces coelicolor* (Uguru *et al*., 2005). Sites bound by AtrA *in vitro* are also known to be occupied under at least some conditions *in vivo* (McArthur & Bibb, 2008). Recently, evidence has emerged through promoter engineering that higher levels of actinorhodin production can be achieved by inducing expression of *act*II-ORF4 early in growth (Sohoni *et al.*, 2014). Thus, the role of AtrA may be primarily to control temporal rather than bulk expression. The details of AtrA-mediated regulation remain unclear, but it has emerged that a number of other transcription factors also bind the *act*II-ORF4 promoter (see Section 1.3). These transcription factors included DasR, which controls the transport and metabolism of the monosaccharide *N*-acetylglucosamine (GlcNAc) (Rigali *et al*., 2006, Colson *et al.*, 2007), coordinating with wider cellular activities including secondary metabolism (Swiatek-Polatynska *et al*., 2015). Indeed, DasR binds directly the promoters of the cluster-situated regulators of actinorhodin biosynthesis and other secondary metabolites (Swiatek-Polatynska *et al*., 2015). The *nagE2* gene, which encodes the high-affinity transporter of GlcNAc, is regulated by AtrA as well as DasR (Nothaft *et al.*, 2010). Thus, AtrA may influence the regulation of actinorhodin production by GlcNAc at both the start and end of the biosynthesis pathway.

Using the position weighted matrix for sites bound by AtrA *in vitro*, the sequence of the genome of *S. coelicolor* has been scanned by us and our collaborators for sites

resembling the consensus binding site for AtrA (see Section 1.4). This revealed, for example, a possible AtrA-binding site in the promoter region of *ssgR* (van Wezel, pers. comm.), the transcriptional activator of *ssgA* (Traag *et al.,* 2004), which encodes a key member of a family of actinomycete-specific proteins that regulate cell division and spore maturation (Traag & van Wezel, 2007). Moreover, analysis of the *ssgR* promoter region using a DNA-affinity capture assay (DACA) (Park *et al.,* 2009) by the laboratory of Prof. Byung-Gee Kim (Seoul National University, Korea) identified AtrA as a potential binding partner. The RNA-seq analysis described in the previous chapter revealed that *ssgA* (and all other *ssg* genes in the same operon) were poorly expressed during submerged growth in liquid culture. This is consistent with the finding that *S. coelicolor*, unlike *S. griseus*, only sporulates on solid media (Daza *et al.*, 1989, Manteca *et al.*, 2010). The expression of *atrA* also appeared to be poor and is consistent with the possibility that AtrA activates expression of *ssgR*.

The function of *atrA* is also being studied beyond *S. coelicolor*. Very recently, AtrA was found using a DACA-type approach to bind to the promoter that directs transcription of the entire cluster for daptomycin (Mao *et al.,* 2015), a cyclic lipopeptide antibiotic produced by *Streptomyces roseosporus* (Miao *et al.*, 2005). Disruption of *atrA-sr* produced a bald phenotype and completely blocked daptomycin production. In the same study, *atrA-sr* expression was shown to be positively regulated by AdpA, an orthologue of the A-factor receptor (Ohnishi *et al.*, 1999). Thus, in *S. roseosporus* AtrA is a mediator of the A-factor signalling pathway. Moreover, it was also shown that AtrA-sr binds to its own promoter and autoregulates its own expression. Interestingly, comparison of the binding sites identified in *S. roseosporus* revealed little similarity with those identified in *S. coelicolor*.

In *S. griseus*, AtrA is not a major mediator of the A-factor regulatory cascade, but can have an influence under some conditions (Hirano *et al.,* 2008). AtrA-sg binds to the promoter of *strR*, the final activator of streptomycin production, between the binding sites of AdpA (Hong *et al.,* 2007, Hirano *et al.,* 2008). Disruption of AtrA in *S. avermitilis* led to an increase in the production of avermectin by *S. avermitilis*, by a mechanism

that directly involves the cluster-situated regulator, AveR (Chen *et al.,* 2008). Transcriptomics revealed that the disruption of AtrA-sa (also called AveI) affected the expression of the biosynthetic clusters for oligomycin and filipin as well as genes encoding enzymes, such as crotonyl-CoA reductase and methylmalonyl-CoA decarboxylase, involved in the supply of precursors for avermectin and other natural products (Mao *et al.,* 2015). Disruption also altered the expression of several genes involved in protein synthesis and fatty acid metabolism suggesting that AtrA may function as a global regulator of flux of carbon from primary to secondary metabolism in *S. avermitilis*. Consistent with such a role, scanning of the *S. coelicolor* genome for potential sites of AtrA binding using the position weighted matrix (see Section 1.4) identified potential targets in the promoters regions of genes encoding acetate kinase, α-isopropylmalate/homocitrate synthase family transferase and acetyl-CoA acetyltransferase/thiolase.

One aspect of AtrA-mediated regulation that had not been reported in the literature prior to the initiation of the work described in this chapter was the possibility that the binding activity of AtrA, like many other members of the TetR family (Cuthbertson & Nodwell, 2013), is regulated by the binding of a small molecule. However, while Bin Hong and colleagues (Peking Union Medical College, China), were investigating the regulation of lidamycin production by *S. globisporus*, they not only discovered that AtrA-gl binds directly to promoter of cluster-situated regulator of lidamycin production, but that the DNA-binding activity of AtrA-gl could be inhibited by the binding of a biosynthetic intermediate of lidamycin. In addition, they also reported that the activity of AtrA-gl could be inhibited, at least *in vitro*, by actinorhodin, a secondary metabolite not produced by *S. globisporus*. This chapter describes the purification of AtrA-gl and its characterisation in direct comparison with *S. coelicolor* AtrA, which from hereon will be referred to as AtrA-c. In addition, it describes the assay used to validate putative AtrA-binding sites (see Section 1.4) and the effects of disrupting *atrA* on the transcription of potential targets of AtrA regulation in *S. coelicolor*.

4.2 RESULTS

4.2.1 Purification and characterisation of AtrA from *S. coelicolor* and *S. globisporus*

4.2.1.1 Constructed Plasmids

All constructs used for the expression of recombinant genes in *E. coli* were based on pET16b (Figure 1, panel A). The genes were inserted between the NdeI and BamHI sites such that their products were tagged with decahistidine (or 10xHis) at the N-terminus to facilitate purification by immobilised metal affinity purification as detailed previously (Uguru *et al.,* 2005). In addition to AtrA from *S. coelicolor* and *S. globisporus*, *S. coelicolor* ActR, which encodes the regulator of the efflux pump for actinorhodin (Tahlan *et al.,* 2007) and *E. coli* TetR, which is the archetypal member of the TetR family and controls the expression of the TetA tetracycline resistance determinant, an efflux pump (Meier *et al.,* 1988) were included as controls (see below, for details). Digestion of the corresponding "expression" plasmids (Table 2.3) with NdeI and BamHI produced fragments of the expected size (Figure 1, panel B). The integrity of each cloned gene was verified by DNA sequencing (data not shown).

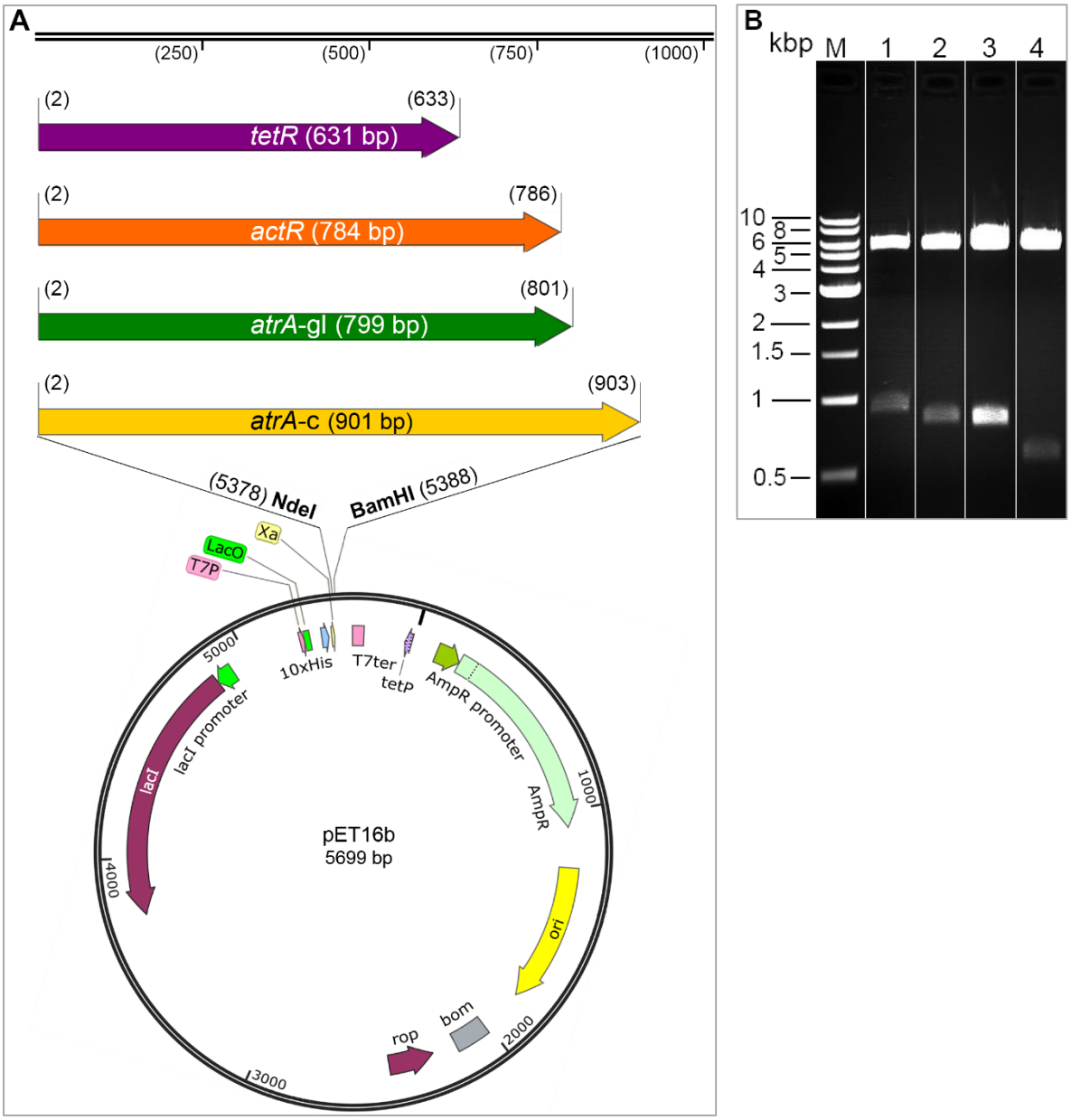

Figure 4.1: Plasmids used to produce AtrA and other members of the TetR family *in E. coli*. Panel A is a schematic illustration of each of the four genes cloned into pET16b. The lengths of the corresponding NdeI-BamHI fragments are indicated in parentheses. Key features of the pET16b plasmid are also indicated, such as the segment that encodes decahistidine (10xHis), the T7 promoter (T7P) that drives transcription of cloned genes, the T7 terminator (T7ter), which terminates transcription by bacteriophage T7 RNA polymerase, the lac operator (lacO), which represses transcription from T7P in the absence of induction, the AmpR, which confers resistance to ampicillin, carbenicillin, and related antibiotics, and factor Xa-protease site (Croxatto *et al.*, 2002), which is useful in cleaving 10xHis tags from engineered His-tagged fusion proteins. The *atrA*-c, *atrA*-gl, *tetR* and *actR* (SCO5082) genes are

coloured gold, green, orange, and purple, respectively. **Panel B** shows the agarose gel electrophoresis (1 %) analysis of the products of digestion with NdeI and BamHI. Lanes 1 to 4 correspond to plasmids encoding AtrA-c (pRA001), AtrA-gl (pET16b-atrA-gl), ActR (pET16b-ActR), and TetR (pET16b-TetR), respectively. The expected sizes of the corresponding insert fragments are 901, 799, 784 and 631 bp, respectively. Lane M corresponds to Quick Load 1 kb DNA ladder (New England Biolabs). Numbers on the left of this panel indicate the sizes of the marker fragments.

4.2.1.2 Over-production and purification of AtrA-c, AtrA-gl, ActR and TetR

AtrA-c, AtrA-gl, ActR and TetR were each purified by IMAC as described previously (Uguru *et al.,* 2005) with only minor modification (see Section 2.14.1). SDS-PAGE analysis of the purification of proteins are shown in Figure 4.2. A polypeptide of the size expected for AtrA-c (29.93 kDa) was clearly induced (cp. lanes 1 and 2), remained in the soluble fraction following lysis of the cells and ultracentrifugation (lane 3), bound to the column with high efficiency (lane 4), eluted as a single major species (lane 5), and remained soluble during dialysis and concentration (lane 6). Preparations of AtrA-gl (27.5 kDa), ActR (28.3 kDa) and TetR (23.3 kDa) of comparable purity were also obtained (see Figure 4.2). The purity of AtrA-gl was comparable to that of AtrA-c (Figure 4.2, panel B, lane 7).

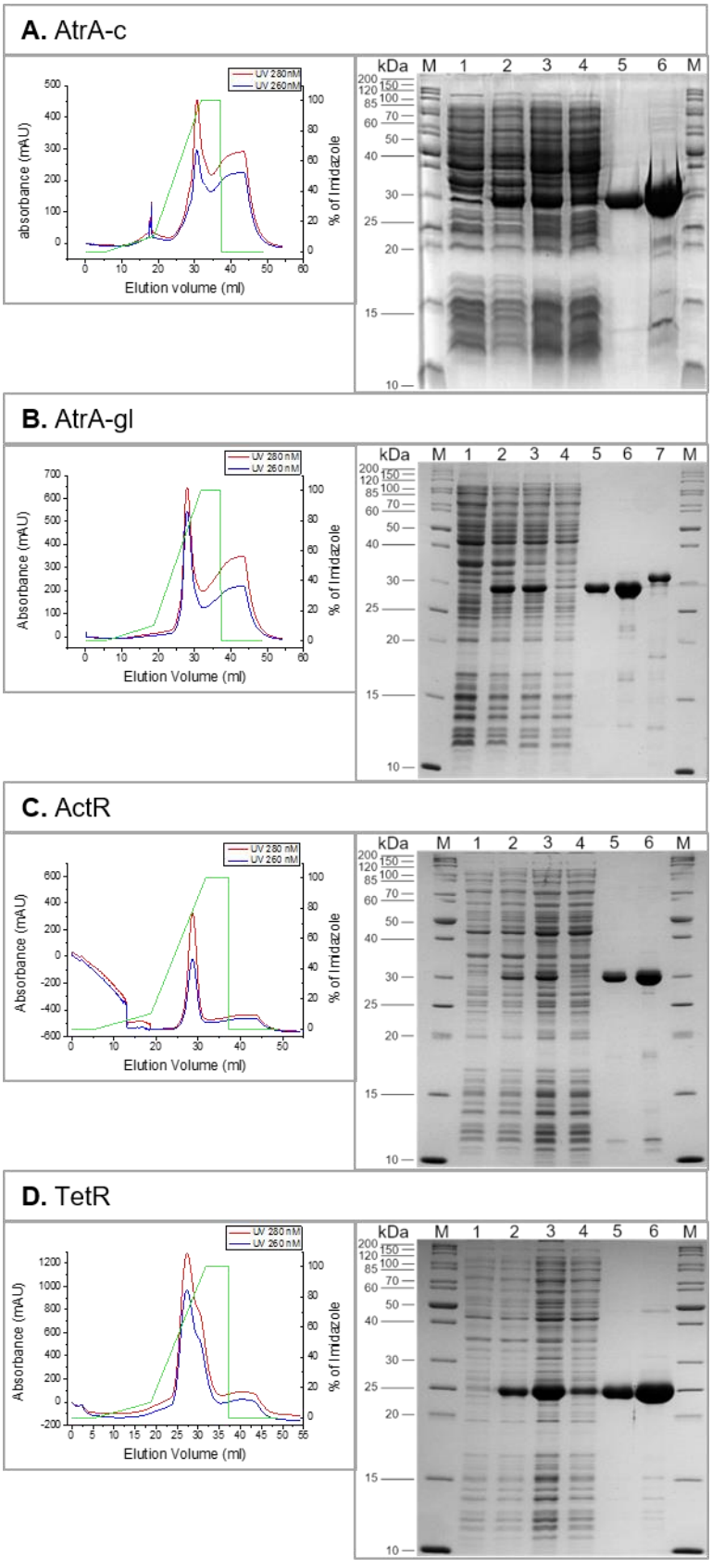
A. AtrA-c
UV 280 nM
UV 260 nM
absorbance (mAU)
% of Imidazole
Elution volume (ml)
kDa M 1 2 3 4 5 6 M
B. AtrA-gl
UV 280 nM
UV 260 nM
Absorbance (mAU)
% of Imidazole
Elution Volume (ml)
kDa M 1 2 3 4 5 6 7 M
C. ActR
UV 280 nM
UV 260 nM
Absorbance (mAU)
% of Imidazole
Elution Volume (ml)
kDa M 1 2 3 4 5 6 M
D. TetR
UV 280 nM
UV 260 nM
Absorbance (mAU)
% of Imidazole
Elution Volume (ml)
kDa M 1 2 3 4 5 6 M

Figure 4.2: Purification of AtrA-c, AtrA-gl, ActR and TetR from *E. coli* by IMAC. Left inset from each panel shows the HPLC profile for the elution step of the purification. The absorbance measure at 260 nm (blue) and 280 nm (red) are plotted against elution volume (mL). The green line shows the gradient of imidazole from 145 mM to 1 M (0 to 100%, respectively). The flow rate was 1 mL/min. Fractions corresponding to the single peak were pooled together prior to dialysis and concentration. Right inset from each panel shows SDS-PAGE analysis of fractions from different steps in the purification. Lanes 1 to 6 are as described in the main body of the text. Lanes labelled M contain PageRuler Unstained Protein Ladder (Fermentas). Numbering on the left of this panel indicates the sizes of the markers (kDa). Lane 7 in panel B contained 3.2 μg of purified *S. coelicolor* AtrA.

4.2.1.3 EMSA analysis

To investigate the possibility that AtrA modulates primary metabolism we generated promoter probes for genes encoding enzymes associate with acetyl-CoA metabolism, acetate kinase, α-isopropylmalate/homocitrate synthase family transferase and acetyl-CoA acetyltransferase/thiolase (SCO5424, *ackA*; SCO5529, *leuA2*; and SCO6027, respectively), as suggested by scanning of the *S. coelicolor* using the PWM for AtrA (see Section 1.4), probes corresponding to these regions were amplified by PCR. Positive controls were provided by amplifying two regions corresponding to the promoter of *act*II-ORF4. The first encompassed the two binding sites for AtrA, while the second only encompassed "Site 2", which of the two sites forms the highest affinity interaction (Uguru *et al.,* 2005). Negative controls were provided by amplifying regions corresponding to the promoters of *act*I-ORF1 and *act*III, which are divergent, and *vdh*. Both regions do not have sequences with significant matches to the PWM for AtrA. The latter, which encodes valine dehydrogenase, was used previously as a negative control for AtrA binding (Uguru *et al.,* 2005). The promoter region between *atrA* (SCO4118) and the gene encoding NADH dehydrogenase (SCO4119) was also included, as contrary to a report by others, previous attempts by the McDowall laboratory to identify AtrA binding to this region had failed (McDowall, pers. comm.). The opportunity was also taken to confirm binding to the promoter region upstream of *ssgR*. Each probe was incubated separately with increasing amounts of purified AtrA-c from 0 to 5 μM (lanes 1 to 15, Figure 4.3).

AtrA bound within the promoter region of *act*II-ORF4 forming distinct complexes (Figure 4.3, panel A). More than 50% of the probe (labelled P) was shifted to higher position (labelled I) with an AtrA concentration of 156 nM (lane 10). As the concentration of the probe was considerably less, 5 nM, the concentration of AtrA at which 50% of the probe was shifted (78 to 156 nM, lanes 9 to 10) provided an estimate of the apparent equilibrium dissociation constant of this interaction (K_d'). With the addition of higher concentrations of AtrA, the probe shifted to a higher position (labelled II) consistent with the presence of a second binding site (Uguru *et al.,* 2005). The majority of the probe was present in a second complex in the presence of 156 nM AtrA (lane 10). Above 156 nM the migration of the probe was increasingly retarded (lanes 11 to 16), a finding that may represent additional binding that less sequence specific. AtrA bound to a region of the promoter of *act*II-ORF4 containing only Site 2 forming a single complex (labelled I; Figure 4.3, panel B). The majority of this probe was bound by 156 nM (lane 10). Interestingly, at higher AtrA concentrations the migration of this complex (corresponding to Site 2) was not increasingly retarded. This suggests that the proposed sites of less sequence-specific binding flank Site 1 and not Site 2. Overall, the results described above were consistent with previous findings both qualitatively and quantitatively (Uguru *et al.,* 2005).

AtrA binding to the promoter region between *act*I-ORF1 and *act*III (panel C) and *vdh* (panel D) occurred at only the highest AtrA concentrations of 2.5 and 5 µM (lanes 14 and 15, respectively). Given the absence of any significant matches to the PWM for AtrA, these interactions are thought to be non-specific. Incubation of AtrA with the probe for SCO5424 (panel E) failed to detect any binding except at the highest protein concentrations (lanes 14 to 15). However, a single complex was readily detected for SCO5529 (panel F) with approximately 50% of the probe being shifted by 78 nM AtrA (lane 9). A complex was also detected for SCO6027 (panel G), but at a concentration (1.25 µM, lane 13) just below that at which AtrA binds to the negative controls *act*I and *vdh* (panels C and D, respectively). Consistent with the observed affinities, the sequence in the promoter of SCO5529 (*leuA2*) is a better match to the PWM of AtrA than that of SCO6027, which is a better match than that of SCO5424. Thus, although it

is clear that genome scanning with the AtrA PWM may produce false positives (e.g. SCO5424), it has identified sites that may regulate the expression of SCO5529 (*leuA2*) and SCO6027, which encode enzymes that directly utilise and produce acetyl-CoA, respectively. Binding to the *ssgRA* promoter region was also detected (panel H), as found by others (Prof. Byung-Gee Kim, pers comm.). However, the affinity of the interaction was clearly weaker than some others, *e.g. act*II-ORF4 and SCO5529. Binding was not detected to the promoter region of *ssgB* (panel I), which was included as an additional negative control, and the promoter region between *atrA* (SCO4118) and the divergent gene upstream (SCO4119) (panel J). The latter is consistent with previous findings of the McDowall lab, but contradicts a report by others (Ahn *et al.,* 2012).

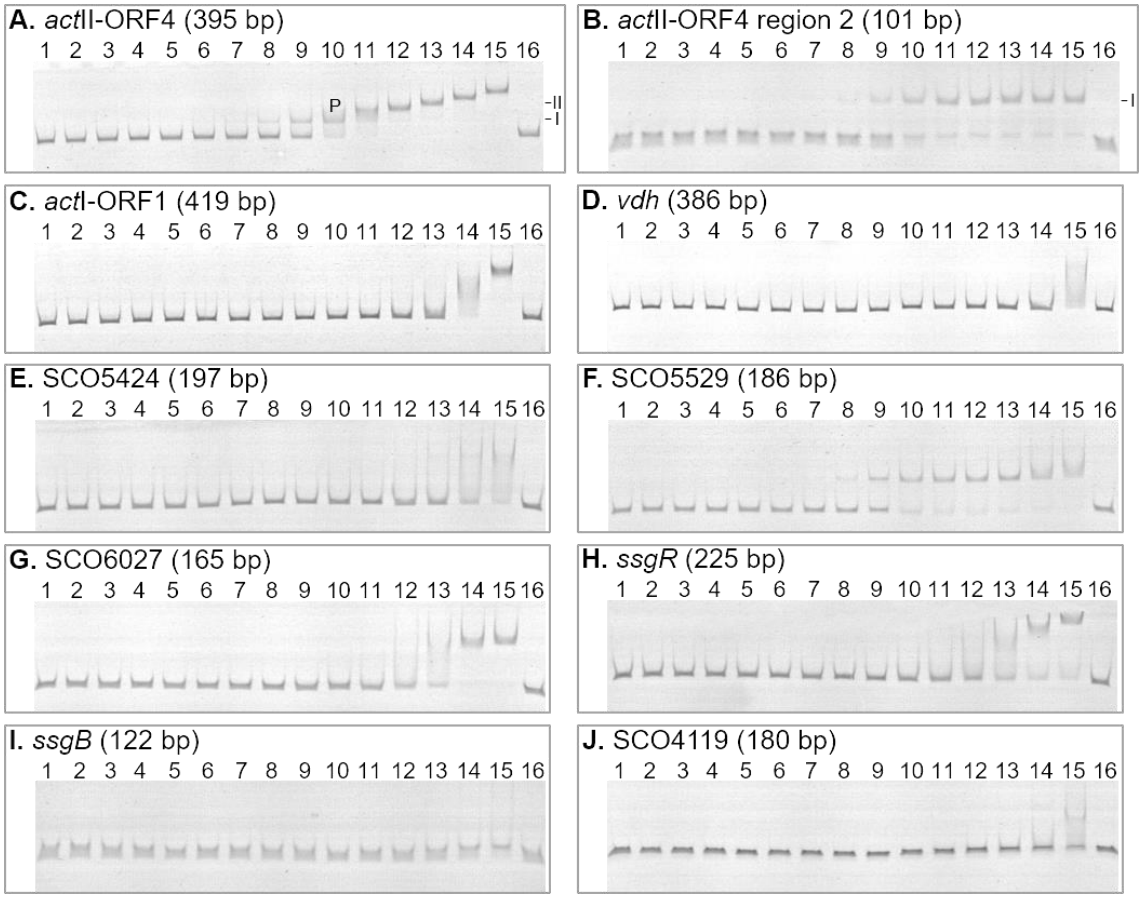

Figure 4.3: *S. coelicolor* AtrA electrophoretic mobility shift assay. Each lane in **panels A** to **J** contain 5 nM of specific fluorescein-labelled probe and 2 µg of Salmon sperm DNA (Invitrogen). Lanes 1 to 16 in each panel contain 0 nM, 0.6 nM, 1.2 nM, 2.4 nM, 4.8 nM, 9.75 nM, 19.5 nM, 39 nM, 78 nM, 156 nM, 312.5 nM, 625 nM, 1.25 µM, 2.5 µM, 5 µM and 0 nM of AtrA-c protein, respectively. The reactions were run in a 4%

(37.5:1) acrylamide: *bis*-acrylamide gel in 1 x TGED buffer for 40 min at 120 V. The gel was imaged using Fujifilm FLA-5000 imaging analyser system. The numbers in parenthesis indicate the size of probe.

To assess the binding activity of *S. globisporus* AtrA, increasing concentrations were incubated with probes for the promoter regions of *act*II-ORF4, *act*II-ORF4 region 2, SCO4119 and SCO5529 (Figure 4.4). The results quantitatively and qualitatively were indistinguishable from those obtained for *S. coelicolor* AtrA. The results confirmed that AtrA binds to the promoter of SCO5529, but not SCO4119 with an affinity similar to that for its interaction with the promoter region of *act*II-ORF4.

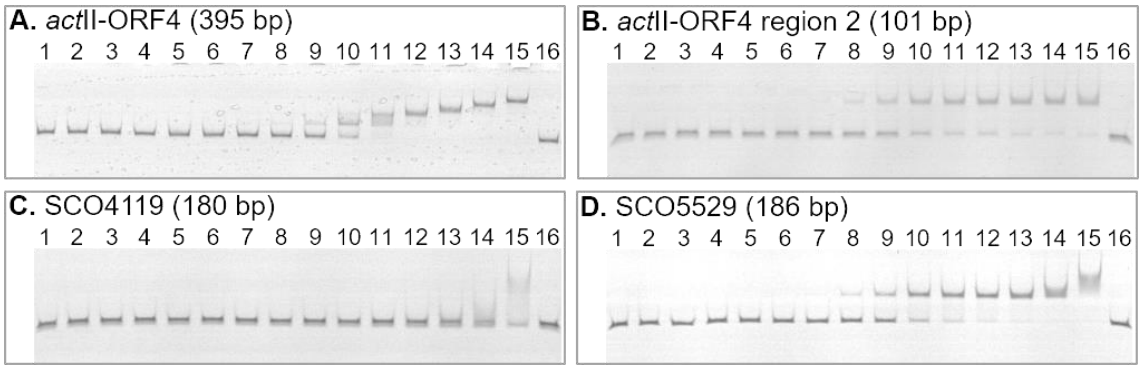

Figure 4.4: *S. globisporus* AtrA electrophoretic mobility shift assay. Each lane in **panels A** to **D** contain 5 nM of specific fluorescein-labeled probe and 2 µg of Salmon sperm DNA (Invitrogen). Lanes 1 to 16 in each panel contain 0 nM, 0.6 nM, 1.2 nM, 2.4 nM, 4.8 nM, 9.75 nM, 19.5 nM, 39 nM, 78 nM, 156 nM, 312.5 nM, 625 nM, 1.25 µM, 2.5 µM, 5 µM and 0 nM of AtrA-gl protein, respectively. The numbers in parenthesis indicate the size of probe. Electrophoresis conditions are identical to that described in Figure 4.3.

4.2.2 Measuring transcript abundance in M145 and L645 strains by RT-qPCR

4.2.2.1 Morphology and RNA purification

Having confirmed *in vitro* binding of AtrA-c to several promoter regions in the previous section, the next objective was to investigate whether disruption of the *atrA* gene affected the transcription of genes putatively regulated by AtrA. RNA for RT-qPCR analysis was isolated from M145 and L645, the congenic *atrA* mutant, grown on the surface of cellophane-overlaid agar plates containing minimal media, with 0.5% [w/v]

mannitol as the sole carbon source (MM). The mycelia were grown as lawns covering the entire surface of the cellophane (diameter of 90 mm) or as distinct patches (diameter of ~16 mm). In each case, the inoculum was 1 x 10^5 CFU of spores/ mycelia fragments. M145 reproducibly generated detectable amounts of actinorhodin when grown as lawns for 4 days and as patches for 9 days (Figure 4.5, panel A). The longer time required for the production of actinorhodin when M145 was grown as patches probably reflected the slower utilisation of nutrients in the plate as a result of limiting the area of inoculation. After being grown as a lawn for 4 days or patches for 9 days, production of actinorhodin by L645 was not observed, despite undergoing normal morphological differentiation (Figure 4.5, panel A). Previously, the time at which actinorhodin production by M145 was first detected offered good discrimination of the effects of *atrA* on transcription of *act*II-ORF4 (Towle, 2007). Representative samples of total RNA preparations before and after treatment with DNase I are shown in Figure 4.5, panels B and C. Distinct bands were evident for 23, 16 and 5S rRNA indicating that degradation of the samples had not occurred during isolation or purification.

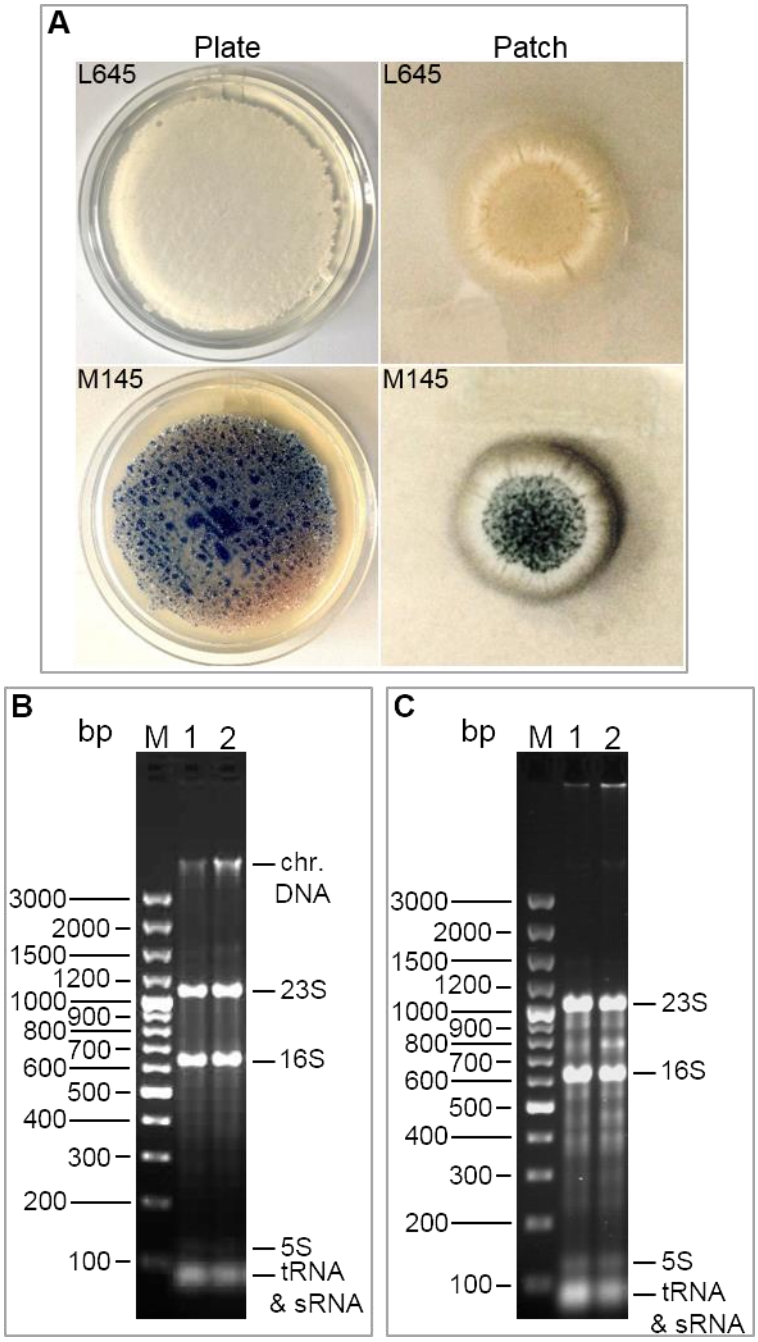

Figure 4.5: Phenotype and RNA analysis of M145 and L645 strains. Panel A shows bacterial growth on cellophane-overlaid MM plates. The mycelial patches and lawns were exposed to ammonia fumes before being photographed. RNA was isolated from additional patches and lawns grown in parallel, not exposed to ammonia. **Panels B** and **C** show RNA samples before and after treatment with DNase I, respectively. Lanes 1 and 2 contain 1 µg of RNA from each of M145 and L645, respectively, grown on MM plates for 96 h. Lane M contains 0.5 µg of GeneRuler 100 bp DNA plus ladder (Thermo Scientific). Numbers on the left of the two panels indicate the sizes of the marker bands (bp). The samples were analysed using 1.2% agarose gels (1X TBE).

4.2.2.2 RT-qPCR

After preliminary investigations, *rpsL*, which encodes ribosomal protein S9, was adopted as an internal control because it is constitutively expressed and the abundance of its mRNA is in the middle of the range for transcripts of interest (Romero *et al.,* 2014). The analysis focused initially on the effects of *atrA* disruption on a selection of *ssg* genes and used RNA isolated from mycelia grown as lawns for 4 days. The integrity of the resulting PCR amplicons were subjected to gel electrophoresis and melt curve analysis to confirm that they were the expected size and the absence of primer dimers, respectively. Representative examples of these analyses for end-products of RT-qPCR are shown in Figure 4.6. In each case, a single species of the expected size was detected.

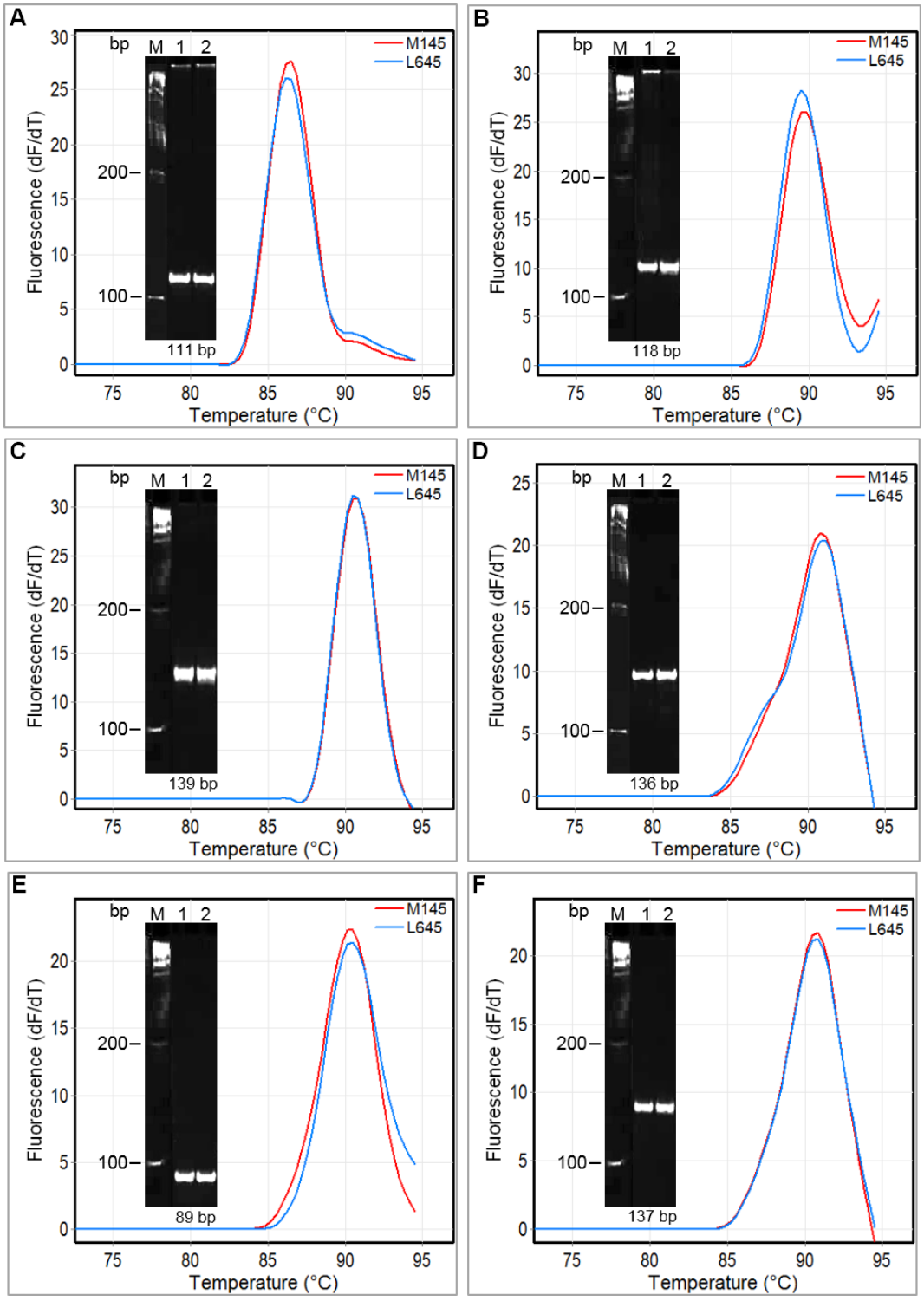

Figure 4.6: Melting curve and polyacrylamide gel electrophoresis analysis of amplified qPCR products. Panels A to **F** represent melting curve analysis of the end product of qPCR exploiting gene-specific primers against *rpsL* (as internal control), *act*I-ORF1, *act*II-ORF4, *ssgR*, SCO5529 and SCO6027, respectively. Raw data were analysed using rotor-gene 6000 series software. Four µL of each qPCR end product was run on 8% polyacrylamide gel (29:1) for 1:30 hr at 12 W (inset). Lanes 1 and 2 contain DNA

from M145 and L645, respectively. Lane M contains 0.5 µg of GeneRuler 100 bp DNA plus ladder (Thermo Scientific) and the sizes of representative bands are shown on the left hand side of inset. The size of each product is shown beneath the gel. Gel electrophoresis revealed one visible band with expected sizes in all PCR products, monitoring a single sharp melting peak.

A threshold common to the exponential phases of all the reactions in a single run was selected manually and the corresponding number of cycles for each reaction recorded. These CT values were then expressed as the difference relative to *rpsL* (SCO4735), which encodes 30S ribosomal subunit protein S9. In turn, the ΔCT values were used to calculate the difference in abundance using the equation $2^{\Delta CT}$. The initial analysis focussed on two independent samples of mycelia grown as lawns for 4 days. To allow statistical analysis, a third sample was added, but of mycelia grown as patches for 9 days. The average fold decreases in the abundance of the transcripts of *act*II-ORF4 and *act*I-ORF1 in L645 were 20.1 (± 1.9) and 7.5 (± 2.9), respectively. Moreover, the results for the individual samples were similar (data not shown), *i.e.* the reported differences were not dominated by the results from one or two of the samples. These results are in line with our previous data showing that transcription of the *act* biosynthetic gene cluster requires AtrA (Uguru *et al.,* 2005) and provide a control for the effects of AtrA. Disruption of *atrA* did not influence transcription of *ssgD*, which is consistent with it neither being directly or indirectly regulated by AtrA. In contrast, the abundance of *ssgR* transcript (a direct target of AtrA) and that of *ssgA*, which is activated by SsgR, decreased on average 2.9 (± 0.5) and 2.8 (± 0.7) fold, respectively. These RT-qPCR results provided *in vivo* evidence consistent with the direct transcriptional activation of *ssgR* by AtrA. The considerably higher abundance in M145 of the *ssgD* transcript relative to those of *ssgA* and *ssgR* (Figure 4.7) was associated with higher transcription as determined by independent promoter probing results (van Wezel, pers. comm.). SsgD is the only SALP (SsgA-like protein) that plays a role during vegetative growth (Noens *et al.*, 2005).

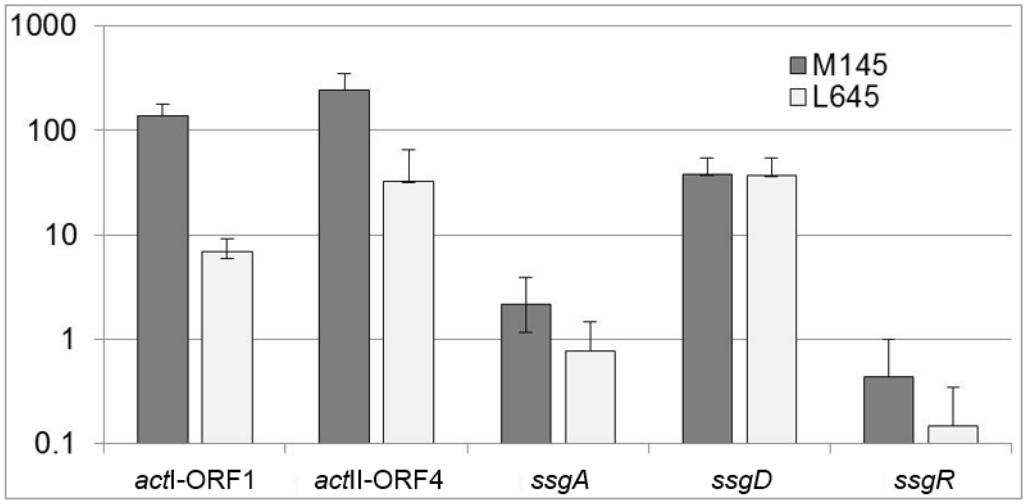

Figure 4.7: Expression levels of transcripts in M145 and L645. Dependency of transcript levels on *atrA in vivo*. Histogram showing the average levels of transcripts in M145 and L645 (the congenic *ΔatrA* partner of M145) expressed as a percentage of the average abundance of the *rpsL* transcript (SCO4735). The values are the average of three independent measurements. Bars indicated the standard error. The Y axis has a log scale, which represents expression level.

To analyse the effects of *atrA* disruption on genes encoding enzymes associate with acetyl-CoA metabolism, in particular those encoding α-isopropylmalate/homocitrate synthase family transferase (SCO5529) and acetyl-CoA acetyltransferase (thiolase) (SCO6027), samples from mycelia grown as lawns for 2, 3 and 4 days were analysed (Table 4.1). A time series was analysed to determine the normal expression profile, such that the significance of any differences between wild-type and *atrA* mutant could be better observed and evaluated. A large difference against a background of constitutive expression is more likely to be of significant than a small difference against a background of changing expression. Between day 2 and 4 the expression of both *act*II-ORF4 and *act*I-ORF1 increased by over 250 fold in M145 (wild-type). This is consistent with the previous finding that *act*II-ORF4 normally peaks at the time actinorhodin is first detected (Uguru *et al.,* 2005). At day 4, the fold decreases in the abundance of the transcripts of *act*II-ORF4 and *act*I-ORF1 in L645 (cf. M145) were 4.5 and 6.3, respectively. The equivalent values obtained in the preceding analysis were 20.1 (± 1.9) and 7.5 (± 2.9), respectively. These quantitative differences may reflect relatively small differences in the growth profile of one or both of the strains between

the analyses, given the large increases in the expression of *act*II-ORF4 and *act*I-ORF1 that accompany growth.

Interestingly, while the expression of *act*I-ORF1 was reduced in the absence of functional *atrA* at days 3 and 4, it was substantially higher (64 fold) at day 2 (L645 vs M145). If reproducible this result would indicate that *atrA* regulation of *act*I-ORF1 expression is more complex than activation of *act*II-ORF4. More specifically, *atrA* may repress the expression of a least some *act* biosynthetic genes earlier in growth. It may also repress the expression of SCO4119, which is adjacent and co-localised throughout evolution. The level of SCO4119 expression was 6.3, 2.0 and 1.4 fold higher at days 2, 3 and 4, respectively in L645 (cf. M145). Given the absence of a good match to the AtrA PWM within the promoter region of SCO4118/9 and the absence of a strong interaction between AtrA and this region, at least *in vitro*, these results suggested that SCO4119 is either repressed indirectly by AtrA or directly by another form of AtrA, perhaps a conformer, that has different binding specificity. The expression of *accA2*, SCO4922, *ackA*, SCO5529, and SCO6027 (and SCO6026, the gene downstream of SCO6027) all increased with growth. However, the expression of these genes did not appear to be affected substantially (>2 fold) by *atrA* disruption.

		Transcriptional Abundance		
Gene Name	**Strain**	**48 hrs**	**72 hrs**	**96 hrs**
*act*II-ORF4 (SCO5085)	M145	2.88	132.51	714.79
	L645	0.69	11.29	153.50
*act*I-ORF1 (SCO5087)	M145	2.61	174.64	713.92
	L645	165.71	61.25	113.50
SCO4119	M145	0.80	3.34	9.83
	L645	5.08	6.98	13.41
accA2 (SCO4921)	M145	1.60	5.84	11.98
	L645	0.76	4.06	9.44
SCO4922	M145	0.01	0.03	0.05
	L645	0.01	0.02	0.08
ackA (SCO5424)	M145	2.45	6.40	9.12
	L645	4.02	7.83	11.46
leuA (SCO5529)	M145	10.03	15.72	15.28
	L645	10.29	13.81	22.18
SCO6027	M145	0.39	1.52	1.43

	L645	0.59	0.67	1.68
SCO6026	M145	1.8	5.94	7.57
	L645	1.81	6.26	10.65
ssgA (SCO3926)	M145	5.20	14.33	18.65
	L645	9.26	54.01	29.49
ssgB (SCO1541)	M145	0.21	2.79	6.00
	L645	17.43	14.33	18.99
ssgC (SCO7289)	M145	0.09	0.87	3.18
	L645	0.32	5.39	9.79
ssgD (SCO6722)	M145	16.25	41.52	176.93
	L645	5.11	30.91	203.25
ssgE (SCO3158)	M145	1.21	4.07	14.89
	L645	10.92	22.22	34.49
ssgF (SCO7175)	M145	0.16	1.29	2.41
	L645	0.45	0.57	2.25
ssgG (SCO2924)	M145	0.07	0.45	0.97
	L645	0.23	0.75	1.38
ssgR (SCO3925)	M145	0.15	0.47	0.91
	L645	0.24	0.98	1.40
rpsL (SCO4735)	M145	99.92	99.09	97.81
	L645	99.78	99.07	98.22

Table 4.1: Expression levels of transcripts in M145 and L645 strains at various time points. Transcriptional abundance for each gene was calculated as mentioned earlier.

The calculation of differences in abundance using the equation $2^{\Delta CT}$ assumes that the abundance of the PCR products during the exponential phase does indeed double with every cycle under the conditions used. To check that this is indeed the case, we diluted serially a PCR product containing the region amplified by the primer pair used to analyse *rpsL*, *i.e.* the binding sites for the primer pair (SCO4735p3 and SCO4735p4) used for analyse *rpsL* were nested within a larger fragment that had itself been produced by PCR. The dilutions were then analysed by qPCR and the Threshold Cycle (CT) determined for each dilution and plotted versus the log of the DNA quantities (Figure 4.8). This produced a straight line with almost no scatter (R^2=0.999). From the gradient of the slope (- 3.29), the efficiency was calculated to be 1.0, which equates to 100%.

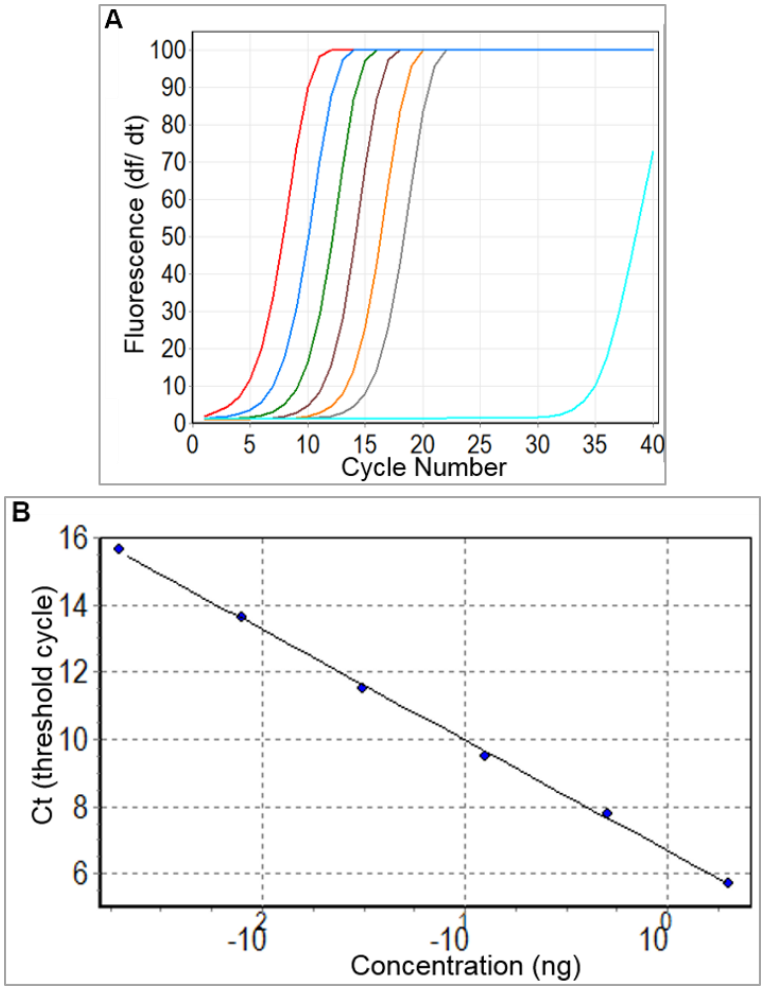

Figure 4.8: A Standard curve from *rpsL* PCR product. Panel A shows kinetic amplification of *rpsL* region, which generated from *rpsL* gene using SCO4735p1 and SCO4735p2 primers. Reactions were done in Corbett Rotor-Gene 6000 system using SensiMix™ SYBR No-ROX kit. The red, blue, green, brown, orange and grey solid lines correspond to reactions to which 2, 0.50, 0.12, 0.03, 0.78 x 10^{-2} and 0.20 x 10^{-2} ng template were used, respectively. The cyan solid line is a "non-template control" (NTC). **Panel B** shows standard curve that obtained from plotting threshold cycle vs 4-fold serial dilution of known concentration of *rpsL* DNA fragments. PCR efficiency was 1.0%, which calculated using the following equation; efficiency = 10(-1/slope) -1.

4.2.3 Actinorhodin Extraction and its effect on AtrA mobility

4.2.3.1 TLC purification of actinorhodin

The finding by Bin Hong and colleagues (Peking Union Medical College, China) that the DNA-binding activity of AtrA-gl could be inhibited by actinorhodin as well as a biosynthetic intermediate of lidamycin (see Section 4.1) prompted study of the regulation of *S. coelicolor* AtrA by small molecule(s). As an obvious candidate was actinorhodin, small molecules were extracted from an overproducing strain, L646, which constitutively expresses *act*II-ORF4 (Towle, 2007). Mycelia were grown on the surface of TSA plates for 6 days, a point at which the production of blue pigment was conspicuous. The mycelia and agar were mashed to a pulp in water before ethyl acetate was added to extract actinorhodin along with other polar organic compounds. The ethyl acetate layer was then concentrated, dried, weighed and the crude extract resuspended in a small volume methanol. For details of the above, see Section 2.17.1. Prior to further purification, the crude extract was analysed for activity that inhibits the binding of AtrA-c and AtrA-gl binding to Site 2 of the *act*II-ORF4 promoter (Figure 4.9 panel A and B). Approximately 50% of the binding activities of both proteins were inhibited when crude extract was added to a final concentration between 1 and 2 ng/ µL (see lanes 4 and 5, respectively). The addition of methanol (the solvent) had no effect on the DNA-binding activity of AtrA (Figure 4.9, panel C).

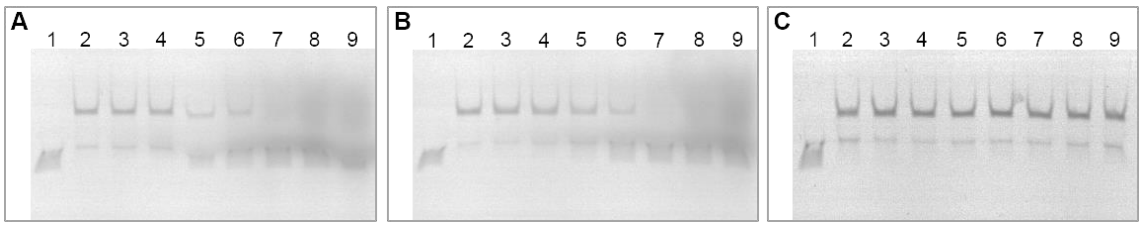

Figure 4.9: The effect of crude chemical extract from L646 on AtrA-c and AtrA-gl binding activities. Panel A illustrates effect of various concentration of crude extract on binding of AtrA-c to *act*II-ORF4 region 2. Lane 1 to 9 contains crude extract at following concentration; 0 ng, 0 ng, 10 ng, 20 ng, 40 ng, 80 ng, 160 ng, 320 ng and 640 ng. Reaction mixture in each lane contains 5 nM of substrate (*act*II-ORF4 region 2), 2 µg salmon sperm DNA and 150 nM of AtrA-c except lane 1. Labels in **panels B** and **C** are as panel A except that the AtrA-c was replaced with AtrA-gl in panel B and crude extract replaced with the solvent in panel C. Conditions are as Figure 4.3.

Having identified an inhibitory activity in the crude extract, actinorhodin was further purified by thin-layer chromatography using a toluene: glacial acetic acid (4:1, v/v) mobile phase. Two distinct red pigmented bands were readily visible (Figure 4.10, see right inset). The top red band, which had an R_f of 0.6, was removed from TLC plate, extracted with EtOAc, dried, weighted and then stored at 4°C. LC-Mass spectrometry analysis revealed that the majority of the purified material had an m/z value of 618.6 (Figure 4.10). However, the molar mass of actinorhodin ($C_{32}H_{26}O_{14)}$ is 634.54 g mol^{-1}. The short fall of 15.9 g mol^{-1} probably corresponds to a single oxygen atom and suggests incomplete tailoring of the backbone. Addition of the $C_{32}H_{26}O_{13}$ derivative of actinorhodin to DNA-binding assays had no discernible effect on the activity of AtrA at the highest concentration of 32 ng/ µL (1 mM) (Figure 4.10, see left inset).

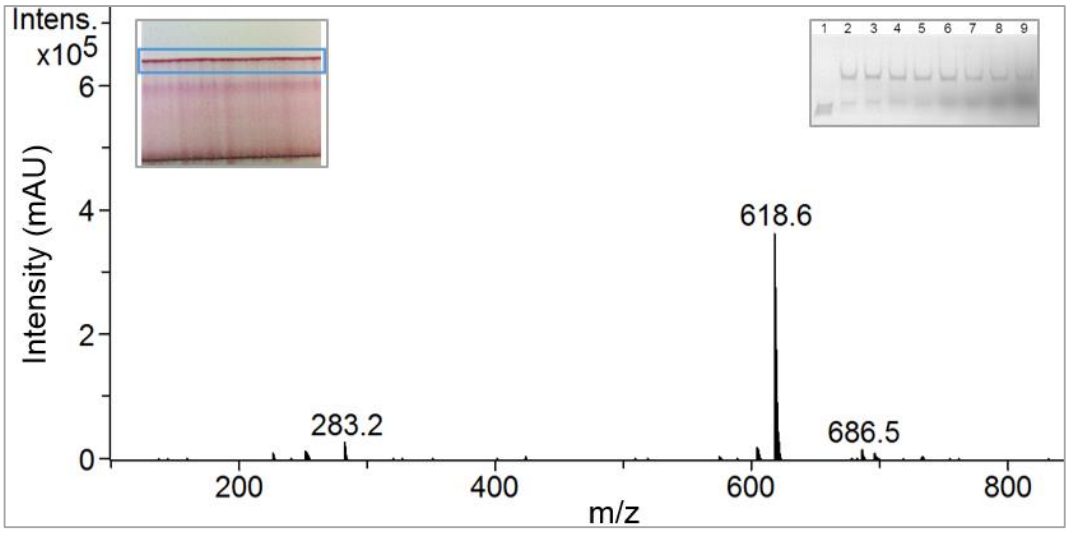

Figure 4.10: Mass spectrometry analysis of actinorhodin extracted from M145 stain. The inset on the top left hand illustrates the position of actinorhodin band that been removed from TLC plate (indicated by blue box). Under our experimental conditions, LC-MS displays a product with retention time of 2.12 min and 618.4 (m/z), which is believed belonging to ACT. The inset on the right hand side shows crude extraction interaction with AtrA-c. Conditions are as Figure 4.3 and labels are as Figure 4.9, panel A.

To determine whether *bona fide* actinorhodin (and perhaps some of its derivatives) was the source of the activity that inhibits AtrA, the crude extract of M1146, a strain that lacks the entire gene cluster for actinorhodin plus those of undecylprodigiosin, the

calcium-dependent antibiotic and the "cryptic" polyketide (Gomez-Escribano & Bibb, 2011). This approach was preferred to assaying the two pigmented bands observed for the extract of L646 by TLC. The analysis was expanded to include two additional controls, the binding of *E. coli* TetR to the promoter region of *tetA* (Meier *et al.,* 1988), which encodes the efflux pump that confers tetracycline resistance (Hillen & Berens, 1994), and the binding of ActR to the promoter region of *actA* (Caballero *et al.,* 1991), which encodes the efflux pump for actinorhodin (Fernandez-Moreno *et al.*, 1991). The ActR/ActA repressor/efflux pump pair is encoded within the *act* biosynthetic cluster (Tahlan *et al.,* 2007). Intermediates in the actinorhodin biosynthetic pathway bind ActR, thereby relieving inhibiting is DNA-binding activity and in turn "inducing" expression of *actA*. The addition of crude extract from M1146 was able to inhibit the DNA-binding activity of AtrA-c, but not TetR nor ActR, at the highest concentration used (32 ng/ µL) (Figure 4.11, panels A, B and C). This result indicated that *S. coelicolor* produces a compound that specifically inhibits AtrA and is not encoded by the *act* cluster (nor indeed the *red*, *cda* and *cpk* clusters). The addition of tetracycline inhibited the DNA-binding activity of TetR, but not AtrA nor ActR (Figure 4. 11, panels D, E and F). The IC_{50} for TetR was 250 nM of tetracycline (in the presence of Mg^{2+}). This value is $1 \pm 0.2 \times 10^{8}\ M^{-1}$ compared with published data (Kamionka *et al.*, 2004).

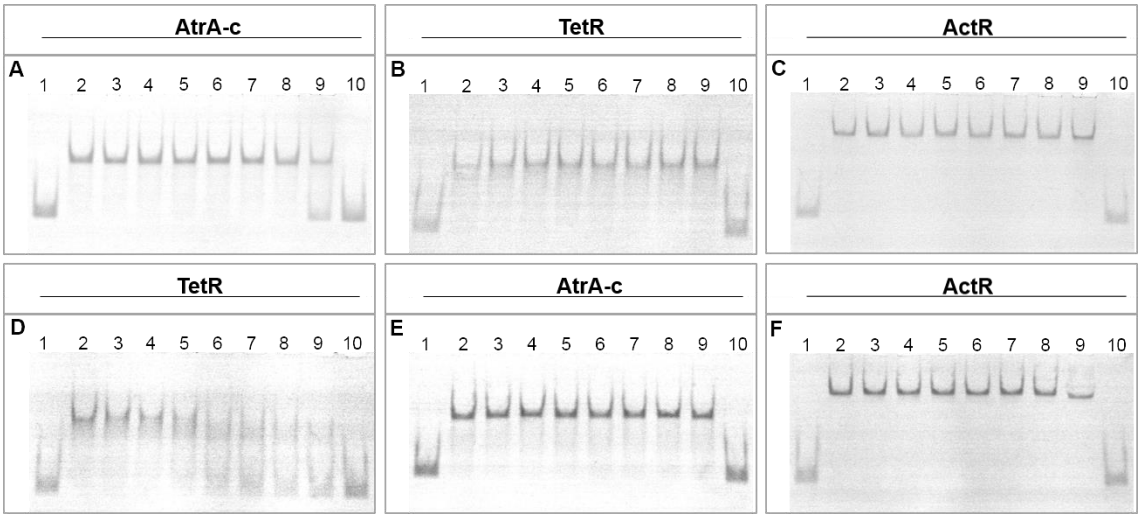

Figure 4.11: The effect of crude chemical extract from M1146 and tetracycline on TetR family regulator binding activities. 150 nM of TetR family regulator protein was used in each reaction with 5 nM of their specific substrate and 2 µg salmon sperm DNA

in 20 µL 1 x TGED reaction buffer. **Panels A**, **B** and **C** represent M1146 crude extract activity. Lanes 1 and 10 serve as negative controls (without protein), lane 2 contains water and lane 3 contains solvent (methanol). Lanes 4 to 9 contain the crude extract as follows; 20 ng, 40 ng, 80 ng, 160 ng, 320 ng and 640 ng, respectively. **Panels D**, **E** and **F** represent tetracycline activity. Lanes 1, 2 and 10 remain same as above. Lanes 3 to 9 contain 31.25 nM, 62.5 nM, 125 nM, 250 nM, 0.5 µM, 1 µM and 2 µM of tetracycline with 5 mM Mg^{2+}, respectively. Conditions are as Figure 4.3.

4.2.3.2 HPLC fractionation of actinorhodin

Encouraged by the finding that M1146 produces a compound that inhibits AtrA, a crude extract was further fractionated by preparative reverse-phase chromatography using a C18 column and a methanol. This fractionation made use of a crude extract from strain L646, which constitutively expresses *S. coelicolor atrA* (Towle, 2007). Four distinct peaks were identified and collected along with the flow through (Figure 4.12, panel A). The material in each fraction was concentrated, dried, weighed and then dissolved in methanol to a concentration of 9.5 µg μL^{-1}. Each fraction was then added to each of the three binding reactions at concentrations ranging from 1 to 32 ng μL^{-1} (Figure 4.13, panels A to F, respectively). In addition to the fractions described above, crude extract from L646 as well as M511 were assayed (lanes 4 and 10, respectively). At the lowest concentration of 1 ng μL^{-1} (panel A), an activity was detected in the crude extract of L646 (lane 4) that could inhibit the DNA-binding of TetR and to a lesser extend AtrA, but not ActR (panels B to F). The inhibition conferred by the crude extract of L646 is not necessarily associated with a single compound. Similar results were obtained for the crude extract of M511 (lane 10) in as much as inhibition of TetR was detected before AtrA and ActR at a concentration of 4 ng μL^{-1} (panel C). At concentrations of 16 ng μL^{-1} and above, the crude extract also inhibited AtrA and ActR (panels E and F). The inhibitory activity(ies) in the crude extract of M511 is lower than that of L646. At concentration of 2 ng μL^{-1} (panel B), an activity(ies) was detected in the fractions corresponding to peak 3, peak 4 and the flow through (lanes 7, 8 and 9 respectively) that inhibited the activity of TetR. Activity in fractions corresponding to peak 4 and the flow through (lanes 8 and 9, respectively) also inhibited to some extent the activity of AtrA and to a lesser extent ActR (panel B). At a concentration of 8 ng μL^{-1}

and above (panels D, E and F), the activity(ies) in peak 3, peak 4 and the flow through (lanes 7, 8 and 9 respectively) completely inhibited TetR, AtrA and ActR.

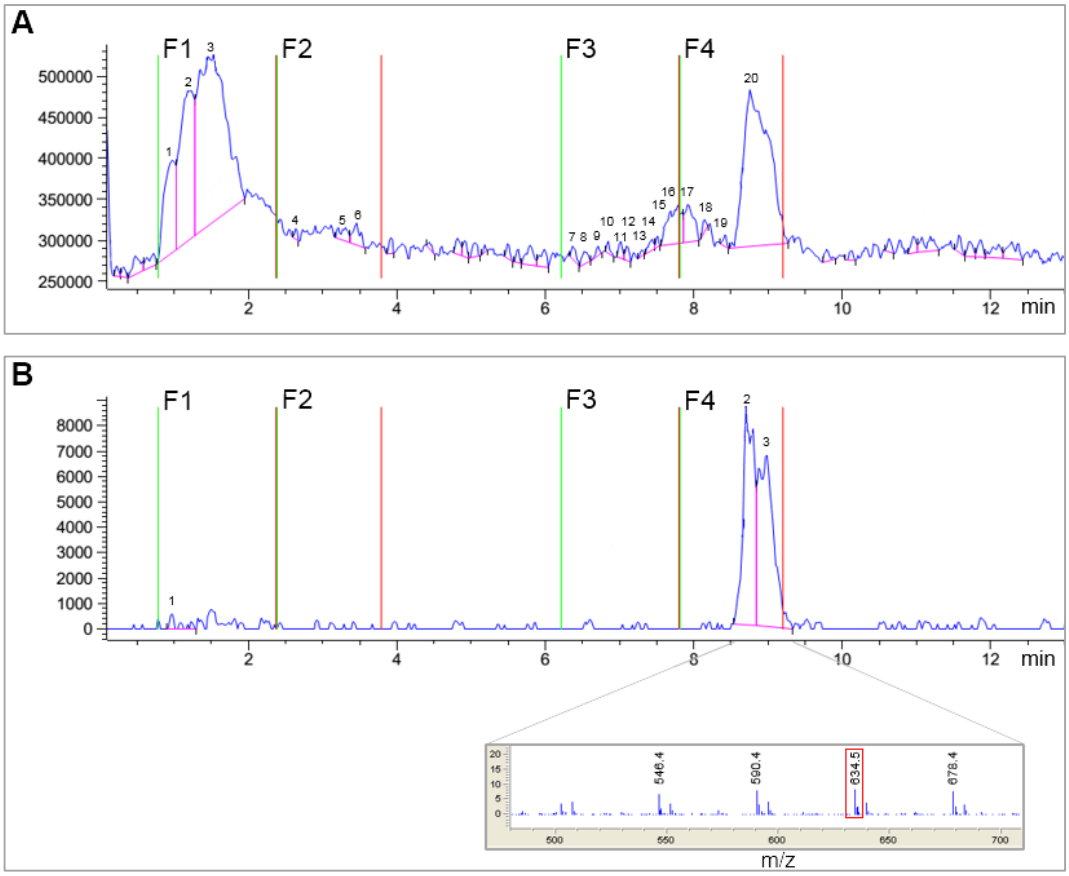

Figure 4.12: Fractionation of crude extract from L646 strain. Panel A represents a total ion chromatogram (TIC) for crude extract fractionations. Numbers from 1 to 20 on top of each peak are refer to retention time in each fraction, which indicated by F as follow; 0.97, 1.186, 1.521, 2.637, 3.203, 3.451, 6.371, 6.524, 6.707, 6.840, 7.011, 7.1, 7,304, 7.435, 7.508, 7.785, 7.924, 8.145, 8.421, 8.76, respectively. Y axis displays the sum of intensities of all ions observed at any point in time. Fractions are collected, when at least one peak detector detects a peak. **Panel B** represents extracted ion chromatogram (EIC) at 635 m/z. Y axis displays the intensity of 635 m/z at any point in time. Number 1, 2 and 3 refer to 0.962, 8.704 and 8.979 min, respectively. Fraction 4 that started from 7.6814 to 9.0739 min has been labelled to show the peak composition. Region around m/z 634.2:635.5 at 8.701 retention time revealed a peak at 634.5 m/z which is likely to be ACT (indicated by red box).

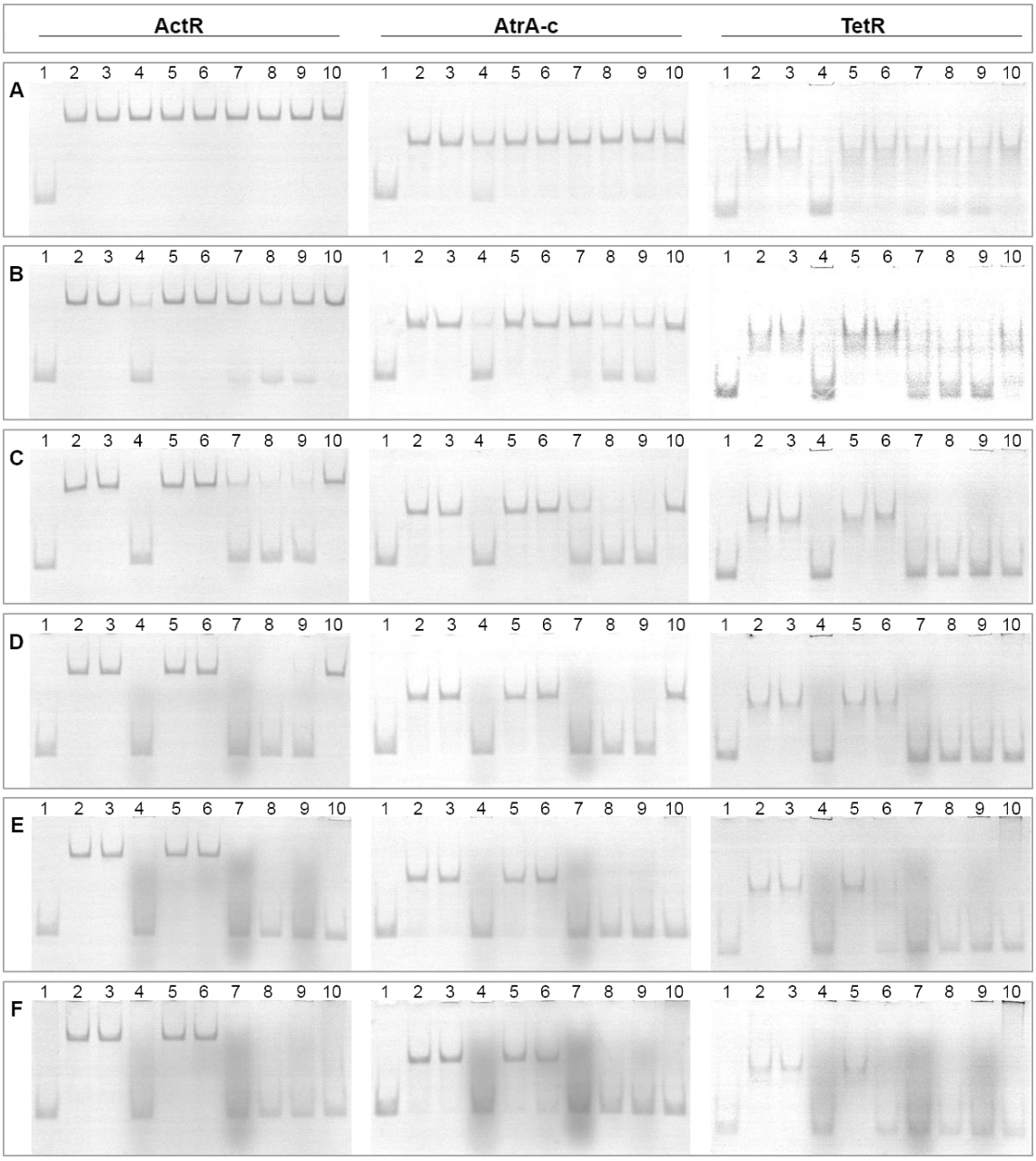

Figure 4.13: Interaction of crude extract HPLC fractions, L646 and M511 crude extract with TetR family regulator binding activities. Each reaction in **panels A** to **F** contain 20 ng, 40 ng, 80 ng, 160 ng, 320 ng and 640 ng of relative fraction, respectively. 20 µL reaction was composed of 150 nM of specific TetR-family protein was used in individual reactions, which labelled at the top of each gel image, with 5 nM of their specific substrate and 2 µg salmon sperm DNA, all made in 1 x TGED buffer. Lane 1 serves as negative control (does not contain any protein), lane 2 serves as positive control (containing water), and lane 3 contains solvent (methanol). Lanes 4 to 10 contain crude extract from L646 strain, fraction 1, fraction 2, fraction 3, fraction 4, flow through and crude extract from M511 strain, respectively. Conditions are as Figure 4.3.

4.3 Discussion

It is shown within this chapter that AtrA binds specifically to sequences in the upstream region of *ssgR*, which the PREDetector algorithm (Hiard *et al.,* 2007) revealed was likely to contain an AtrA-responsive element (GGAACCACCGGTTCC; −289 to −275 relative to the predicted start of *ssgR*). The inverted repeats in this example are perfect with 5 out of 6 matches to the original consensus GGAATG/C for the repeats (Uguru *et al.,* 2005). The translational start codon of *ssgR* is incorrect in current annotation of the genome (Kim *et al.*, 2015) and is likely the TTG triplet 123 nt farther upstream (start at chromosomal position 4318125). Proteomics experiments have unequivocally shown that the SsgR protein is at least 19 aa longer than previously predicted (Van Wezel, pers. comm.). In addition to identifying AtrA binding to the promoter region of *ssgR* (Figure 4.3, panel H), disruption of *atrA* was shown to reduce transcription of *ssgR* and *ssgA in vivo* (Figure 4.7). Taken together with the DACA data, our data indicate that AtrA activates the transcription of *ssgR*, the gene product which is in turn required for the transcriptional activation of *ssgA*, the best-studied SALP, which has a crucial role on septation and the morphology aerial hyphae (Traag & van Wezel, 2008).

Initially it was thought that *ssgR* and consequently *ssgA* would be under the regulation of the *whi* (white) genes are unable to complete the developmental process to form mature chains of spores; however, as part of a report that included the above findings (Kim *et al.,* 2015), it was found that this is not the case. The likely explanation is that SsgA is not only needed for sporulation-specific cell division, but during germination and tip growth and branching of the vegetative hyphae when the *whi* genes are not yet expressed (Noens *et al.*, 2007). However, as originally reported, activation of *ssgR* by *atrA* is not essential under at least some conditions for the production of aerial hyphae and spores. However, others have found that morphological development appears to be impeded (van Wezel, pers. comm.). AtrA can probably be best described as a moderator of *ssgR* expression, functioning much like the c-AMP receptor protein of *E. coli* to enhance, rather than determine, the expression of genes within its regulon.

It was also shown within this chapter that AtrA binds specifically to a sequence in the 186 bp upstream region of SCO5529/*leuA2* with an affinity (K_d' value of ~ 150 nM; Figure 4.3, panel F) similar to that for Site 2 in the *act*II-ORF4 promoter region. The predicted binding site for AtrA in the promoter region of SCO5529/*leuA2* (CCGGAATGACCGGTTCCAC) is located −34 to −52 relative to the predicted start. The inverted repeats in this example perfectly matches the original consensus for the repeats (Uguru *et al.,* 2005). Binding of AtrA was also detected in the 165-bp upstream region of SCO6027 (-251 to -87 relative to the start of the gene), but the affinity was lower (K_d' value of 1.25 µM; Figure 4.3, panel G). This was not unexpected as the sequence in the promoter region of SCO6027 is not as good of a match as that of *leuA2* to the PWM of AtrA. The binding of AtrA to the promoter of genes encoding enzymes that directly utilise and produce acetyl-CoA, respectively, suggest that it has a direct role in coordinating the utilisation of acetyl-CoA for primary and secondary metabolism. The reach of AtrA regulon may also extend to SCO4922/*aacA2*, which encodes a subunit of acyl-CoA carboxylase complex, if as predicted *act*II-ORF4 binds the promoter region of SCO4922/*aacA2*. The potential points at which AtrA regulates acetyl-CoA metabolism are shown in Figure 4.14. Analysis of the expression of SCO6027 and *leuA2* by RT-qPCR did not reveal any more than 1.5-fold differences between M145 and L645 (Table 4.1). Thus, while the expression of these genes does not appear to be highly dependent on AtrA under at least the growth conditions of the analysis, this is not evidence that AtrA does not have a physiologically important role in acetyl-CoA metabolism. What does appear to be clear is that not all genes with promoters bound by AtrA are regulated in the same way as *act*II-ORF4. This is not surprising given that the activities of most promoters, at least in *E. coli*, are governed by multiple transcription factors (Kao *et al.*, 2004). One of the next steps would be to insert promoters bound by AtrA (with and without the predicted site disrupted) in front of reporter genes as a transcriptional fusion and assay their activity under a variety of conditions. Conditions in which normal promoter activity is found to be dependent on AtrA would be selected for further study by ChIP-seq, for example.

One of the most exciting findings stemming from this study is evidence that a small molecule(s) regulates AtrA. A preparation of solvent extracted material from strain M1146 was found to inhibit the DNA-binding activity of AtrA, but not that of two controls: *E. coli* TetR (Ahn *et al.*, 2007) or *S. coelicolor* ActR (Figure 4.13). Both controls were also members of the TetR family (Cuthbertson & Nodwell, 2013). The extent to which the inhibitor was specific for AtrA is as yet unclear as inhibition of the DNA-binding activity of AtrA was only detected at highest concentration of crude extract used in the analysis. Nevertheless, the source of the inhibition is not actinorhodin nor undecylprodigiosin, the calcium-dependent antibiotic and the "cryptic" polyketide as the clusters for these secondary metabolites are deleted in M1146 (Gomez-Escribano & Bibb, 2011). This finding meets with initial expectations as, while AtrA is found within all streptomycetes, at least to our knowledge, these clusters are not. Further fractionation of a solvent-extracted material from L646 was conducted, this revealed what appears to be a second inhibitor; one that is more potent against TetR than either AtrA or ActR (Figure 4.13). These results are in agreement with the findings reported by Yim *et al.* (2007) that some antibiotics act as signal molecules to modulate gene transcription related to SMs production and physiological function.

One of the next steps would be to fractionate solvent extracted material from M1146 and M114 with the aim of confirming and identifying the inhibitor with specificity for AtrA. This would be approached using reverse-phase chromatography in combination with mass spectrometry as was used to identify a derivative of actinorhodin (Figure 4.12). The identification of a small molecule inhibitor of AtrA would provide insight into the physiology that regulates secondary metabolism and morphological development and may allow as simple means of altering the profiles of secondary metabolites produced by streptomycetes, perhaps even stimulating the production of metabolites that would otherwise remain cryptic. It would also provide a powerful tool to dissect further the role of AtrA in *S. coelicolor*. With hindsight priority should have been given to the analysis of the effects of inhibitors on *S. globisporus* AtrA. An open question is whether or not all orthologues of *S. coelicolor* AtrA are sensitive to the same small molecule(s). As indicated in the Introduction to this chapter, collaborators

have shown that the activity of *S. globisporus* AtrA is regulated by actinorhodin as well as a biosynthetic intermediate of lidamycin (C-1027). Binding of the latter, which is produced by *S. globisporus*, provides negative feedback. Moreover, as the feedback occurs at the level of AtrA and not a cluster-situated regulator it is highly that it provide a mechanism for coordinating the production of lidamycin with that of other secondary metabolites. As AtrA is evolutionarily conserved, negative feedback of the type described here may be widespread within the streptomycetes.

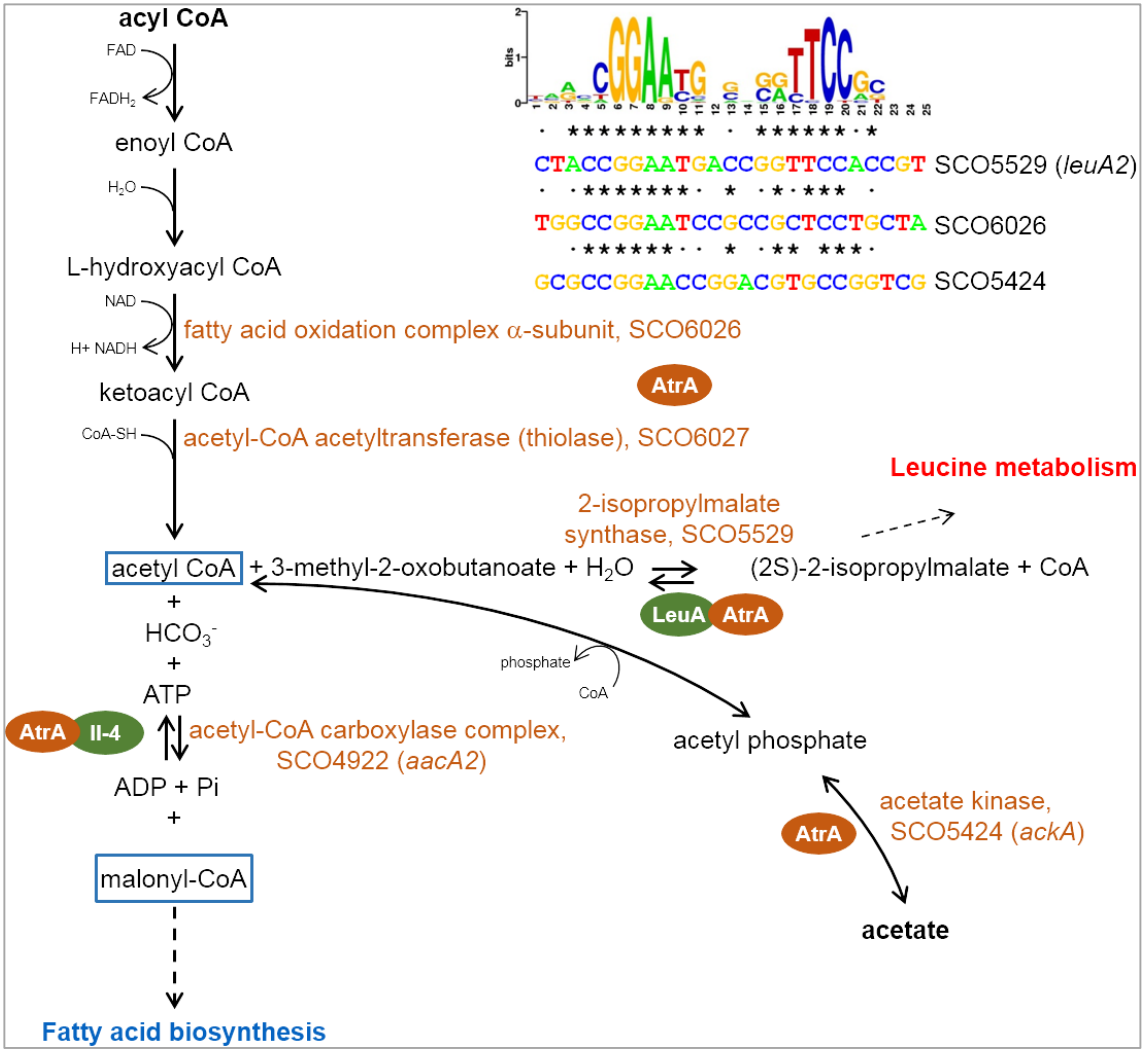

Figure 4.14: Possible link to acetyl-CoA metabolism. The position of the corresponding enzymes within the pathways linking to acetyl-CoA are shown here. An alignment of the predicted sites to the consensus sequences is shown in the inset.

Chapter 5

5 Epitope tagging of AtrA

5.1 Introduction

As part of the previous chapter it was shown that AtrA binds *in vitro* as predicted by a position weighted matrix to the promoter regions of genes associated with morphological development and primary metabolism. The latter connection is of particular interest to the natural product community as an understanding of the co-regulation of primary and secondary metabolism may inform strategies for increasing or even activating the expression of silent / cryptic biosynthetic pathways harboured by actinomycetes. However, the *in vivo* evidence for the regulation of acetyl-CoA metabolism by AtrA on its own was tenuous rather than convincing. The disruption of *atrA* in *S. coelicolor* did not produce obvious changes in the expression of genes with promoters bound *in vitro* by AtrA. While it was argued that not all promoters bound by AtrA need be as dependent on *atrA* as *act*II-ORF4 *in vivo*, it was clear that evidence for the predicted interactions occurring *in vivo* was required. Indeed, prior to the initiation of this thesis, an attempt was made to map sites of AtrA binding *in vivo* using polyclonal antibodies that have been generated with aid of Freund's adjuvant raised against AtrA as part of the ChIP-chip approach (Davis *et al.*, 2011), whereby microarrays (or chips) were used to identify DNA regions purified by chromatin immunoprecipitation. Unfortunately, these antibodies cross-reacted considerably with proteins in the extract of L465, the *atrA*-disrupted strain. In hindsight, this cross-reactivity is unsurprising since Freund's adjuvant contains crushed mycobacterial cells; its use will therefore generate polyclonal antibodies capable of binding mycobacterial proteins, many of which are similar to those found in other actinobacteria (Scherr & Nguyen, 2009). It was presumed that the cross-reacting proteins were some of the many other members of the TetR family present in *S. coelicolor* (Cuthbertson & Nodwell, 2013, Ahn *et al.,* 2012). While in theory it should have been possible to subtract this "background" to identify sites of AtrA binding in M145, the accuracy of

any subtraction would have been critically dependent on precisely matching the physiology of M145 and L645. Therefore, it was decided to explore here the tagging of AtrA with the 3 X FLAG™ tag (Einhauer & Jungbauer, 2001), an epitope that can be recognised by highly specific antibodies available commercially. The FLAG-tag is reported to have the advantage that it was designed to be hydrophilic, which in turn reduces its capacity to denature or interfere with the activity of a protein to which it is attached. Monoclonal antibodies were then raised against the designed tag. In the first iteration of epitope tagging, the tag was derived by identifying a peptide recognised by a monoclonal antibody raised against a larger protein (Hopp *et al.*, 1988). In this chapter the successful tagging of AtrA from *S. coelicolor* is reported.

5.2 Results

5.2.1 The ability of AtrA with the 3 x FLAG tag™ to complement disruption of *atrA* in the chromosome

5.2.1.1 Construction and description of plasmid

Although the 3 x FLAG tag™ is relatively unimposing, it is well established that transcription factors interact physically with each other and RNA polymerases (Dyson *et al.*, 2004). To safeguard against adverse effects of the 3 x FLAG tag™ we generated C-terminal and N-terminal 3 x FLAG tageed variants of AtrA expressed from the native *atrA* promoter (Figure 5.1, panel A). The latter was produced by the laboratory of Colin Smith (Surrey) in collaboration with the McDowall group. The coding sequence of the 3 x FLAG tag™ was codon optimised for streptomycetes (for details, see Materials and Methods, Section 2.16.1). To provide a baseline control, an untagged version of *atrA* was also produced. All versions of the recombinant genes incorporated the promoter region of *atrA*. In the case of the N-terminally tagged version generated at Surrey, the region upstream of *atrA* extended to position -426 (relative to the start codon), while for the others it only extended to position -332. The constructs designed in Leeds were shortened to facilitate their synthesis by Invitrogen. The 49-bp region immediately downstream of *atrA* was also included with the C-terminally tagged gene to allow the option of replacing the 3 x FLAG tag™ using SLiCE, a cell extract-based DNA cloning method (Zhang *et al.*, 2012). This option was never utilised, however, BamHI and EcoRI sites at the 5'and 3'ends of the *atrA* genes designed at Leeds were also included to facilitate cloning. EcoRV cutting sites were added in the front of BamHI and EcoRI sites at the 5'and 3'ends of the *atrA* genes designed at Surrey to allow the option of blunt cloning.

All three version of *atrA* were cloned into the BamHI and EcoRI of vector pAU3-45 (Bignell *et al.,* 2005), a derivative of pSET152 (Bierman *et al.,* 1992) that confers resistance to thiostrepton. The pSET152 vector could not be used to set up a complementation assay as it shares the apramycin resistance gene with the cassette used to disrupt the chromosomal *atrA* gene (Uguru *et al.,* 2005), i.e. L645 is already resistant to apramycin (Uguru *et al.,* 2005). The constructs with N and C tagged and untagged *atrA* were designated pAU-3xatrA, -atrA3x and -atrA. Putative plasmid clones were verified by BamHI and EcoRI restriction digest (Figure 5.1, panel B) and the integrity of cloned inserts was verified by sequencing (Figure S2, S3 and S4). All three of these construct and the empty pAU3-45 vector were separately introduced into L645 by interspecies conjugation, as described previously (Gust *et al.,* 2003).

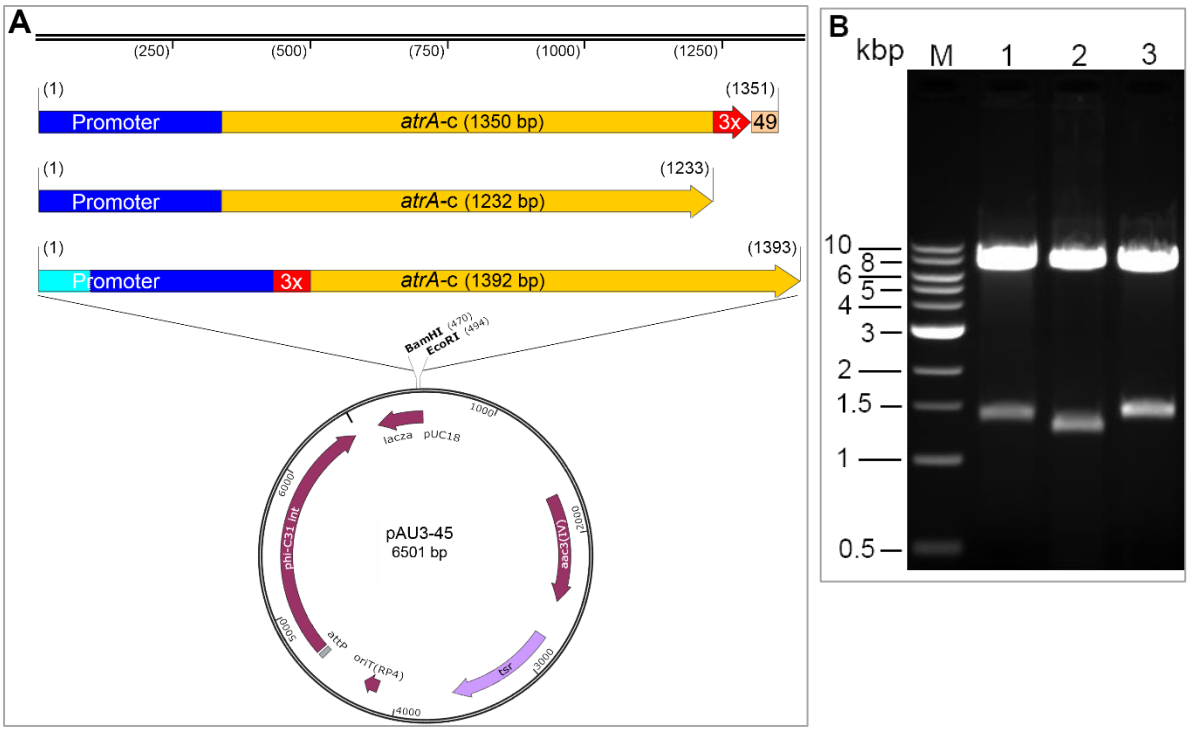

Figure 5.1: Schematic representation and enzymatic analysis of constructs based on pAU3-45. (A) The position of the 3 x FLAG tag™ is labelled "3x" and depicted by a box filled red, while the promoter region and coding regions are depicted by boxes filled blue and gold, respectively. The boxes filled turquoise and tan indicates the position of the 94 bp of upstream sequence present only in pAU-3xatrA, and the 49 bp of downstream sequence present only in pAU-atrA3x, respectively. The numbers in

parenthesis indicated the length of corresponding BamHI and EcoRI fragments (for further details, see Section 2.16.1). **(B)** Products of digestion with BamHI and EcoRI. Lanes 1 to 3 correspond to pAU-atrA3x, -atrA and -3xatrA, respectively. Lane M corresponds to 1kb DNA ladder (New England Biolabs). Numbering on the left indicate the sizes of representative markers.

5.2.1.2 Analysis of transconjugants and chromosomal integration

Spores and mycelia from the colonies of multiple putative transconjugants of L645 were streaked individually onto the surface of SFM plates containing thiostrepton to eliminate any false positives. In our hands thiostrepton is not as selective as apramycin when using *S. coelicolor*. Transconjugants confirmed as thiostrepton resistant were then isolated and analysed by PCR to confirm their integration at the expected site and their identity. Integration at the expected site was confirmed using a forward primer that targets a region upstream of the ΦC31 *attB* site and a reverse primer that targets a region downstream of the ΦC31 *attP* site in the pAU3-45 plasmid. Using this primer pair (Figure 2, panel A), the expected amplicon (1,635 bp) was produced for each of the four transconjugants (lanes 7 to 10), but not the parental strain L645 nor M145 (lanes 6 to 1). The presence of the original *atrA* disruption was confirmed using a forward primer that binds a site downstream of chromosomal *atrA*, but outside the fragment inserted into pAU3-45, and a reverse primer than binds within the apramycin cassette. Using this pair (panel B), an amplicon (800 bp) was produced for all four of the transconjugants (lanes 7 to 10) and L645 (lane 6), but not M145 (lane 1).

The status of the 3' ends of the recombinant *atrA* genes delivered by pAU3-45 was confirmed using a forward primer that binds with the 3'end of the coding region of *atrA* and a reverse primer that binds within the pSET152 backbone of the delivery plasmid (panel C). Amplicons were produced for transconjugants containing pAU-atrA3x, -atrA and -3xatrA (lanes 8 to 10), but not pAU3-45 (lane 7) nor L645 (lane 6) and M145 (lane 1). Moreover, the amplicon corresponding to pAU-atrA3x (lane 8) migrated more slowly confirming the presence of the 3 x FLAG tag™ at the 3' end of *atrA* (304 vs 186 bp). The difference is size reflects the presence of downstream sequences of *atrA* as well as the 3 x FLAG tag™ in pAU-atrA3x. Similarly, the status of

the 5'ends of the recombinant *atrA* genes delivered by pAU3-45 was confirmed using a forward primer than binds with the promoter region of *atrA* and a reverse primer that binds within the coding region of *atrA* at a site disrupted in L645 (panel D). Amplicons were produced for transconjugants containing pAU-atrA3x, -atrA and -3xatrA (lanes 8 to 10), but not pAU3-45 (lane 7) nor L645 (lane 6) and M145 (lane 1). Moreover, the amplicon corresponding to pAU-3xatrA (lane 10) migrated more slowly confirming the presence of the 3 x FLAG tag™ at the 5'end of *atrA* (274 vs 205 bp).

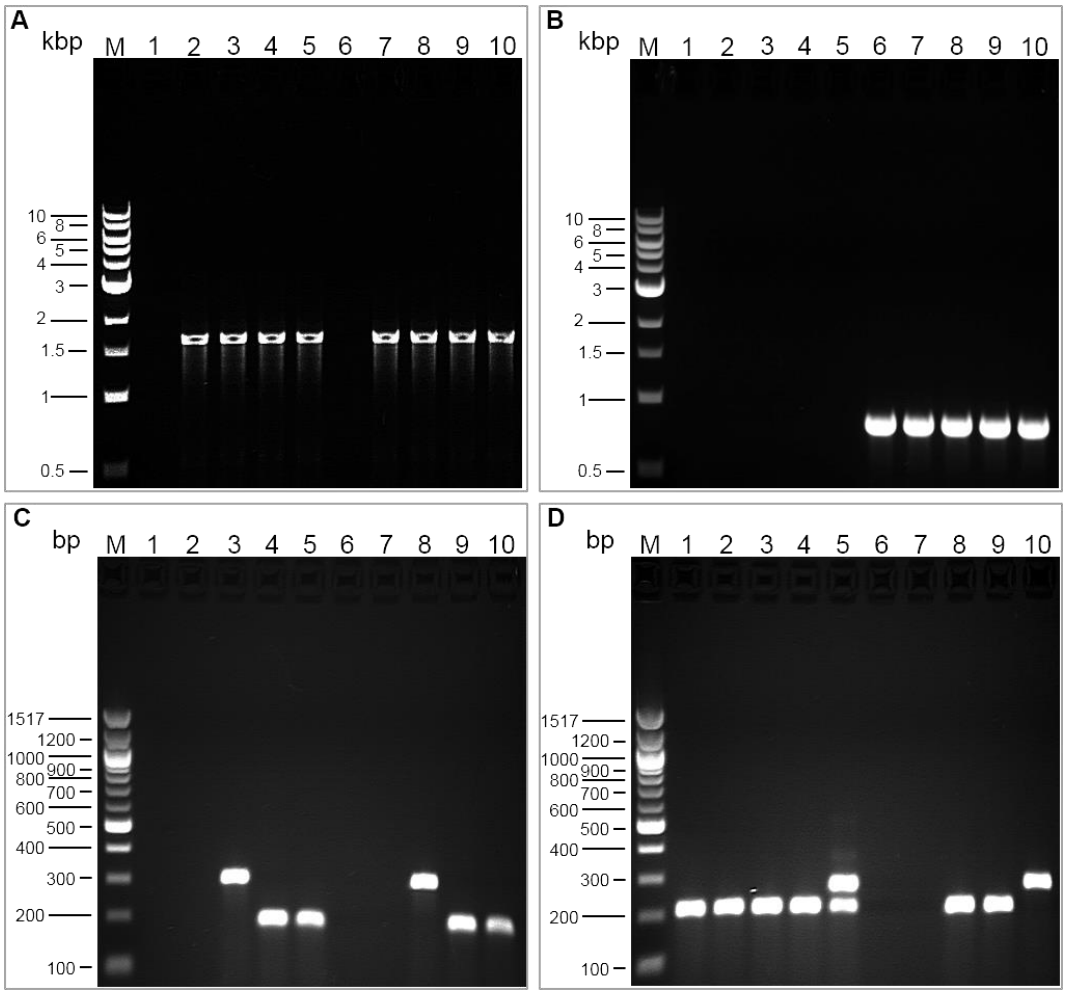

Figure 5.2: Confirmation of integrated plasmids construct into L645 by PCR and gel electrophoresis. (A), **(B)**, **(C)** and **(D)** assay integration of pAU3-45-based constructs at the *attB* site, the presence of *atrA* disrupted by the apramycin cassette, and the status of the 3'and 5'ends of the recombinant atrA genes, respectively. Lane 1 corresponds to M145, while lane 6 corresponds to L645. Lanes 7 to 10 correspond to transconjugants (L645) containing pAU3-45, pAU-atrA3x, pAU-atrA and pAU-3xatrA, respectively. Lanes 2 to 5 correspond to transconjugants (M145) containing pAU3-45, pAU-atrA3x, pAU-atrA and pAU-3xatrA, respectively. Lane M corresponds to the 1kb DNA ladder (New

England Biolabs). Numbering on the left indicate the sizes of representative markers. The primers used to generate the results shown in (A), (B), (C) and (D) were attB-F and attP-R, atrA56Dn-R and Apra-F, atrAF3 and pSET-R2, and 3xatrA-R and 3xatrA159-F, respectively (for sequences, see Table S2). (A) and (B) were analysed using 1% [w/v] agarose gels, while (C) and (D) were analysed using 2% [w/v] gels.

5.2.1.3 Analysis of the phenotypes of the transconjugants

The phenotypes of the transconjugants were compared against L645 and M145 when grown as circular patches (12-15 mm diameter) on R5 plates. Representative results are shown following incubation at 30 ^{o}C for 2, 3, 4 and 5 days (Figure 5.3). As reported initially (Uguru *et al.,* 2005), disruption of *atrA* blocks production of blue-pigmented actinorhodin (cf. L645 and M145). Moreover, the production of actinorhodin was restored by introducing untagged or N-terminally tagged *atrA*. The presence of 3 x FLAG tag™ at the C terminus of AtrA appears to interfere with the function of *atrA*. The presence of untagged or N-terminally tagged *atrA* also seemed to restore the production of aerial hyphae which were first obvious around the edge of patches as a white ring (see 72 hours). Thus, all the indications are that the addition of a 3 x FLAG tag™ to the N-terminus of AtrA does not interfere with the function of AtrA *in vivo*.

As part of the phenotypic analysis, we also included M145 into which pAU3-45 had been introduced. The reason for doing so was a report by a previous PhD student that the introduction of pSET152 into *S. coelicolor* substantially reduced, if not blocked, the production of actinorhodin (Jane Towle, pers. comm.). In agreement with this report, it was found that transconjugants of M145 containing pAU3-45 did not produce actinorhodin. The plasmids pAU3-atrA and pAU-3xatrA were also introduced into M145. Their inclusion showed that the introduction of genes encoding untagged or N-terminally tagged AtrA overcame the negative effects of pSET152 on actinorhodin production (Figure 5.3). The results of this analysis also suggested that pAU3-45 also interferes with morphological development as evidence by the bald phenotype (i.e. lack of white aerial hyphae) associated with patches of M145 (pAU3-45), and that the insertion of genes encoding untagged or N-terminally tagged AtrA can overcome this additional effect of pAU3-45, see M145 (pAU-atrA) and M145 (pAU-3xatrA) (Figure

5.3). PCR analysis confirmed the integrity of all of the M145 transconjugants (Figure 5.2). It should be noted that as expected no amplicon was produced for any of the M145 transconjugants when the *atrA* disruption was assayed (panel B) and two amplicons were produced for transconjugants containing pAU-3xatrA when the 5'status of *atrA* was assayed (panel D).

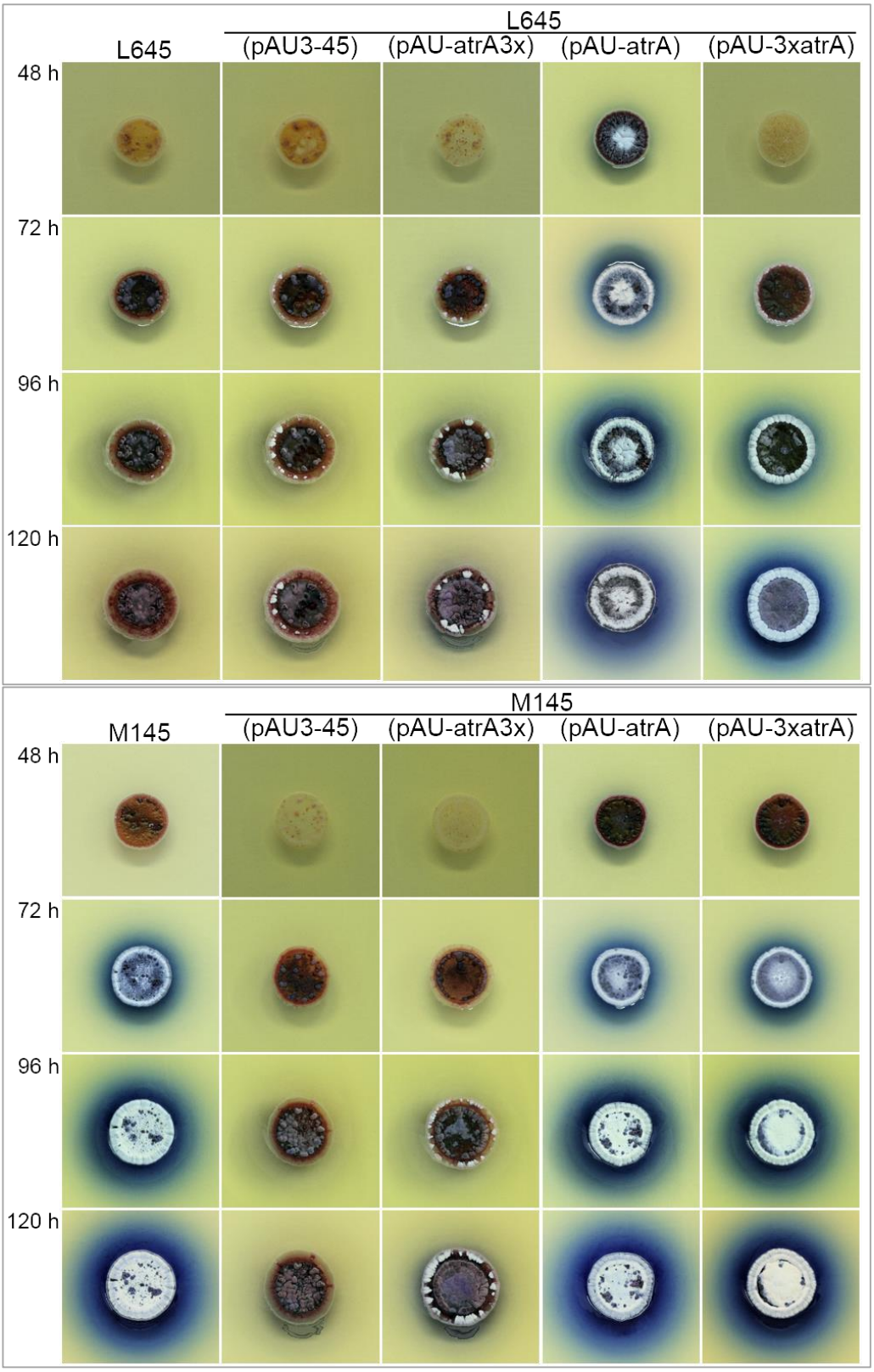

Figure 5.3: Complementation of L645 with untagged, N- and C-terminally 3 x Flag tagged AtrA. Labelling at the top indicates the strains. The plasmids introduced by

conjugation are indicated in parenthesis, where appropriate. Labelling on the left indicates the length of incubation prior to photographing of the patches. Images representative of at least five independent replicates.

5.2.2 Identification of a "phenotype neutral" vector for the delivery of tagged genes

The finding that the 3 x FLAG tag™ to the N-terminus of AtrA does not interfere with the function of AtrA *in vivo* was consistent with the finding that the addition of a His-tag™ (or decahistidine-tag) to the N-terminus of AtrA does not block DNA-binding *in vivo*. At this juncture, it was learnt that our collaborators in Beijing were planning on using AtrA tagged at the C-terminus with 3 x FLAG tag™ to map binding sites in *S. globisporus*, and delivering the corresponding gene via pKC1139, a non-integrative plasmid (Figure 5.4, panel A). To test whether pKC1139, unlike pAU3-45, could deliver recombinant genes into M145 without interfering with actinorhodin production or morphological development, the pKC1139 vector was requested and introduced into M145 by conjugation. A pKC1139-based construct encoding AtrA from *S. globisporus* with the 3 x FLAG tag™ at the C-terminus (a gift from Bin Hong) was also introduced, see pKC-atrA3x-gl (Figure 5.4, panel B). The integrity of these plasmids was confirmed by PCR and sequencing (data not shown). The vector pSET152 was also introduced to confirm the original observation that this widely used, integrative vector blocks actinorhodin production. The integration of pSET152 at the *attB* site was confirmed by PCR analysis, as described above (data not shown). Transconjugants containing the pKC1139 vector and the *atrA*-gl-containing construct, but not those containing pSET152 produced actinorhodin and developed normally (Figure 5.4, panel B).

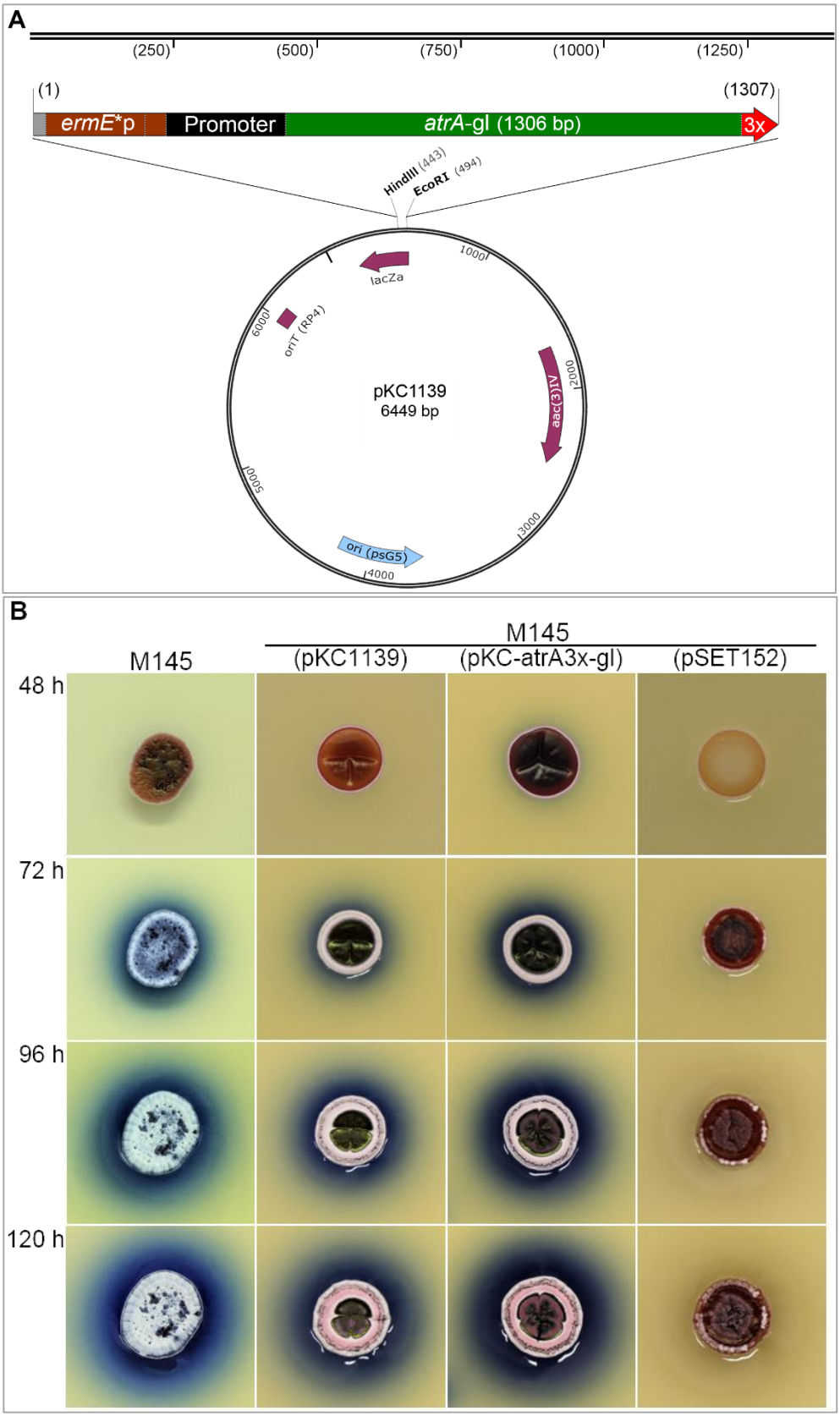
A
(250)
(500)
(750)
(1000)
(1250)
(1)
(1307)
ermE*p
Promoter
atrA-gl (1306 bp)
3x
HindIII (443)
EcoRI (494)
lacZa
oriT (RP4)
pKC1139
6449 bp
aac(3)IV
ori (pSG5)
B
M145
M145
(pKC1139)
(pKC-atrA3x-gl)
(pSET152)
48 h
72 h
96 h
120 h

Figure 5.4: Schematic representation of pKC1139 and pKC-atrA3x-gl and the phenotype of M145 when these plasmids are resident. (A) As labelled, the plasmid pKC1139 contains the origin of transfer from RP4, the origin of replication from pSG5 and the promoter and coding region of *atrA* from S. globisporus, which was inserted into pKC1139 between the NdeI and BamHI cloning sites. **(B)** Labelling at the top indicates the strains. The plasmids introduced by conjugation are indicated in parenthesis. Labelling on the left indicates the length of incubation prior to photographing of the patches. Each of the results are representative at least five independent repeats.

5.2.3 AtrA tagged with 3x Flag™ tag at C-terminus activity test

Having found that the addition of the 3x FLAG tag™ to the C-terminus of AtrA blocks complementation of the disruption of chromosomal *atrA* in *S. coelicolor*, the DNA-binding activity of AtrA tagged at the C-terminus with 3 x FLAG tag™ was assayed *in vitro*. The corresponding gene was expressed from pET16b and its product purified (Figure S5), as described previously for *atrA* without a segment encoding a tag at its 3' end (see Chapter 4). The DNA-binding activity of AtrA3x was then assayed using "Site 2" of the *act*II-ORF4 promoter region, as described previously (see Chapter 4). This revealed an apparent K_d' of 78 nM (Figure 5.5) indicating that the DNA-binding activity of AtrA in isolation is unaffected by a 3 x FLAG tag™ on its C-terminus.

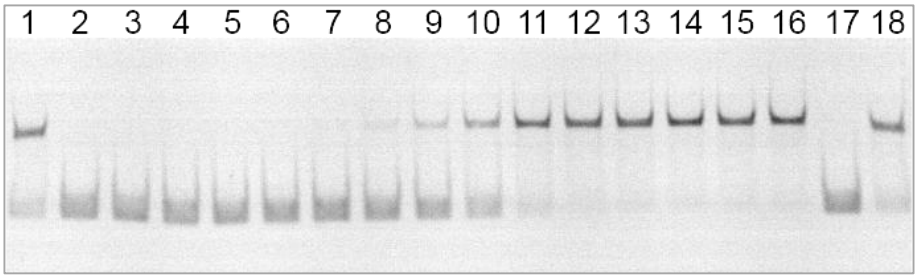

Figure 5.5: *S. coelicolor* AtrA tagged at C-terminus electrophoretic mobility shift assay. Each reaction of 20 µL 1 x TGEK contains 5 nM of fluorescein labelled *act*II-ORF4 site 2 region and 2 µg salmon sperm DNA (Invitrogen). Lanes 2 to 17 contain 0 nM, 0.6 nM, 1.2 nM, 2.4 nM, 4.8 nM, 9.75 nM, 19.5 nM, 39 nM, 78 nM, 156 nM, 312.5 nM, 625 nM, 1.25 µM, 2.5 µM, 5 µM and 0 nM of AtrA3x protein, respectively. Lane 1 and 18 contain 150 nM of AtrA, which used as positive control. The reactions were run in a 4% (37.5:1) acrylamide: *bis*-acrylamide gel in 1 x TGED buffer for 40 min at 120V. The gel was imaged using Fujifilm FLA-5000 imaging analyser system.

5.3 Discussion

The work described in this chapter was successful in showing that the addition of the 3 x FLAG tag™ to the N-terminus of AtrA does not hinder its ability to functional substitute for untagged AtrA in *S. coelicolor* (Figure 5.3). Thus, it should be possible to proceed with a ChIP analysis of AtrA using antibodies against the FLAG epitope. In terms of ChIP analysis, there are two existing strain pairs that could form part of a study; M145 with pAU-atrA or pAU-3xatrA and L645 with pAU-atrA or pAU-3xatrA. A possible disadvantage of using M145 as the background for comparison is the expression of untagged *atrA* from the chromosome, which might diminish the signal significantly. However, an advantage is that it might be easier to match the physiology of strains containing pAU-atrA or pAU-3xatrA. This would be particularly important if other 'omic analyses were integrated with a ChIP analysis of AtrA. Repeated observations suggest that in L645, while actinorhodin production and morphological development proceed with N-terminally tagged AtrA, both these process might be delayed at earlier time points (Figure 5.3).

Interestingly, the addition of the 3 x FLAG tag™ to the C-terminus of AtrA hindered its ability to substitute functionally for untagged AtrA in *S. coelicolor* (Figure 5.3). The basis of this difference is not known, but it would be relatively straightforward using polyclonal antibodies against AtrA, which have been raised by our laboratory already, or monoclonal antibodies against the FLAG epitope to check that C-terminally tagged AtrA is produced to normal levels. If so, then it would be interesting to compare DNA binding profiles of N- and C-terminally tagged AtrA. It might be that the C-terminal tag disrupts cooperative interactions with other factors that facilitate its binding to some sites. Electrophoretic mobility shift assays showed that DNA binding of AtrA3x was indistinguishable to that of untagged AtrA, thus a C-terminal 3 x FLAG tag™ does not abrogate binding to DNA, at least *in vitro* (Figure 5.5).

Consideration should be given also to the vector used to deliver recombinant *atrA* genes. The results of this work clearly shows that the widely used, integrative vector pSET152 (Bierman *et al.,* 1992), but not the non-integrative vector pKC1139 (Bierman

et al., 1992), blocks actinorhodin and morphological development (Figures 5.3 and 5.4, panel B). The finding that pSET152 causes defective or decreased antibiotic production is not new (Baltz, 1998) and has been described in the context of landomycin E production by *S. globisporus* (Luzhetskii *et al.*, 2001), tylosin production by *S. fradiae* (Voeikova, 1999), and kanamycin production by *S. kanamyceticus* (Luzhetskii *et al.*, 2000). A study by (Combes *et al.*, 2002) suggested that pSET152 can integrate at several position of genomic DNA of streptomycetes and cause chromosomal rearrangements, which may alter the production of secondary metabolites. The remarkable finding from this study is that expression of *atrA* from its own promoter, but within chromosomally integrated pSET152 can reverse the effects of pSET152 on *S. coelicolor* M145. Although the integrity of the chromosome of transconjugants was not checked as part of this study, it is unlikely that the presence of *atrA* within pSET152 would prevent pSET152-induced chromosomal rearrangements. The presence of a gene coding C-terminal tagged AtrA was certainly unable to block the effects of pSET152. An alternative explanation is that pSET152 in some ways blocks expression of *atrA* at its normal chromosomal position, either directly or further back in the regulatory pathways, but is unable to do so when *atrA* is repositioned at the *attB* site. Given pSET152 is not neutral with regard to the physiology of M145, it is recommended that pKC1139 is considered as the vector for the delivery the genes encoding untagged and N-terminally AtrA for ChIP analysis. It would also be interesting to compare the M145 and M145 (pSET152), by transcriptomics and proteomics in an attempt to better understand the effect of pSET152.

Chapter 6

6 Concluding remarks and future work

Combining differential and global RNA sequencing approaches, nucleotide-resolution transcriptome maps detailing the positions of transcriptional start sites and the boundaries of transcription have been identified in *S. coelicolor* using the UCSC microbial genome browser (Schneider *et al.*, 2006, Chan *et al.*, 2012) (see Chapter 3). These maps have already allowed us to better define the nature of one class of promoter (Kazakov *et al.*, 2007) (Figure 3.2, table S2); identify small RNAs (Figure 3.4, table S5), many of which are expected to be regulatory; and confirm that leaderless transcripts are widespread in *S. coelicolor* (Figure 3.3, table S3). They have also allowed us to identify which of the regulatory genes that have been previously annotated (Castro-Melchor *et al.,* 2010) were expressed within our sample (Figure 3.9) and are thus likely to have contributed to the observed transcription. Sequences upstream of TSSs associated with the translational machinery have been analysed and revealed two conserved hexanucleotide regions (Figure 3.10, panel A and figure S1), which are similar to the consensuses sequences reported previously for 'vegetative' promoters of streptomycetes (Strohl, 1992) and *Escherichia coli* (Figure 3.10, panel B) (Harley & Reynolds, 1987, Lisser & Margalit, 1993). These sequences can be used as a guide to predict the location of other vegetative promoters.

It should be noted that several of the RNAs annotated here as 5' regions of mRNA were previously listed as small RNAs (Vockenhuber *et al.,* 2011). The detection by us of longer transcription units may reflect an increase in the sensitivity of transcript detection using FRT RNA-seq or possibly the attenuation of transcription of the corresponding genes in the mycelial sample studied previously (Figure 3.7, panel C). Attenuation is a mechanism by which transcription from a promoter can be terminated in a regulated manner before reaching structural genes (Henkin & Yanofsky, 2002). The best studied attenuation system is the tryptophan operon of *E. coli* (Merino *et al.*, 2008). By comparing our approach with information from hundreds of transcriptome

studies, most of which to date have used microarray, allowed for the identification of genes with expression patterns that match closely those of regulators either directly or inversely, suggesting positive or negative regulation, respectively (Castro-Melchor *et al.,* 2010).

It will be interesting to analyse the network modules, groups of genes that appear to be co-regulated (Castro-Melchor *et al.,* 2010), and identify those that contain a regulator that is expressed in our sample. This would be followed by the analysis of the sequences in the vicinity of the TSSs of proposed target genes for conserved sequences that might represent a shared *cis*-regulatory element (*i.e.* the binding sites of transcription factor(s)). The latter can be done using MEME (Bailey *et al.,* 2009). It would also be interesting to determine whether binding sites predicted with knowledge of TSSs differs from those predicted by simply scanning a fixed distance upstream of the predicted start codons of target genes (Castro-Melchor *et al.,* 2010). The binding of transcription factors to predicted *cis*-elements could be investigated experimentally. Analysis of individual regulators that link to either morphological development or secondary metabolism (van Wezel & McDowall, 2011) has already revealed new layers of transcriptional complexity (see Figure 3.8). The goal in the end would be to develop automated, computer-based 'piplines' that produce maps of TSSs and TUs from raw RNAseq data because it is extremely labour intensive to assign these features manually.

Another aspect of my work was the study of the AtrA regulon and its impact on metabolism and morphological development (Chapter 4). EMSA (Figure 4.3, panel H) combined with transcript analysis (Figure 4.7) showed that *ssgR* transcription is directly activated by the developmental regulator AtrA, which also activates actinorhodin biosynthesis. This provides a novel and direct link between development and antibiotic production and also highlights the first developmental target for AtrA itself. Furthermore, our results (Figure 4.3, panel G) suggest that AtrA plays an important role in regulating precursor supply via the SCO5529/ *leuA2* gene, which encodes enzymes that directly utilise acetyl-CoA for primary and secondary

metabolism (for more details see Section 4.3). In addition to the identification of AtrA binding in the promoter regions of *ssgR* and *leuA2* (Figure 4.3, panels H and G, respectively), disruption of *atrA* was shown to reduce transcription of *ssgR* and its target *ssgA*, the best-studied SALP *in vivo* (Figure 4.7). However, analysis of the expression of *leuA2* showed lower differences between wild type (M145) and *atrA* knockout (L645) strains (Table 4.1). One of the next steps would be to construct reporter plasmids for each of the predicted AtrA binding sites by cloning the DNA-fragment containing the sequence of interest into pMU1* (Craney *et al.*, 2007) in an orientation such that *lux* expression would be driven by the inserted promoter. The activity could then be measured (bioluminescence) in the host organism under a variety of conditions.

The remarkable result that we obtained from studying small molecules was the ability of crude extract from M1146 (Figure 4.11, panel A), in which four metabolite gene clusters have been deleted (Gomez-Escribano & Bibb, 2011), to inhibit DNA-binding activity of AtrA at high concentration (32 ng/µL). However as yet, the small molecule(s) inhibition specificity for AtrA is unclear. The next step would be to fractionate a crude extract from the M1146 strain and screen the fractions against AtrA to identify the fraction containing the inhibition activity. The most active fraction could be analysed using analytical LC-MS to obtain a high quality molecular weight that can be used to acquire a formula. Finally, to determine the physical and chemical properties of the small molecule, NMR could be used.

The goal for the final results chapter was to gain a wider and better knowledge about the regulon of AtrA by performing ChIP-sequencing. The approach we undertook was successful. A 3 x FLAG tag™ was added to the N-terminus of AtrA, which does not hinder its ability to functionally substitute for untagged AtrA in *S. coelicolor* (L645 strain, Figure 5.3). Limited time prevented me from completing the experiment. In the future, it would be of interest to immunoprecipitate AtrA/DNA complexes from L645 (with pAU-3xatrA) cells that have been treated with formaldehyde by using ChIP-grade antibodies against the FLAG epitope (anti-Flag), which are available commercially.

Finally, it would be interesting to conduct ribosome profiling for the wild type (M145) and *atrA* knockout (L645) strains. This approach would give us a wider knowledge about the transcriptome and proteome. Ribosome profiling targets only mRNA sequences protected by the ribosome unlike RNA-seq, which sequences all of the mRNA in a given sample. The advantages of this technique includes the determination of actively translated mRNAs at a particular moment, provides the precise position of the ribosome on the mRNA, estimates the speed of translating ribosomes and quantitative measurement of gene expression (Ingolia *et al.*, 2012, Ingolia, 2014). Briefly, this could be preceded by nuclease digestion of the mRNA bound by ribosomes and followed by ribosome footprints, reverse transcription and deep sequencing of mRNA fragments that are protected by the ribosome (Ingolia *et al.*, 2009).

7 Supplementary figures and tables

Score	Gene ID	Function	Position	Cis-element
17.27	SCO6059	hypothetical protein SC9B1.06	-74	gcggaatctcccgttccgt
15.56	SCO0775	conserved hypothetical protein	400	acggaacgggagattcccg
15.16	SCO1657c; SCO1658, *gylR*	putative methionine synthase, glycerol operon regulatory protein	-205	ccggaacgtcccactccgg
14.93	SCO5285; *lon*	ATP-dependent protease	2395	gcggagcgtgaggttccgg
14.48	SCO4109	putative oxidoreductase	-39	ccggaatcccccgctccgg
14.18	**SCO5529, *leuA2***	**putative 2-isopropylmalate synthase**	**-52**	**gtggaaccggtcattccgg**
13.95	SCO2507, *zurM*	putative metal transport ABC transporter	850	gcggagtgcaagattccgg
13.89	SCO0509, *glk2*	glycerol kinase 2 (ATP:glycerol 3-phosphotransferase) (EC 2.7.1.30)	-17	gcggaatgtgagcttccat
13.89	SCO5987c; SCO5988	conserved hypothetical proteins	-223	gtggaacccctcattccga
13.84	SCO4029	hypothetical protein	672	ccggaatgtcaaactccgg
13.53	SCO2907, *nagE2*	putative PTS transmembrane component	-90	gcggaatcacgggttcctt
13.17	SCO6054c	putative transmembrane transport protein	17	cgggaacctccggttccgg
13.14	SCO3526	conserved hypothetical protein	1063	acggaacggatccttccgc
13.11	SCO3217, *cdaR*	transcriptional regulator of calcium-dependent antibiotic	734	acggaatgcggtattccat
13.06	SCO1251	putative integral membrane protein	159	acggaacctgaccttccag
12.85	SCO1839c	putative transcriptional regulator	-143	acggaacccctgattccag
12.66	SCO1842c	conserved hypothetical protein SCI8.27c	70	ccggagcgcgccattccgg
12.41	SCO6429	hypothetical protein	128	acggaatcagcggttccat
12.4	SCO2273	FecCD-family membrane transport protein	240	gtggaacgtgcggttcccc
12.27	SCO5461	putative secreted protein	-37	ccggaatgcgtcgctccgc
12.2	**SCO5085, *actII-ORF4***	***act* cluster (and possibly *aacA2)* activator protein**	**43**	**gcgggatgtgtaattccgc**
12.13	SCO0098c	putative transposase	-43	ccggaaggctccgttccgc
12.13	SCO0566c; SCO0567	putative membrane protein, insertion element transposase	-308	gcggaacggagccttccgg
12.13	SCO3467	transposase	-42	gcggaacggagccttccgg
12.13	SCO3538c; SCO3539	hypothetical protein, putative transposase	-445	gcggaacggagccttccgg
12.13	SCO6910c	insertion element transposase	-43	ccggaaggctccgttccgc
11.94	SCO1623	conserved hypothetical protein SCI41.06	703	ctggagcgcgagattcccc
11.86	SCO6956	putative oxidoreductase	1290	ctggaacctcgcgttctgg
11.54	SCO5084, *act*II-3	Act membrane protein	2084	caggaatgccagattctat

11.52	SCO2869	putative regulatory protein	298	caggaatcggttattccac
11.47	SCO2007	hypothetical protein	456	ccggaatcccggactccct
11.35	SCO6840c, SCO6841	hypothetical protein SC3D9.08c, conserved hypothetical protein SC3D9.09	-247	gcggaatgggagatttccc
11.32	SCO1226c	putative bifunctional protein	-169	ccggaatcaacggctccgc
11.13	SCO3867c, SCO3868	putative ferredoxin, hypothetical protein SCH18.05	-158	cggcaatcccacattccgc
11.06	SCO1403	putative membrane protein	94	ccggatcgcgacattccgg
10.99	SCO6039c	putative flavoprotein oxidoreductase	918	gcggaatgtgccgctcctg
10.97	**SCO3925, *ssgR***	**putative transcriptional regulator**	**-291**	**caggaaccaccggttccct**
10.97	SCO5445c	conserved hypothetical protein	14	cgggaaccgctggttccgg
10.96	SCO4462	putative integral membrane protein	1157	ccggaacccgcggttcccg
10.94	SCO6103c	putative acetyltransferase	-27	agggaatcatccgctccgc
10.92	SCO3968c	putative integral membrane protein	-103	acggaagatcacattctgt
10.91	SCO2917,	conserved hypothetical protein	1297	ggggaacccgtcattccga
10.87	SCO4113c	conserved hypothetical protein	698	ccggaacgggacgatccgc
10.8	SCO7139	putative aldehyde dehydrogenase	486	gtggaacttcccgctccag
10.63	SCO3472c	putative transposase remnant	113	ctggaacggttcactccgg
10.6	**SCO6027c;** SCO6028c	**acetyl-coA acetyltransferase (thiolase), putative ribonuclease**	**-194**	**ccggaatccgccgctcctg**
10.55	SCO1706c	putative aldehyde dehydrogenase	434	gtggagcgggtagttccag
10.55	SCO2948c; SCO2949, *murA*	conserved hypothetical protein, UDP-N-acetylglucosamine transferase (MurA)	-228	acggaacgcgagattcatg
10.55	SCO5666	putative aldehyde dehydrogenase	498	gtggaactacccgctccag
10.52	SCO2116	putative secreted protein	193	gaggagttcgacgttccgc
10.5	SCO6780	putative secreted protein	711	gtggaaggtcccgttcccg
10.38	SCO0646	putative TetR-family transcriptional regulator	10	ccggaacccccgccccgc
10.38	SCO3920c, *cysA;* SCO3921c	putative cystathionine/methionine gamma-synthase/lyase, low molecular weight protein-tyrosine-phosphatase	-6	gcggaatcgctcatgccgc
10.33	SCO5875c	putative potassium uptake protein	258	gcggaacgccgtactccgt
10.32	SCO4828	betaine aldehyde dehydrogenase	483	ctggaactacccgctcctt
10.31	SCO1840c	putative ABC transporter ATP binding protein.	1518	gcagaatgatccgttccgg
10.3	SCO5479	oligopeptide ABC transporter ATP-binding protein	871	caggaacgcctcatcccgg
10.26	SCO1862c	putative integral membrane protein	-57	acggaagccacgattccgg
10.26	SCO4775	serine/threonine protein kinase	992	gcggagcgctgcgttccgt
10.24	SCO3189	putative membrane protein	-238	ggggaacgccacgttccgg

10.24	SCO4350c	putative integrase	524	ccggaacgagacgtcccgc
10.21	SCO1434	putative CbxX/CfqX family protein	1674	ggggaacgggcggttcctg
10.2	SCO6840c; SCO6841	hypothetical protein SC3D9.08c, conserved hypothetical protein SC3D9.09	-188	ttggaacgaaaaattccgt
10.17	SCO6226c	hypothetical protein SC2H4.08c	1030	gcggaacgcggcagtccgg
10.15	SCO0197c	conserved hypothetical protein	5	gcggaatccgtcgttgcgg
10.13	SCO0372c	putative secreted protein	2306	gcggaacgtcacgtcctgg
10.13	SCO5352	putative arginyl tRNA synthetase	681	ggggaatcccctgttccgc
10.13	SCO6841	conserved hypothetical protein SC3D9.09	4	gaggaatcgcggattcccg
10.12	SCO5091	cyclase	363	gccgaacgtgaccttccgc
10.11	SCO3330c	putative acetoin utilization protein	903	ccggaatcaccgtctccgg
10.09	SCO1555c	putative methyltransferase	1169	caggaacgcaacgctccgc
10.05	SCO4283c	putative sugar kinase	733	gcggagtcgcccgctccgg
10.02	SCO5260c (the other *atrA*); SCO5261	secreted protein, putative malate oxidoreductase	-268	cgggaatctgcccttccgg
10.02	SCO6407	putative gamma-glutamyltranspeptidase (putative secreted protein)	18	acggaatctgacggtcctg

Table S1: PREDetector output 19-mer matrix aptamers. PWM scores express the strength of the identified motifs. Sites that have been experimentally verified are in bold.

Primer Name	Binding site	Sequence (5'→3')	Product Size (bp)	Template	Tm
Used for EMSA (Reverse primers "p2" were labelled at 5' end with fluorescent (FAM))					
actII-4p1	*act*II-ORF4	TCTCGATGTCGGCCGGTGGATGTGG	395	*S. coelicolor*	55°C
actII-4p2		FAM-TCGTGCCGCCTGAGGAGCAGCAGC			
actII-4(s2)p1		TTGGGACGTGTCCATGTAATCACC	101		
actII-4p2		FAM-TCGTGCCGCCTGAGGAGCAGCAGC			
actIp1	*act*I-ORF1	AAGAACGAGATCCGGCGCGTC	419		
actIp2		FAM-ACCCCTCCTTACCGAGCCTGCGGGC			
vdhp1	*vdh* (SCO4089)	CGATCACCGAGGACACC	386		
vdhp2		FAM-GGAGTGGAGGGCGATCACGGCCTT			
SCO4119p1	SCO4119	CTACACACCGTGGGAAAGAA	180		
SCO4119p2		FAM-AGCCGCGGGAACCAACGTCG			
SCO4921p1	SCO4921	GCCCTTACAGAGACCTTGAC	195		57°C
SCO4921p2		FAM-AGGCTCCCTCCTTGAAAC			
SCO5424p1	SCO5424	AGCAGCTCGCCGACATC	197		60°C
SCO5424p2		FAM-CGCGTCGTACTGGATCG			
SCO5529p1	SCO5529	CGCGCGCAGACCTCTC	186		
SCO5529p2		FAM-CGAAGACGTGGAACGAAT			
SCO6027p1	SCO6027	GCAGGTGGAGCAGGTG	165		57°C
SCO6027p2		FAM-ACGTAGCTGCCCAGTCC			
ssgRp1	SCO3925	GTGAGAGCCGTGGGACAC	225		58°C

ssgRp2		FAM-AGCGAATGCGGTGTCAG			
ssgBp1	SCO1541	ACGGCCAAACACTACCG	122		
ssgBp2		FAM-CGGTGCCGTTGTAAACC			
actAp1	*actA* (SCO5083)	CGTGCTCCTCATCGTATGG	102		
actAp2		FAM-CGGGTCCTCGACTATTGG			
tetAp1	*tetA*	CATTAATTCCTAATTTTTG	84	XL1-Blue MRF'	45°C
tetAp2		FAM-CATTTCACTTTTCTCTATC			
Used for gene amplification (restriction sites are underlined)					
atrAF2	*atrA*	AAGCTTGGGCTGCAGGT	1270	pAU-atrA3x plasmid	65°C
atrAEcoRI- R2		GGACTGAATTCTCACACCGGCCGCGACC			
NdelatrA-F	*atrA*3x	AAAACATATGCATGTTCAGGATTCT	981		61°C
atrABamHI-R		ACGAAGGATCCTCACTTGTCGTCGTCGT			
Used for for generation of riboprobe templates (The T7 promoter sequence is bold)					
SRP-F	SRP RNA	CCGAAAAAATAATCGCCAAC	95	*S. coelicolor*	55°C
SRP-Rt7		ATCC**TAATACGACTCACTATA** GGGACGCACCCGCCAGAG			
tmRNA-F	tmRNA	CCGAAAAAATAATCGCCAAC	301		
tmRNA-Rt7		ATCC**TAATACGACTCACTATA** GGGAATCAGGGCTTCTCCGTGT			
LDS2-F	scr3871 (Xu *et al.,* 2012)	ACACCAACAAGGCTGGACA	169		57°C
LDS2-Rt7		ATCC**TAATACGACTCACTATA** GGGGTGGGGCTAACAGGATGTGA			
LDS3-F	scr2822 (Touzain et al., 2008)	ATCACGGGGGCCAACT	114		55°C
LDS3-Rt7		ATCC**TAATACGACTCACTATA** GGGGCCTTGCACCTGCATTTC			
LDS6-F	scr2100 (Touzain *et al.,* 2008)	GGCGGGAGCGAGACAC	106		
LDS6-Rt7		ATCC**TAATACGACTCACTATA** GGGTCTCCTGCTCCCCCGTA			
Used for QPCR					
SCO3926p1	*ssgA*	GACCGCACCGACAAGCTG	105	cDNA from RT	60°C
SCO3926p2		GCTCTGTTCCTCCGCCAG			
SCO1541p1	*ssgB*	GTCGTGTGCATCGCTCT	129		55°C
SCO1541p2		ATCGAAGTGCCGGTGTT			
SCO7289p1	*ssgC*	CCCACACGAACACCACCG	85		60°C
SCO7289p2		GGCATGTGGGACGAGCTG			
SCO6722p1	*ssgD*	GGGAGAAGGTCCAGCAGA	100		
SCO6722p2		GCCACCCTGCACTACGAC			
SCO3158p1	*ssgE*	AGTTCGCGGGAGAAGGT	121		
SCO3158p2		GAGCGATACCCGTCGTG			
SCO7175p1	*ssgF*	AACGCACACGGTGACATA	139		55°C
SCO7175p2		CGAACAGCTTGGAGAGGA			
SCO2924p1	*ssgG*	CTTCTGGTGGAGGGCGTCTTC	99		60°C
SCO2924p2		GCTCAGCGCGACGAGTAC			
SCO3925p1	*ssgR*	GCTGTTCTTCCTCGGTGA	136		55°C
SCO3925p2		CCGGTACGTGGCGTAGTA			
SCO4119p3	SCO4119	CTGCAACGGAATCTCAAGG	93		60°C
SCO4119p4		GAGGAACGGCTGATAGGTC			
SCO5529p3	SCO5529	CATCGCACGGCACCTGGA	118		
SCO5529p4		GTGCTTGAAGTCGATCTCCTG			
SCO4118p1	*atrA*	GCACGTGCTCCAGATTGC	98		
SCO4118p2		CGGCGGTGCGATGAGTA			
SCO5085p1	*act*II-ORF4	TACTTCGCCCTGATCGAC	139		

SCO5085p2		CTCGCCAGAGATCCAGTC			
SCO5087p1	*act*I-ORF1	CAACGCGTACCACATGAC	118		
SCO5087p2		GTGCGCGTTGATGTAGTC			
SCO4735p3	*rpsL*	GGCAAGTGGAAGGTCAAC	111		55/ 60°C
SCO4735p4		GTCGTAGCGGTTGTCCAG			
SCO5424p3	*ackA*	GAAGGACTCGTGATGGGTAC	118		55°C
SCO5424p4		CGCTCCTCTTGTTGAGAAG			
SCO4921p3	*accA2*	AGGCGATGAAGATGGAACAG	119		60°C
SCO4921p4		TCAGTCCTTGATCTCGCAGA			
SCO4922p1	SCO4922	GCGCTGAACTGGTTCCT	85		55°C
SCO4922p2		AGGACACGGACATCACG			
SCO6026p1	SCO6026	GCCAAGAAGCTGAAGAAGAC	105		60°C
SCO6026p2		GTCGATGACGTTCTGGATCT			
SCO6027p3	SCO6027	GTTCCTGGAGCACTACGG	137		
SCO6027p4		GCTGCTCCTCGAACTGAC			
SCO4735p1	*rpsL*	AGCTACACCACCGAGTCC	429	*S. coelicolor*	55°C
SCO4735p2		CTTCAGACCGGCCTTCTT			
Used for confirmation of integrated plasmids construct into *S. coelicolor*					
attB-F	*attB* site in *S. coelicolor*	CGGTGCGGGTGCCAGGGC	1635	Transconjugants	65°C
attP-R	*attP* site in pSET152	TTCGGCGGCTTCAAGTTCGG			
Apra-F	Apramycin gene	TGGGCCACTTGGACTGAT	800		60°C
atrA56Dn-R	downstream *atrA*	CCGCGGTAATAACGCTCA			
Apra-F	Apramycin gene	TGGGCCACTTGGACTGAT	519 bp		58°C
Apra-R		CCGACTGGACCTTCCTTCTG			
atrAF3	*atrA*	ACGTGCTGCTGGTCATAG	304/ 186		
pSET-R2	pSET152	TGTTGTGTGGAATTGTGAGC			
3xatrA159-F	upstream *atrA*	AACGTGCCCCTTCTGAAT	274/ 205		
3xatrA-R	*atrA*	CGGACGACCAATGAGAAT			
Used for Sequencing confirmation					
Bec-F	*atrA*	AGCGCAATCTGGAGCAC	137 bp	*atrA*	55°C
Bec-R		GTCCTTGCTCGGAAAGC			
gl-atrA-F	*atrA*-gl	GACCGAGCAGGCGCGTA	189 bp	*atrA*-gl	61°C
gl-atrA-R		GTCTCGTCCACCGATTCC			

Table S2: Names, binding sites, sequences, sizes, templates and annealing temperatures of primers used in this thesis. All primers were designed using primer3 web-based tool.

Table S3: Transcription start sites identified for *S. coelicolor.* The corresponding values where above the upper envelope boundary (see Figure 4.1). The first column shows the TSS id, which named based on their direction forward (FWD) or reverse (RVS), followed by the TSS class (I, II or III) and first nucleotide of the TSS. Second column represent TSS direction. Fourth and fifth columns represent nucleotide positions within 8 nt of each other were classified as belonging to the same TSS. (*)

refers to data extracted from (Cipriano *et al.,* 2013). (**) refers to data extracted from (Vockenhuber *et al.,* 2011). See Excel file on attached CD.

Table S4: Leaderless mRNAs identified for *S. coelicolor*. See Excel file on attached CD.

Table S5: Identifiable 3' processing associated with tRNAs in *S. coelicolor*. See Excel file on attached CD.

Table S6: Comparison with prior RNA-seq analysis of *S. coelicolor*. (*) indicates expression verified by northern blotting (Vockenhuber *et al.,* 2011). See Excel file on attached CD.

Table S7: List of annotated and small RNAs in *S. coelicolor*. Labelling in parentheses indicates whether the sRNA is upstream (u) or downstream (d) of the nearest protein-coding gene and whether on the same (+) or opposite (-) strand. The prefix 'scr' is used for discrete RNAs of unknown function, 'as' is reserved for those that are antisense to genes, and 'rs' is served for riboswitches that appear to be active. (*) sRNA probed by northern blotting. (**) indicates sRNA detected by northern blotting. See Excel file on attached CD.

Table S8: Examples of *S. coelicolor* genes whose annotation merits review. See Excel file on attached CD.

Figure S1: Alignment of upstream of TSSs associated with genes of the translational machinery. Gaps have been introduced 6 nt upstream of the '-10 boxes' (TAnnnT) to maximise alignment to a second conserved hexanucleotide sequence (GnTTTG), which labelled as '-35 box'. TSS loci are presented as bold alongside each sequence. Underlined sequences were identified using MEME (Bailey *et al.,* 2009), while highlighting indicates nucleotide matches to the consensus sequences. See word file on attached CD.

Figure S2: Sequence of recombinant 3xatrA (pAU-3xatrA). The ruler above uppercase nucleotide sequences, which are referred to the original DNA sequences, indicates position of a particular base pair in the constructed plasmid. Lowercase nucleotide sequences indicate sequenced DNA fragment, which aligned with the original sequences (uppercase). pAU-3xatrA construct was sequenced by Beckman Coulter Genomics using primers Bec-R and Bec-F (in violent). Start codon is indicated in pink, *Lacz*-α in plum, BamHI and EcoRI sites in yellow (for further details, see Section 2.16.1). Labels are as Figure 5.1. See word file on attached CD.

Figure S3: Sequence of recombinant atrA3x (pAU-atrA3x). Labelling are as Figure S2. See word file on attached CD.

Figure S4: Sequence of atrA in pAU-atrA (pAU-atrA). Labelling are as Figure S2. See word file on attached CD.

Figure S5: Purification of AtrA3x from *E. coli* by IMAC. Panel A shows the HPLC profile for the elution step of the purification. The absorbance measure at 260 nm (blue) and 280 nm (red) are plotted against elution volume (mL). The green line shows the gradient of imidazole from 145 mM to 1 M (0 to 100%, respectively). The flow rate was 1 mL/min. **Panel B** shows SDS-PAGE analysis of fractions from different steps in the purification. Lane 1 contains cell lysate from BL21 (DE3) cells prior to IPTG induction, lane 2 contains IPTG inducted BL21 (DE3) cell lysates, lane 3 contains ultracentrifuge supernatant, lane 4 column flow through, lane 5 before dialysis, lane 6 after dialysis, lane 7 contains 3.2 µg of *S. coelicolor* AtrA. Lanes labelled M contain PageRuler Unstained Protein Ladder (Fermentas). Numbering on the left of this panel indicates the sizes of the markers (kDa). Lane 7 contained 3.2 µg of purified *S. coelicolor* AtrA. See word file on attached CD.

8 References

Aceti, D. J. & W. C. Champness, (1998) Transcriptional regulation of *Streptomyces coelicolor* pathway-specific antibiotic regulators by the *absA* and *absB* loci. *J. Bacteriol.* **180**: 3100-3106.

Ahn, S. K., L. Cuthbertson & J. R. Nodwell, (2012) Genome Context as a Predictive Tool for Identifying Regulatory Targets of the TetR Family Transcriptional Regulators. *PloS one* **7**.

Ahn, S. K., K. Tahlan, Z. Yu & J. Nodwell, (2007) Investigation of transcription repression and small-molecule responsiveness by TetR-like transcription factors using a heterologous *Escherichia coli*-based assay. *J. Bacteriol.* **189**: 6655-6664.

Ainsa, J. A., H. D. Parry & K. F. Chater, (1999) A response regulator-like protein that functions at an intermediate stage of sporulation in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **34**: 607-619.

Ainsa, J. A., N. J. Ryding, N. Hartley, K. C. Findlay, C. J. Bruton & K. F. Chater, (2000) WhiA, a protein of unknown function conserved among gram-positive bacteria, is essential for sporulation in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **182**: 5470-5478.

Aldridge, M., P. Facey, L. Francis, S. Bayliss, R. Del Sol & P. Dyson, (2013) A novel bifunctional histone protein in *Streptomyces*: a candidate for structural coupling between DNA conformation and transcription during development and stress? *Nucleic Acids Res.* **41**: 4813-4824.

Allenby, N. E. E., E. Laing, G. Bucca, A. M. Kierzek & C. P. Smith, (2012) Diverse control of metabolism and other cellular processes in *Streptomyces coelicolor* by the PhoP transcription factor: genome-wide identification of *in vivo* targets. *Nucleic Acids Res.* **40**: 9543-9556.

Anderson, T. B., P. Brian & W. C. Champness, (2001) Genetic and transcriptional analysis of *absA*, an antibiotic gene cluster-linked two-component system that regulates multiple antibiotics in *Streptomyces coelicolor*. *Mol. Microbiol.* **39**: 553-566.

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

Arias, P., M. A. Fernandez-Moreno & F. Malpartida, (1999) Characterization of the pathway-specific positive transcriptional regulator for actinorhodin biosynthesis in *Streptomyces coelicolor* A3(2) as a DNA-binding protein. *J. Bacteriol.* **181**: 6958-6968.

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner & H. Mori, (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* **2**: 2006 0008.

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Y. Ren, W. W. Li & W. S. Noble, (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**: W202-W208.

Baltz, R. H., (1998) Genetic manipulation of antibiotic-producing *Streptomyces*. *Trends in microbiology* **6**: 76-83.

Baltz, R. H., (2008) Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharmacol.* **8**: 557-563.

Barrett, T., S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L. Robertson, N. Serova, S. Davis & A. Soboleva, (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* **41**: D991-995.

Barrick, J. E., K. A. Corbino, W. C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J. K. Wickiser & R. R. Breaker, (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci. USA* **101**: 6421-6426.

Batchelor, E., D. Walthers, L. J. Kenney & M. Goulian, (2005) The *Escherichia coli* CpxA-CpxR envelope stress response system regulates expression of the porins *ompF* and *ompC*. *J. Bacteriol.* **187**: 5723-5731.

Beceiro, A., M. Tomas & G. Bou, (2013) Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin. Microbiol. Rev.* **26**: 185-230.

Beever, L., R. Bond, P. A. Graham, B. Jackson, D. H. Lloyd & A. Loeffler, (2015) Increasing antimicrobial resistance in clinical isolates of *Staphylococcus intermedius* group bacteria and emergence of MRSP in the UK. *The Veterinary record* **176**: 172.

Bentley, S. D., S. Brown, L. D. Murphy, D. E. Harris, M. A. Quail, J. Parkhill, B. G. Barrell, J. R. McCormick, R. I. Santamaria, R. Losick, M. Yamasaki, H. Kinashi, C. W. Chen, G. Chandra, D. Jakimowicz, H. M. Kieser, T. Kieser & K. F. Chater, (2004) SCP1, a 356 023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **51**: 1615-1628.

Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill & D. A. Hopwood, (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141-147.

Betat, H., C. Rammelt & M. Morl, (2010) tRNA nucleotidyltransferases: ancient catalysts with an unusual mechanism of polymerization. *Cell. Mol. Life Sci.* **67**: 1447-1463.

Bibb, M. J., (2005) Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* **8**: 208-215.

Bierman, M., R. Logan, K. Obrien, E. T. Seno, R. N. Rao & B. E. Schoner, (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116**: 43-49.

Bignell, D. R., K. Tahlan, K. R. Colvin, S. E. Jensen & B. K. Leskiw, (2005) Expression of *ccaR*, encoding the positive activator of cephamycin C and clavulanic acid production in *Streptomyces clavuligerus*, is dependent on *bldG*. *Antimicrob. Agents Chemother.* **49**: 1529-1541.

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

Breter, H. J. & R. E. Rhoads, (1979) Analysis of NaIO4-oxidized/NaBH4-reduced mRNA cap analogs by high-performance liquid anion-exchange chromatography and tobacco acid pyrophosphatase (EC 3.6.1.9). *Hoppe-Seylers Zeitschrift Fur Physiologische Chemie* **360**: 240-240.

Britton, R. A., T. Y. Wen, L. Schaefer, O. Pellegrini, W. C. Uicker, N. Mathy, C. Tobin, R. Daou, J. Szyk & C. Condon, (2007) Maturation of the 5 ' end of *Bacillus subtilis* 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol. Microbiol.* **63**: 127-138.

Bucca, G., Z. Hindle & C. P. Smith, (1997) Regulation of the *dnaK* operon of *Streptomyces coelicolor* A3(2) is governed by HspR, an autoregulatory repressor protein. *J. Bacteriol.* **179**: 5999-6004.

Bucca, G., E. Laing, V. Mersinias, N. Allenby, D. Hurd, J. Holdstock, V. Brenner, M. Harrison & C. P. Smith, (2009) Development and application of versatile high density microarrays for genome-wide analysis of *Streptomyces coelicolor*: characterization of the HspR regulon. *Genome Biol.* **10**: e5.

Bystrykh, L. V., M. A. FernandezMoreno, J. K. Herrema, F. Malpartida, D. A. Hopwood & L. Dijkhuizen, (1996) Production of actinorhodin-related ''blue pigments'' by *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **178**: 2238-2244.

Caballero, J. L., F. Malpartida & D. A. Hopwood, (1991) Transcriptional organization and regulation of an antibiotic export complex in the producing *Streptomyces* culture. *Mol. Gen. Genet.* **228**: 372-380.

Caffrey, P., J. F. Aparicio, F. Malpartida & S. B. Zotchev, (2008) Biosynthetic engineering of polyene macrolides towards generation of improved antifungal and antiparasitic agents. *Curr. Top. Med. Chem.* **8**: 639-653.

Carpousis, A. J., B. F. Luisi & K. J. McDowall, (2009) Endonucleolytic Initiation of mRNA Decay in *Escherichia coli*. *Prog Mol Biol Transl* **85**: 91-135.

Carreras, C. W. & C. Khosla, (1998) Purification and *in vitro* reconstitution of the essential protein components of an aromatic polyketide synthase. *Biochemistry* **37**: 2084-2088.

Castro-Melchor, M., S. Charaniya, G. Karypis, E. Takano & W. S. Hu, (2010) Genome-wide inference of regulatory networks in *Streptomyces coelicolor*. *Bmc Genomics* **11**.

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

Challis, G. L., (2008) Genome mining for novel natural product discovery. *J. Med. Chem.* **51**: 2618-2628.

Challis, G. L. & K. F. Chater, (2001) Incorporation of [U-C-13]glycerol defines plausible early steps for the biosynthesis of methylenomycin A in *Streptomyces coelicolor* A3(2). *Chem Commun*: 935-936.

Chan, P. P., A. D. Holmes, A. M. Smith, D. Tran & T. M. Lowe, (2012) The UCSC Archaeal Genome Browser: 2012 update. *Nucleic Acids Res.* **40**: D646-652.

Chater, K. F., (1993) Genetics of Differentiation in *Streptomyces*. *Annu. Rev. Microbiol.* **47**: 685-713.

Chater, K. F., (2001) Regulation of sporulation in *Streptomyces coelicolor* A3(2): a checkpoint multiplex? *Curr. Opin. Microbiol.* **4**: 667-673.

Chater, K. F., (2006) *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**: 761-768.

Chater, K. F., G. Bucca, P. Dyson, K. Fowler, B. Gust, P. Herron, A. Hesketh, G. Hotchkiss, T. Kieser, V. Mersinias & C. P. Smith, (2002) *Streptomyces coelicolor* A3(2): from genome sequence to function. *Functional Microbial Genomics* **33**: 321-336.

Chater, K. F. & G. Chandra, (2006) The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol. Rev.* **30**: 651-672.

Chater, K. F. & G. Chandra, (2008) The use of the rare UUA codon to define "expression space" for genes involved in secondary metabolism, development and environmental adaptation in *Streptomyces*. *J. Microbiol.* **46**: 1-11.

Chater, K. F. & S. Horinouchi, (2003) Signalling early developmental events in two highly diverged *Streptomyces* species. *Mol. Microbiol.* **48**: 9-15.

Chatterji, D. & A. K. Ojha, (2001) Revisiting the stringent response, ppGpp and starvation signaling. *Curr. Opin. Microbiol.* **4**: 160-165.

Chen, L., Y. H. Lu, J. Chen, W. W. Zhang, D. Shu, Z. J. Qin, S. Yang & W. H. Jiang, (2008) Characterization of a negative regulator AveI for avermectin biosynthesis in *Streptomyces avermitilis* NRRL8165. *Appl. Microbiol. Biotechnol.* **80**: 277-286.

Cho, B. K., K. Zengler, Y. Qiu, Y. S. Park, E. M. Knight, C. L. Barrett, Y. Gao & B. O. Palsson, (2009) The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotech.* **27**: 1043-U1115.

Cipriano, M. J., P. N. Novichkov, A. E. Kazakov, D. A. Rodionov, A. P. Arkin, M. S. Gelfand & I. Dubchak, (2013) RegTransBase--a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes. *Bmc Genomics* **14**: 213.

Claessen, D., W. de Jong, L. Dijkhuizen & H. A. B. Wosten, (2006) Regulation of *Streptomyces* development: reach for the sky! *Trends in microbiology* **14**: 313-319.

Claessen, D., R. Rink, W. de Jong, J. Siebring, P. de Vreugd, F. G. H. Boersma, L. Dijkhuizen & H. A. B. Wosten, (2003) A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in *Streptomyces coelicolor* by forming amyloid-like fibrils. *Genes Dev.* **17**: 1714-1726.

Claessen, D., I. Stokroos, H. J. Deelstra, N. A. Penninga, C. Bormann, J. A. Salas, L. Dijkhuizen & H. A. B. Wosten, (2004) The formation of the rodlet layer of streptomycetes is the result of the interplay between rodlins and chaplins. *Mol. Microbiol.* **53**: 433-443.

Colson, S., J. Stephan, T. Hertrich, A. Saito, G. P. van Wezel, F. Titgemeyer & S. Rigali, (2007) Conserved *cis*-acting elements upstream of genes composing the chitinolytic system of streptomycetes are DasR-responsive elements. *J. Mol. Microbiol. Biotech.* **12**: 60-66.

Combes, P., R. Till, S. Bee & M. C. M. Smith, (2002) The *Streptomyces* genome contains multiple pseudo-*attB* sites for the phi C31-encoded site-specific recombination system. *J. Bacteriol.* **184**: 5746-5752.

Condon, C., (2010) What is the role of RNase J in mRNA turnover? *Rna Biology* **7**: 316-321.

Corre, C. & G. L. Challis, (2005) Evidence for the unusual condensation of a diketide with a pentulose in the methylenomycin biosynthetic pathway of *Streptomyces coelicolor* A3(2). *ChemBioChem* **6**: 2166-2170.

Corre, C. & G. L. Challis, (2009) New natural product biosynthetic chemistry discovered by genome mining. *Nat. Prod. Rep.* **26**: 977-986.

Craney, A., S. Ahmed & J. Nodwell, (2013) Towards a new science of secondary metabolism. *The Journal of antibiotics* **66**: 387-400.

Craney, A., T. Hohenauer, Y. Xu, N. K. Navani, Y. F. Li & J. Nodwell, (2007) A synthetic luxCDABE gene cluster optimized for expression in high-GC bacteria. *Nucleic Acids Res.* **35**.

Croxatto, A., V. J. Chalker, J. Lauritz, J. Jass, A. Hardman, P. Williams, M. Camara & D. L. Milton, (2002) VanT, a homologue of Vibrio harveyi LuxR, regulates serine, metalloprotease, pigment, and biofilm production in Vibrio anguillarum. *J. Bacteriol.* **184**: 1617-1629.

Cudny, H. & M. P. Deutscher, (1986) High-level overexpression, rapid purification, and properties of *Escherichia coli* tRNA nucleotidyltransferase. *J. Biol. Chem.* **261**: 6450-6453.

Cuthbertson, L. & J. R. Nodwell, (2013) The TetR Family of Regulators. *Microbiol. Mol. Biol. Rev.* **77**: 440-475.

Davies, J., (2011) How to discover new antibiotics: harvesting the parvome. *Curr. Opin. Chem. Biol.* **15**: 5-10.

Davis, N. K. & K. F. Chater, (1990) Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Mol. Microbiol.* **4**: 1679-1691.

Davis, N. K. & K. F. Chater, (1992) The *Streptomyces coelicolor whiB* gene encodes a small transcription factor-like protein dispensable for growth but essential for sporulation. *Mol. Gen. Genet.* **232**: 351-358.

Davis, S. E., R. A. Mooney, E. I. Kanin, J. Grass, R. Landick & A. Z. Ansari, (2011) Mapping *E. coli* RNA Polymerase and Associated Transcription Factors and Identifying Promoters Genome-Wide. *Method Enzymol* **498**: 449-471.

Daza, A., J. F. Martin, A. Dominguez & J. A. Gil, (1989) Sporulation of Several Species of *Streptomyces* in Submerged Cultures after Nutritional Downshift. *J. Gen. Microbiol.* **135**: 2483-2491.

Deana, A., H. Celesnik & J. G. Belasco, (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* **451**: 355-358.

Debono, M., B. J. Abbott, R. M. Molloy, D. S. Fukuda, A. H. Hunt, V. M. Daupert, F. T. Counter, J. L. Ott, C. B. Carrell, L. C. Howard, L. D. Boeck & R. L. Hamill, (1988) Enzymatic and Chemical Modifications of Lipopeptide Antibiotic A21978c - the Synthesis and Evaluation of Daptomycin (Ly146032). *J. Antibiot.* **41**: 1093-1105.

Deutscher, J., (2008) The mechanisms of carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.* **11**: 87-93.

Deutscher, M. P., (2009) Maturation and degradation of ribosomal RNA in bacteria. *Progress in molecular biology and translational science* **85**: 369-391.

Donadio, S., P. Monciardini & M. Sosio, (2007) Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat. Prod. Rep.* **24**: 1073-1109.

Dutta, T. & M. P. Deutscher, (2010) Mode of action of RNase BN/RNase Z on tRNA precursors: RNase BN does not remove the CCA sequence from tRNA. *J. Biol. Chem.* **285**: 22874-22881.

Dutta, T., A. Malhotra & M. P. Deutscher, (2012) Exoribonuclease and endoribonuclease activities of RNase BN/RNase Z both function *in vivo*. *J. Biol. Chem.* **287**: 35747-35755.

Dyson, M. R., S. P. Shadbolt, K. J. Vincent, R. L. Perera & J. McCafferty, (2004) Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC biotechnology* **4**: 32.

Einhauer, A. & A. Jungbauer, (2001) The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *Journal of biochemical and biophysical methods* **49**: 455-465.

Elliot, M. A., M. J. Bibb, M. J. Buttner & B. K. Leskiw, (2001) BldD is a direct regulator of key developmental genes in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **40**: 257-269.

Elliot, M. A., N. Karoonuthaisiri, J. Q. Huang, M. J. Bibb, S. N. Cohen, C. M. Kao & M. J. Buttner, (2003) The chaplins: a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in *Streptomyces coelicolor*. *Genes Dev.* **17**: 1727-1740.

Embley, T. M. & E. Stackebrandt, (1994) The Molecular Phylogeny and Systematics of the Actinomycetes. *Annu. Rev. Microbiol.* **48**: 257-289.

Feitelson, J. S. & D. A. Hopwood, (1983) Cloning of a *Streptomyces* Gene for an O-Methyltransferase Involved in Antibiotic Biosynthesis. *Mol. Gen. Genet.* **190**: 394-398.

Fernandez-Moreno, M. A., J. L. Caballero, D. A. Hopwood & F. Malpartida, (1991) The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. *Cell* **66**: 769-780.

Fernandez-Moreno, M. A., E. Martinez, L. Boto, D. A. Hopwood & F. Malpartida, (1992) Nucleotide sequence and deduced functions of a set of cotranscribed genes of *Streptomyces coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin. *J. Biol. Chem.* **267**: 19278-19290.

Fernandez-Moreno, M. A., E. Martinez, J. L. Caballero, K. Ichinose, D. A. Hopwood & F. Malpartida, (1994) DNA sequence and functions of the *act*VI region of the

actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor* A3(2). *J. Biol. Chem.* **269**: 24854-24863.

Flardh, K. & M. J. Buttner, (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature Rev. Microbiol.* **7**: 36-49.

Flardh, K., K. C. Findlay & K. F. Chater, (1999) Association of early sporulation genes with suggested developmental decision points in *Streptomyces coelicolor* A3(2). *Microbiol-Uk* **145**: 2229-2243.

Flardh, K., E. Leibovitz, M. J. Buttner & K. F. Chater, (2000) Generation of a non-sporulating strain of *Streptomyces coelicolor* A3(2) by the manipulation of a developmentally controlled ftsZ promoter. *Mol. Microbiol.* **38**: 737-749.

Floriano, B. & M. Bibb, (1996) *afsR* is a pleiotropic but conditionally required regulatory gene for antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **21**: 385-396.

Fujii, T., H. C. Gramajo, E. Takano & M. J. Bibb, (1996) *redD* and *act*II-ORF4, pathway-specific regulatory genes for antibiotic production in *Streptomyces coelicolor* A3(2), are transcribed *in vitro* by an RNA polymerase holoenzyme containing sigma (*hrdD*). *J. Bacteriol.* **178**: 3402-3405.

Gao, C., Hindra, D. Mulder, C. Yin & M. A. Elliot, (2012) Crp is a global regulator of antibiotic production in *Streptomyces*. *Mbio* **3**.

Gatewood, M. L., P. Bralley & G. H. Jones, (2011) RNase III-dependent expression of the rpsO-pnp operon of *Streptomyces coelicolor*. *J. Bacteriol.* **193**: 4371-4379.

Gold, L., (1995) Oligonucleotides as Research, Diagnostic, and Therapeutic Agents. *J. Biol. Chem.* **270**: 13581-13584.

Goldman, S. R., J. S. Sharp, I. O. Vvedenskaya, J. Livny, S. L. Dove & B. E. Nickels, (2011) NanoRNAs prime transcription initiation *in vivo*. *Mol. Cell* **42**: 817-825.

Gomez-Escribano, J. P. & M. J. Bibb, (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol* **4**: 207-215.

Gorke, B. & J. Stulke, (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature reviews. Microbiology* **6**: 613-624.

Gossringer, M. & R. K. Hartmann, (2012) 3 '-UTRs as a source of regulatory RNAs in bacteria. *EMBO J.* **31**: 3958-3960.

Gottelt, M., S. Kol, J. P. Gomez-Escribano, M. Bibb & E. Takano, (2010) Deletion of a regulatory gene within the *cpk* gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). *Microbiol-Sgm* **156**: 2343-2353.

Gramajo, H. C., E. Takano & M. J. Bibb, (1993) Stationary phase production of the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2) is transcriptionally regulated. *Mol. Microbiol.* **7**: 837-845.

Graur, D., Y. Zheng, N. Price, R. B. Azevedo, R. A. Zufall & E. Elhaik, (2013) On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol* **5**: 578-590.

Graziani, E. I., (2009) Recent advances in the chemistry, biosynthesis and pharmacology of rapamycin analogs. *Nat. Prod. Rep.* **26**: 602-609.

Gross, H., (2009) Genomic mining - a concept for the discovery of new bioactive natural products. *Curr. Opin. Drug Di. De.* **12**: 207-219.

Gust, B., G. L. Challis, K. Fowler, T. Kieser & K. F. Chater, (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.

Gutgsell, N. S. & C. Jain, (2012) Role of precursor sequences in the ordered maturation of *E. coli* 23S ribosomal RNA. *RNA* **18**: 345-353.

Guthrie, E. P., C. S. Flaxman, J. White, D. A. Hodgson, M. J. Bibb & K. F. Chater, (1998) A response-regulator-like activator of antibiotic synthesis from *Streptomyces coelicolor* A3(2) with an amino-terminal domain that lacks a phosphorylation pocket. *Microbiol-Uk* **144**: 2007-2007.

Hallam, S. E., F. Malpartida & D. A. Hopwood, (1988) Nucleotide sequence, transcription and deduced function of a gene involved in polyketide antibiotic synthesis in *Streptomyces coelicolor*. *Gene* **74**: 305-320.

Haneishi, T., A. Terahara, M. Arai, T. Hata & C. Tamura, (1974) New antibiotics, methylenomycins A and B. II. Structures of methylenomycins A and B. *The Journal of antibiotics* **27**: 393-399.

Harley, C. B. & R. P. Reynolds, (1987) Analysis of *Escherichia coli* Promoter Sequences. *Nucleic Acids Res.* **15**: 2343-2361.

Hartmann, R. K., M. Gossringer, B. Spath, S. Fischer & A. Marchfelder, (2009) The Making of tRNAs and More - RNase P and tRNase Z. *Prog Mol Biol Transl* **85**: 319-368.

Haugen, S. P., M. B. Berkmen, W. Ross, T. Gaal, C. Ward & R. L. Gourse, (2006) rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell* **125**: 1069-1082.

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

Hertweck, C., A. Luzhetskyy, Y. Rebets & A. Bechthold, (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **24**: 162-190.

Hiard, S., R. Maree, S. Colson, P. A. Hoskisson, F. Titgemeyer, G. P. van Wezel, B. Joris, L. Wehenkel & S. Rigali, (2007) PREDetector: A new tool to identify regulatory elements in bacterial genomes. *Biochem. Biophys. Res. Commun.* **357**: 861-864.

Higo, A., H. Hara, S. Horinouchi & Y. Ohnishi, (2012) Genome-wide Distribution of AdpA, a Global Regulator for Secondary Metabolism and Morphological Differentiation in *Streptomyces*, Revealed the Extent and Complexity of the AdpA Regulatory Network. *DNA Res* **19**: 259-273.

Hillen, W. & C. Berens, (1994) Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu. Rev. Microbiol.* **48**: 345-369.

Hillerich, B. & J. Westpheling, (2006) A new GntR family transcriptional regulator in *Streptomyces coelicolor* is required for morphogenesis and antibiotic production and controls transcription of an ABC transporter in response to carbon source. *J. Bacteriol.* **188**: 7477-7487.

Hirano, S., K. Tanaka, Y. Ohnishi & S. Horinouchi, (2008) Conditionally positive effect of the TetR-family transcriptional regulator AtrA on streptomycin production by *Streptomyces griseus*. *Microbiol-Sgm* **154**: 905-914.

Hobbs, G., C. M. Frazer, D. C. J. Gardner, J. A. Cullum & S. G. Oliver, (1989) Dispersed Growth of *Streptomyces* in Liquid Culture. *Appl. Microbiol. Biotechnol.* **31**: 272-277.

Hojati, Z., C. Milne, B. Harvey, L. Gordon, M. Borg, F. Flett, B. Wilkinson, P. J. Sidebottom, B. A. M. Rudd, M. A. Hayes, C. P. Smith & J. Micklefield, (2002) Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem. Biol.* **9**: 1175-1187.

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

Hopp, T. P., K. S. Prickett, V. L. Price, R. T. Libby, C. J. March, D. P. Cerretti, D. L. Urdal & P. J. Conlon, (1988) A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Bio-Technol* **6**: 1204-1210.

Hopwood, D. A., (1967) Genetic analysis and genome structure in *Streptomyces coelicolor*. *Bacteriological reviews* **31**: 373-403.

Hopwood, D. A., (1986) Cloning and analysis of antibiotic biosynthetic genes in *Streptomyce,* In: Biological, Biochemical and Biomedical aspects of Actinomycetes. G. Szabo, S. Biro, M. Goodfellow (Eds). *Akademiai Kiaido, Budapest*.

Hopwood, D. A., (1999) Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*. *Microbiology* **145 (Pt 9)**: 2183-2202.

Horinouchi, S., (2002) A microbial hormone, A-factor, as a master switch for morphological differentiation and secondary metabolism in *Streptomyces griseus*. *Front. Biosci.* **7**: D2045-D2057.

Horinouchi, S., (2007) Mining and polishing of the treasure trove in the bacterial genus *Streptomyces*. *Biosci. Biotechnol. Biochem.* **71**: 283-299.

Horinouchi, S. & T. Beppu, (2007) Hormonal control by A-factor of morphological development and secondary metabolism in *Streptomyces*. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **83**: 277-295.

Horinouchi, S., M. Kito, M. Nishiyama, K. Furuya, S. K. Hong, K. Miyake & T. Beppu, (1990) Primary structure of *afsR*, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2) *Gene* **95**: 49-56.

Hoskisson, P. A., S. Rigali, K. Fowler, K. C. Findlay & M. J. Buttner, (2006) DevA, a GntR-like transcriptional regulator required for development in *Streptomyces coelicolor*. *J. Bacteriol.* **188**: 5014-5023.

Hsiao, N. H., M. Gottelt & E. Takano, (2009) Regulation of antibiotic production by bacterial hormones. In: Complex Enzymes in Microbial Natural Product Biosynthesis, Part A: Overview Articles and Peptides. pp. 143-157.

Hsiao, N. H., J. Soding, D. Linke, C. Lange, C. Hertweck, W. Wohlleben & E. Takano, (2007) ScbA from *Streptomyces coelicolor* A3(2) has homology to fatty acid

synthases and is able to synthesize α-butyrolactones. *Microbiol-Sgm* **153**: 1394-1404.

Hunt, A. C., L. Servin-Gonzalez, G. H. Kelemen & M. J. Buttner, (2005) The *bldC* developmental locus of *Streptomyces coelicolor* encodes a member of a family of small DNA-binding proteins related to the DNA-binding domains of the MerR family. *J. Bacteriol.* **187**: 716-728.

Ingolia, N. T., (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.* **15**: 205-213.

Ingolia, N. T., G. A. Brar, S. Rouskin, A. M. McGeachy & J. S. Weissman, (2012) The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7**: 1534-1550.

Ingolia, N. T., S. Ghaemmaghami, J. R. Newman & J. S. Weissman, (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218-223.

Inoue, H., H. Nojima & H. Okayama, (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**: 23-28.

Irnov, I., C. M. Sharma, J. Vogel & W. C. Winkler, (2010) Identification of regulatory RNAs in *Bacillus subtilis*. *Nucleic Acids Res.* **38**: 6637-6651.

Jakimowicz, D., S. Mouz, J. Zakrzewska-Czerwinska & K. F. Chater, (2006) Developmental control of a *parAB* promoter leads to formation of sporulation-associated ParB complexes in *Streptomyces coelicolor*. *J. Bacteriol.* **188**: 1710-1720.

Javidpour, P., J. Bruegger, S. Srithahan, T. P. Korman, M. P. Crump, J. Crosby, M. D. Burkart & S. C. Tsai, (2013) The determinants of activity and specificity in actinorhodin type II polyketide ketoreductase. *Chem. Biol.* **20**: 1225-1234.

Jung, D., A. Rozek, M. Okon & R. E. W. Hancock, (2004) Structural transitions as determinants of the action of the calcium-dependent antibiotic daptomycin. *Chem. Biol.* **11**: 949-957.

Kaberdina, A. C., W. Szaflarski, K. H. Nierhaus & I. Moll, (2009) An Unexpected Type of Ribosomes Induced by Kasugamycin: A Look into Ancestral Times of Protein Synthesis? *Mol. Cell* **33**: 227-236.

Kaiser, B. K. & B. L. Stoddard, (2011) DNA recognition and transcriptional regulation by the WhiA sporulation factor. *Scientific reports* **1**.

Kamionka, A., J. Bogdanska-Urbaniak, O. Scholz & W. Hillen, (2004) Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Res.* **32**: 842-847.

Kang, S. G., W. Jin, M. Bibb & K. J. Lee, (1998) Actinorhodin and undecylprodigiosin production in wild-type and *relA* mutant strains of *Streptomyces coelicolor* A3(2) grown in continuous culture. *FEMS Microbiol. Lett.* **168**: 221-226.

Kao, K. C., Y. L. Yang, R. Boscolo, C. Sabatti, V. Roychowdhury & J. C. Liao, (2004) Transcriptome-based determination of multiple transcription regulator activities in *Escherichia coli* by using network component analysis. *Proc Natl Acad Sci U S A* **101**: 641-646.

Karr, J. R., J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, Jr., N. Assad-Garcia, J. I. Glass & M. W. Covert, (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* **150**: 389-401.

Kato, J. Y., N. Funa, H. Watanabe, Y. Ohnishi & S. Horinouchi, (2007) Biosynthesis of γ-butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. *Proc. Natl. Acad. Sci. USA* **104**: 2378-2383.

Kazakov, A. E., M. J. Cipriano, P. S. Novichkov, S. Minovitsky, D. V. Vinogradov, A. Arkin, A. A. Mironov, M. S. Gelfand & I. Dubchak, (2007) RegTransBase - a database of regulatory sequences and interactions in a wide range of prokaryotic genomes. *Nucleic Acids Res.* **35**: D407-D412.

Kelemen, G. H., G. L. Brown, J. Kormanec, L. Potuckova, K. F. Chater & M. J. Buttner, (1996) The positions of the sigma-factor genes, *whiG* and *sigF*, in the hierarchy controlling the development of spore chains in the aerial hyphae of *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **21**: 593-603.

Kempter, C., D. Kaiser, S. Haag, G. Nicholson, V. Gnau, T. Walk, K. H. Gierling, H. Decker, H. Zahner, G. Jung & J. W. Metzger, (1997) CDA: Calcium-Dependent antibiotic from *Streptomyces coelicolor* A3(2) containing unusual residues. *Chem. Int. Ed. Engl.* **36**: 498–501.

Schneider, K. L., K. S. Pollard, R. Baertsch, A. Pohl & T. M. Lowe, (2006) The UCSC Archaeal Genome Browser. *Nucleic Acids Res* **34**: D407-410.

Khokhlov, A. S., L. N. Anisova, I. I. Tovarova, E. M. Kleiner, I. V. Kovalenko, O. I. Krasilnikova, E. Y. Kornitskaya & S. A. Pliner, (1973) Effect of A-factor on the growth of asporogenous mutants of *Streptomyces griseus*, not producing this factor. *Zeitschrift für allgemeine Mikrobiologie* **13**: 647-655.

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

Kim, S. H., B. A. Traag, A. H. Hasan, K. J. McDowall, B. G. Kim & G. P. van Wezel, (2015) Transcriptional analysis of the cell division-related ssg genes in *Streptomyces coelicolor* reveals direct control of *ssgR* by AtrA. *Antonie van Leeuwenhoek*.

Kime, L., S. S. Jourdan & K. J. McDowall, (2008) Identifying and Characterizing Substrates of the RNase E/G Family of Enzymes. *Rna Turnover in Bacteria, Archaea and Organelles* **447**: 215-241.

Kirby, R. & D. A. Hopwood, (1977) Genetic Determination of Methylenomycin Synthesis by Scp1 Plasmid of *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* **98**: 239-252.

Kodani, S., M. E. Hudson, M. C. Durrant, M. J. Buttner, J. R. Nodwell & J. M. Willey, (2004) The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene *ramS* in *Streptomyces coelicolor*. *Proc. Natl. Acad. Sci. USA* **101**: 11448-11453.

Korner, H., H. J. Sofia & W. G. Zumft, (2003) Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol. Rev.* **27**: 559-592.

Kwan, D. H. & F. Schulz, (2011) The stereochemistry of complex polyketide biosynthesis by modular polyketide synthases. *Molecules* **16**: 6092-6115.

Labeda, D. P., (2011) Multilocus sequence analysis of phytopathogenic species of the genus *Streptomyces*. *International journal of systematic and evolutionary microbiology* **61**: 2525-2531.

Lakey, J. H., E. J. A. Lea, B. A. M. Rudd, H. M. Wright & D. A. Hopwood, (1983) A New Channel-Forming Antibiotic from *Streptomyces coelicolor* A3(2) Which Requires Calcium for Its Activity. *J. Gen. Microbiol.* **129**: 3565-3573.

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

Lawlor, E. J., H. A. Baylis & K. F. Chater, (1987) Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). *Genes Dev.* **1**: 1305-1310.

Lee, H. N., J. S. Kim, P. Kim, H. S. Lee & E. S. Kim, (2013) Repression of antibiotic downregulator WblA by AdpA in *Streptomyces coelicolor*. *Appl. Environ. Microbiol.* **79**: 4159-4163.

Lewis, R. A., E. Laing, N. Allenby, G. Bucca, V. Brenner, M. Harrison, A. M. Kierzek & C. P. Smith, (2010) Metabolic and evolutionary insights into the closely-related species *Streptomyces coelicolor* and *Streptomyces lividans* deduced from high-resolution comparative genomic hybridization. *Bmc Genomics* **11**: 682.

Li, X., T. Yu, Q. He, K. J. McDowall, B. Jiang, Z. Jiang, L. Wu, G. Li, Q. Li, S. Wang, Y. Shi, L. Wang & B. Hong, (2015) Binding of a biosynthetic intermediate to AtrA modulates the production of lidamycin by *Streptomyces globisporus*. *Mol. Microbiol.*

Liang, H., Y. T. Zhao, J. Q. Zhang, X. J. Wang, R. X. Fang & Y. T. Jia, (2011) Identification and functional characterization of small non-coding RNAs in *Xanthomonas oryzae* pathovar oryzae. *Bmc Genomics* **12**.

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

Lisser, S. & H. Margalit, (1993) Compilation of *Escherichia coli* mRNA Promoter Sequences. *Nucleic Acids Res.* **21**: 1507-1516.

Liu, G., K. F. Chater, G. Chandra, G. Niu & H. Tan, (2013) Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol. Mol. Biol. Rev.* **77**: 112-143.

Loria, R., D. Bignell, J. Huguet, E. Johnson, M. Joshi, S. Moll & R. Seipke, (2008) *Streptomyces* from the dark side: Mechanisms and emergence of pathogenicity. In: 100th Annual Meeting of the American-Phytopathological-Society. Minneapolis, MN, pp. S191-S191.

Loria, R., D. R. D. Bignell, S. Moll, J. C. Huguet-Tapia, M. V. Joshi, E. G. Johnson, R. F. Seipke & D. M. Gibson, (2007) Thaxtomin biosynthesis: the path to plant pathogenicity in the genus *Streptomyces*. In: 14th Meeting of the International-Symposia-on-the-Biology-of-Actinomycetes (ISBA). Newcastle upon Tyne, ENGLAND, pp. 3-10.

Luzhetskii, A., A. Mazepa & O. Marina, (2000) Intergeneric Conjugation *Escherichia coli–Streptomyces*: The Use to Introduce Genetic Information into *Streptomyces kanamyceticus* I and *Streptomyces globisporus* 1912 Strains. *Visn. L'viv. Univ., Ser. Biol.* **25** 74–80.

Luzhetskii, A. N., B. E. Ostash & V. A. Fedorenko, (2001) Intergeneric conjugation *Escherichia coli-Streptomyces globisporus* 1912 using integrative plasmid pSET152 and its derivatives. *Russ J Genet+* **37**: 1123-1129.

Malpartida, F., J. Niemi, R. Navarrete & D. A. Hopwood, (1990) Cloning and expression in a heterologous host of the complete set of genes for biosynthesis of the *Streptomyces coelicolor* antibiotic undecylprodigiosin. *Gene* **93**: 91-99.

Mamanova, L., R. M. Andrews, K. D. James, E. M. Sheridan, P. D. Ellis, C. F. Langford, T. W. B. Ost, J. E. Collins & D. J. Turner, (2010) FRT-seq: amplification-free, strand-specific transcriptome sequencing. *Nat. Methods* **7**: 130-U163.

Mandal, M., M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo & R. R. Breaker, (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**: 275-279.

Manteca, A., H. R. Jung, V. Schwammle, O. N. Jensen & J. Sanchez, (2010) Quantitative proteome analysis of *Streptomyces coelicolor* nonsporulating liquid cultures demonstrates a complex differentiation process comparable to that occurring in sporulating solid cultures. *J. Proteome Res.* **9**: 4801-4811.

Manteca, A., U. Mader, B. A. Connolly & J. Sanchez, (2006) A proteomic analysis of *Streptomyces coelicolor* programmed cell death. *Proteomics* **6**: 6008-6022.

Mao, X. M., S. Luo, R. C. Zhou, F. Wang, P. Yu, N. Sun, X. X. Chen, Y. Tang & Y. Q. Li, (2015) Transcriptional regulation of the daptomycin gene cluster in *Streptomyces roseosporus* by an autoregulator, AtrA. *J. Biol. Chem.* **290**: 7992-8001.

Matharu, A. L., R. J. Cox, J. Crosby, K. J. Byrom & T. J. Simpson, (1998) MCAT is not required for *in vitro* polyketide synthesis in a minimal actinorhodin polyketide synthase from *Streptomyces coelicolor*. *Chem. Biol.* **5**: 699-711.

Mathy, N., L. Benard, O. Pellegrini, R. Daou, T. Y. Wen & C. Condon, (2007) 5 '-to-3 ' exoribonuclease activity in bacteria: Role of RNase J1 in rRNA maturation and 5 ' stability of mRNA. *Cell* **129**: 681-692.

McArthur, M. & M. J. Bibb, (2008) Manipulating and understanding antibiotic production in *Streptomyces coelicolor* A3(2) with decoy oligonucleotides. *Proc Natl Acad Sci U S A* **105**: 1020-1025.

McCormick, J. R. & K. Flardh, (2012) Signals and regulators that govern *Streptomyces* development. *FEMS Microbiol. Rev.* **36**: 206-231.

McDaniel, R., S. Ebert-Khosla, D. A. Hopwood & C. Khosla, (1994) Engineered Biosynthesis of Novel Polyketides: *act*VII and *act*IV Genes Encode Aromatase and Cyclase Enzymes, Respectively. *J. Am. Chem. Soc.* **116**.

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

McKenzie, N. L. & J. R. Nodwell, (2007) Phosphorylated AbsA2 negatively regulates antibiotic production in *Streptomyces coelicolor* through interactions with pathway-specific regulatory gene promoters. *J. Bacteriol.* **189**: 5284-5292.

Meier, I., L. V. Wray & W. Hillen, (1988) Differential regulation of the Tn10-encoded tetracycline resistance genes *tetA* and *tetR* by the tandem tet operators O1 and O2. *EMBO J.* **7**: 567-572.

Merino, E., R. A. Jensen & C. Yanofsky, (2008) Evolution of bacterial *trp* operons and their regulation. *Curr. Opin. Microbiol.* **11**: 78-86.

Miao, V., M. F. Coeffet-LeGal, P. Brian, R. Brost, J. Penn, A. Whiting, S. Martin, R. Ford, I. Parr, M. Bouchard, C. J. Silva, S. K. Wrigley & R. H. Baltz, (2005) Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiol-Sgm* **151**: 1507-1523.

Mistry, B. V., R. Del Sol, C. Wright, K. Findlay & P. Dyson, (2008) FtsW is a dispensable cell division protein required for Z-ring stabilization during sporulation septation in *Streptomyces coelicolor*. *J. Bacteriol.* **190**: 5555-5566.

Mo, S., P. K. Sydor, C. Corre, M. M. Alhamadsheh, A. E. Stanley, S. W. Haynes, L. Song, K. A. Reynolds & G. L. Challis, (2008) Elucidation of the *Streptomyces coelicolor* pathway to 2-undecylpyrrole, a key intermediate in undecylprodiginine and streptorubin B biosynthesis. *Chem. Biol.* **15**: 137-148.

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

Mulle, J. G., V. C. Patel, S. T. Warren, M. R. Hegde, D. J. Cutler & M. E. Zwick, (2010) Empirical evaluation of oligonucleotide probe selection for DNA microarrays. *PloS one* **5**: e9921.

Nakagawa, S., Y. Niimura, K. Miura & T. Gojobori, (2010) Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc Natl Acad Sci U S A* **107**: 6382-6387.

Narva, K. E. & J. S. Feitelson, (1990) Nucleotide sequence and transcriptional analysis of the *redD* locus of *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **172**: 326-333.

Nechooshtan, G., M. Elgrably-Weiss, A. Sheaffer, E. Westhof & S. Altuvia, (2009) A pH-responsive riboregulator. *Genes Dev.* **23**: 2650-2662.

Neidhardt, F. C., (1996) *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. **1 and 2**.

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

Nguyen, K. T., J. Tenor, H. Stettler, L. T. Nguyen, L. D. Nguyen & C. J. Thompson, (2003) Colonial differentiation in *Streptomyces coelicolor* depends on translation of a specific codon within the *adpA* gene. *J. Bacteriol.* **185**: 7291-7296.

Nicholson, A. W., (2003) The ribonuclease III superfamily: forms and functions in RNA maturation, decay, and gene silencing In: RNAi: A Guide to Gene Silencing. G. J. Hannon (ed). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp.

Nickels, B. E. & S. L. Dove, (2011) NanoRNAs: a class of small RNAs that can prime transcription initiation in bacteria. *J. Mol. Biol.* **412**: 772-781.

Nodwell, J. R., R. Yang, D. Kuo & R. Losick, (1999) Extracellular complementation and the identification of additional genes involved in aerial mycelium formation in *Streptomyces coelicolor*. *Genetics* **151**: 569-584.

Noens, E. E., V. Mersinias, J. Willemse, B. A. Traag, E. Laing, K. F. Chater, C. P. Smith, H. K. Koerten & G. P. van Wezel, (2007) Loss of the controlled localization of growth stage-specific cell-wall synthesis pleiotropically affects developmental gene expression in an *ssgA* mutant of *Streptomyces coelicolor*. *Mol. Microbiol.* **64**: 1244-1259.

Noens, E. E. E., V. Mersinias, B. A. Traag, C. P. Smith, H. K. Koerten & G. P. van Wezel, (2005) SsgA-like proteins determine the fate of peptidoglycan during sporulation of *Streptomyces coelicolor*. *Mol. Microbiol.* **58**: 929-944.

Nothaft, H., S. Rigali, B. Boomsma, M. Swiatek, K. J. McDowall, G. P. van Wezel & F. Titgemeyer, (2010) The permease gene *nagE2* is the key to N-acetylglucosamine sensing and utilization in *Streptomyces coelicolor* and is subject to multi-level control. *Mol. Microbiol.* **75**: 1133-1144.

O'Rourke, S., A. Wietzorrek, K. Fowler, C. Corre, G. L. Challis & K. F. Chater, (2009) Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in *Streptomyces coelicolor*. *Mol. Microbiol.* **71**: 763-778.

Ohnishi, Y., Y. Furusho, T. Higashi, H. K. Chun, K. Furihata, S. Sakuda & S. Horinouchi, (2004) Structures of grixazone A and B, A-factor-dependent yellow pigments produced under phosphate depletion by *Streptomyces griseus*. *J. Antibiot.* **57**: 218-223.

Ohnishi, Y., S. Kameyama, H. Onaka & S. Horinouchi, (1999) The A-factor regulatory cascade leading to streptomycin biosynthesis in *Streptomyces griseus*: identification of a target gene of the A-factor receptor. *Mol. Microbiol.* **34**: 102-111.

Ohnishi, Y., H. Yamazaki, J. Y. Kato, A. Tomono & S. Horinouchi, (2005) AdpA, a central transcriptional regulator in the A-factor regulatory cascade that leads to morphological development and secondary metabolism in *Streptomyces griseus*. *Biosci. Biotechnol. Biochem.* **69**: 431-439.

Okamoto, S., T. Taguchi, K. Ochi & K. Ichinose, (2009) Biosynthesis of actinorhodin and related antibiotics: discovery of alternative routes for quinone formation encoded in the *act* gene cluster. *Chem. Biol.* **16**: 226-236.

Olano, C., F. Lombo, C. Mendez & J. A. Salas, (2008) Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metab. Eng.* **10**: 281-292.

Olano, C., C. Mendez & J. A. Salas, (2009) Antitumor compounds from actinomycetes: from gene clusters to new derivatives by combinatorial biosynthesis. *Nat. Prod. Rep.* **26**: 628-660.

Onaka, H., N. Ando, T. Nihira, Y. Yamada, T. Beppu & S. Horinouchi, (1995) Cloning and characterisation of the A-factor receptor protein from *Streptomyces griseus*. *J. Bacteriol.* **177**: 6083-6092.

Onaka, H. & S. Horinouchi, (1997) DNA-binding activity of the A-factor receptor protein and its recognition DNA sequences. *Mol. Microbiol.* **24**: 991-1000.

Paget, M. S., V. Molle, G. Cohen, Y. Aharonowitz & M. J. Buttner, (2001) Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Mol. Microbiol.* **42**: 1007-1020.

Panek, J., J. Bobek, K. Mikulik, M. Basler & J. Vohradsky, (2008) Biocomputational prediction of small non-coding RNAs in *Streptomyces*. *Bmc Genomics* **9**: 217.

Park, S. S., Y. H. Yang, E. Song, E. J. Kim, W. S. Kim, J. K. Sohng, H. C. Lee, K. K. Liou & B. G. Kim, (2009) Mass spectrometric screening of transcriptional regulators involved in antibiotic biosynthesis in *Streptomyces coelicolor* A3(2). *J Ind Microbiol Biotechnol* **36**: 1073-1083.

Persson, J., K. F. Chater & K. Flardh, (2013) Molecular and cytological analysis of the expression of *Streptomyces* sporulation regulatory gene *whiH*. *FEMS Microbiol. Lett.* **341**: 96-105.

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

Piette, A., A. Derouanx, P. Gerkens, E. E. E. Noens, G. Mazzucchelli, S. Vion, H. K. Koerten, F. Titgemeyer, E. De Pauw, P. Leprince, G. P. van Wezel, M. Galleni & S. B. Rigali, (2005) From dormant to germinating spores of *Streptomyces coelicolor* A3(2): New perspectives from the *crp* null mutant. *J. Proteome Res.* **4**: 1699-1708.

Poot, R. A., C. W. Pleij & J. van Duin, (1996) The central pseudoknot in 16S ribosomal RNA is needed for ribosome stability but is not essential for 30S initiation complex formation. *Nucleic Acids Res.* **24**: 3670-3676.

Price, B., T. Adamidis, R. Q. Kong & W. Champness, (1999) A *Streptomyces coelicolor* antibiotic regulatory gene, *absB*, encodes an RNase III homolog. *J. Bacteriol.* **181**: 6142-6151.

Price, M. N., K. H. Huang, E. J. Alm & A. P. Arkin, (2005) A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res.* **33**: 880-892.

Pruitt, K. D., T. Tatusova & D. R. Maglott, (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**: D61-D65.

Raghavan, R., D. B. Sloan & H. Ochman, (2012) Antisense transcription is pervasive but rarely conserved in enteric bacteria. *Mbio* **3**.

Ramos, J. L., M. Martinez-Bueno, A. J. Molina-Henares, W. Teran, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan & R. Tobes, (2005) The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* **69**: 326-356.

Rampersaud, A., S. L. Harlocker & M. Inouye, (1994) The OmpR protein of *Escherichia coli* binds to sites in the *ompF* promoter region in a hierarchical manner determined by its degree of phosphorylation. *J. Biol. Chem.* **269**: 12559-12566.

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

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

Rico, S., R. I. Santamaria, A. Yepes, H. Rodriguez, E. Laing, G. Bucca, C. P. Smith & M. Diaz, (2014) Deciphering the regulon of *Streptomyces coelicolor* AbrC3, a positive response regulator of antibiotic production. *Appl. Environ. Microbiol.* **80**: 2417-2428.

Rigali, S., H. Nothaft, E. E. E. Noens, M. Schlicht, S. Colson, M. Muller, B. Joris, H. K. Koerten, D. A. Hopwood, F. Titgemeyer & G. P. van Wezel, (2006) The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Mol. Microbiol.* **61**: 1237-1251.

Rigali, S., F. Titgemeyer, S. Barends, S. Mulder, A. W. Thomae, D. A. Hopwood & G. P. van Wezel, (2008) Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO reports* **9**: 670-675.

Robinson, J. A., (1991) Polyketide synthase complexes: their structure and function in antibiotic biosynthesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **332**: 107-114.

Rodriguez-Garcia, A., C. Barreiro, F. Santos-Beneit, A. Sola-Landa & J. F. Martin, (2007) Genome-wide transcriptomic and proteomic analysis of the primary response to phosphate limitation in *Streptomyces coelicolor* M145 and in a delta *phoP* mutant. *Proteomics* **7**: 2410-2429.

Rodriguez, E., C. Banchio, L. Diacovich, M. J. Bibb & H. Gramajo, (2001) Role of an essential acyl coenzyme A carboxylase in the primary and secondary metabolism of *Streptomyces coelicolor* A3(2). *Appl. Environ. Microbiol.* **67**: 4166-4176.

Romero, D. A., A. H. Hasan, Y. F. Lin, L. Kime, O. Ruiz-Larrabeiti, M. Urem, G. Bucca, L. Mamanova, E. E. Laing, G. P. van Wezel, C. P. Smith, V. R. Kaberdin & K. J. McDowall, (2014) A comparison of key aspects of gene regulation in *Streptomyces coelicolor* and *Escherichia coli* using nucleotide-resolution transcription maps produced in parallel by global and differential RNA sequencing. *Mol. Microbiol.*

Rudd, B. A. & D. A. Hopwood, (1979) Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* **114**: 35-43.

Ryding, N. J., G. H. Kelemen, C. A. Whatling, K. Flardh, M. J. Buttner & K. F. Chater, (1998) A developmentally regulated gene encoding a repressor-like protein is essential for sporulation in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **29**: 343-357.

Ryu, Y. G., M. J. Butler, K. F. Chater & K. J. Lee, (2006) Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*. *Appl. Environ. Microbiol.* **72**: 7132-7139.

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

Sallet, E., B. Roux, L. Sauviac, M. F. Jardinaud, S. Carrere, T. Faraut, F. de Carvalho-Niebel, J. Gouzy, P. Gamas, D. Capela, C. Bruand & T. Schiex, (2013) Next-

generation annotation of prokaryotic genomes with EuGene-P: application to *Sinorhizobium meliloti* 2011. *DNA Res* **20**: 339-353.

Scherr, N. & L. Nguyen, (2009) *Mycobacterium* versus *Streptomyces*-we are different, we are the same. *Curr. Opin. Microbiol.* **12**: 699-707.

Schluenzen, F., C. Takemoto, D. N. Wilson, T. Kaminishi, J. M. Harms, K. Hanawa-Suetsugu, W. Szaflarski, M. Kawazoe, M. Shirouzu, K. H. Nierhaus, S. Yokoyama & P. Fucini, (2006) The antibiotic kasugamycin mimics mRNA nucleotides to destabilize tRNA binding and inhibit canonical translation initiation. *Nat. Struct. Mol. Biol.* **13**: 871-878.

Schuwirth, B. S., J. M. Day, C. W. Hau, G. R. Janssen, A. E. Dahlberg, J. H. Cate & A. Vila-Sanjurjo, (2006) Structural analysis of kasugamycin inhibition of translation. *Nat. Struct. Mol. Biol.* **13**: 879-886.

Schwedock, J., J. R. McCormick, E. R. Angert, J. R. Nodwell & R. Losick, (1997) Assembly of the cell division protein FtsZ into ladder-like structures in the aerial hyphae of *Streptomyces coelicolor*. *Mol. Microbiol.* **25**: 847-858.

Serganov, A., L. L. Huang & D. J. Patel, (2009) Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* **458**: 233-U210.

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

Shin, J. H., A. K. Singh, D. J. Cheon & J. H. Roe, (2011) Activation of the SoxR Regulon in *Streptomyces coelicolor* by the Extracellular Form of the Pigmented Antibiotic Actinorhodin. *J. Bacteriol.* **193**: 75-81.

Shine, J. & L. Dalgarno, (1974) 3'-terminal sequence of *Escherichia coli* 16S rRNA: possible role in initiation and termination of protein synthesis. *P. Aust. Biochem. Soc.* **7**: 72-72.

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

Shiomi, K. & S. Omura, (2004) Antiparasitic agents produced by microorganisms. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **80**: 245-258.

Shu, D., L. Chen, W. H. Wang, Z. Y. Yu, C. Ren, W. W. Zhang, S. Yang, Y. H. Lu & W. H. Jiang, (2009) *afsQ1-Q2-sigQ* is a pleiotropic but conditionally required signal transduction system for both secondary metabolism and morphological development in *Streptomyces coelicolor*. *Appl. Microbiol. Biotechnol.* **81**: 1149-1160.

Sohoni, S. V., A. Fazio, C. T. Workman, I. Mijakovic & A. E. Lantz, (2014) Synthetic promoter library for modulation of actinorhodin production in *Streptomyces coelicolor* A3(2). *PloS one* **9**: e99701.

Soliveri, J., K. L. Brown, M. J. Buttner & K. F. Chater, (1992) Two promoters for the *whiB* sporulation gene of *Streptomyces coelicolor* A3(2) and their activities in relation to development. *J. Bacteriol.* **174**: 6215-6220.

Songia, S., A. Mortellaro, S. Taverna, C. Fornasiero, E. A. Scheiber, E. Erba, F. Colotta, A. Mantovani, A. M. Isetta & J. Golay, (1997) Characterization of the new

immunosuppressive drug undecylprodigiosin in human lymphocytes - Retinoblastoma protein, cyclin-dependent kinase-2, and cyclin-dependent kinase-4 as molecular targets. *J. Immunol.* **158**: 3987-3995.

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

Staunton, J. & K. J. Weissman, (2001) Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**: 380-416.

Storz, G., J. Vogel & K. M. Wassarman, (2011) Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol. Cell* **43**: 880-891.

Strauch, E., E. Takano, H. A. Baylis & M. J. Bibb, (1991) The stringent response in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **5**: 289-298.

Strohl, W. R., (1992) Compilation and Analysis of DNA-Sequences Associated with Apparent Streptomycete Promoters. *Nucleic Acids Res.* **20**: 961-974.

Sun, J. H., A. Hesketh & M. Bibb, (2001) Functional analysis of *relA* and *rshA*, two *relA/spoT* homologues of *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **183**: 3488-3498.

Swiatek-Polatynska, M. A., G. Bucca, E. Laing, J. Gubbens, F. Titgemeyer, C. P. Smith, S. Rigali & G. P. van Wezel, (2015) Genome-Wide Analysis of *In Vivo* Binding of the Master Regulator DasR in *Streptomyces coelicolor* Identifies Novel Non-Canonical Targets. *PloS one* **10**.

Swiatek, M. A., J. Gubbens, G. Bucca, E. Song, Y. H. Yang, E. Laing, B. G. Kim, C. P. Smith & G. P. van Wezel, (2013) The ROK family regulator Rok7B7 pleiotropically affects xylose utilization, carbon catabolite repression, and antibiotic production in *Streptomyces coelicolor*. *J. Bacteriol.* **195**: 1236-1248.

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

Taguchi, T., T. Ebihara, A. Furukawa, Y. Hidaka, R. Ariga, S. Okamoto & K. Ichinose, (2012) Identification of the actinorhodin monomer and its related compound from a deletion mutant of the *act*VA-ORF4 gene of *Streptomyces coelicolor* A3(2). *Bioorg. Med. Chem. Lett.* **22**: 5041-5045.

Taguchi, T., K. Itou, Y. Ebizuka, F. Malpartida, D. A. Hopwood, C. M. Surti, K. I. Booker-Milburn, G. R. Stephenson & K. Ichinose, (2000) Chemical characterisation of disruptants of the *Streptomyces coelicolor* A3(2) actVI genes involved in actinorhodin biosynthesis. *The Journal of antibiotics* **53**: 144-152.

Taguchi, T., S. Okamoto, K. Hasegawa & K. Ichinose, (2011) Epoxyquinone formation catalyzed by a two-component flavin-dependent monooxygenase involved in biosynthesis of the antibiotic actinorhodin. *ChemBioChem* **12**: 2767-2773.

Taguchi, T., M. Yabe, H. Odaki, M. Shinozaki, M. Metsa-Ketela, T. Arai, S. Okamoto & K. Ichinose, (2013) Biosynthetic conclusions from the functional dissection of oxygenases for biosynthesis of actinorhodin and related *Streptomyces* antibiotics. *Chem. Biol.* **20**: 510-520.

Tahlan, K., S. K. Ahn, A. Sing, T. D. Bodnaruk, A. R. Willems, A. R. Davidson & J. R. Nodwell, (2007) Initiation of actinorhodin export in *Streptomyces coelicolor*. *Mol. Microbiol.* **63**: 951-961.

Takano, E., (2006) γ-Butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. *Curr. Opin. Microbiol.* **9**: 287-294.

Takano, E., H. Kinoshita, V. Mersinias, G. Bucca, G. Hotchkiss, T. Nihira, C. P. Smith, M. Bibb, W. Wohlleben & K. Chater, (2005a) A bacterial hormone (the SCB1) directly controls the expression of a pathway-specific regulatory gene in the cryptic type I polyketide biosynthetic gene cluster of *Streptomyces coelicolor*. *Mol. Microbiol.* **56**: 465-479.

Takano, E., M. Tao, F. Long, M. J. Bibb, L. Wang, W. Li, M. J. Buttner, Z. X. Deng & K. F. Chater, (2003) A rare leucine codon in *adpA* is implicated in the morphological defect of *bldA* mutants of *Streptomyces coelicolor*. *Mol. Microbiol.* **50**: 475-486.

Takano, H., S. Obitsu, T. Beppu & K. Ueda, (2005b) Light-induced carotenogenesis in *Streptomyces coelicolor* A3(2): identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster. *J. Bacteriol.* **187**: 1825-1832.

Thomason, M. K., T. Bischler, S. K. Eisenbart, K. U. Forstner, A. X. Zhang, A. Herbig, K. Nieselt, C. M. Sharma & G. Storz, (2015) Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in Escherichia coli. *J. Bacteriol.* **197**: 18-28.

Touzain, F., S. Schbath, I. Debled-Rennesson, B. Aigle, G. Kucherov & P. Leblond, (2008) SIGffRid: A tool to search for sigma factor binding sites in bacterial genomes using comparative approach and biologically driven statistics. *Bmc Bioinformatics* **9**.

Towle, J. E., (2007) AtrA-mediated transcriptional regulation in *Streptomyces* secondary metabolite production and development. In: Faculty of Biological Sciences. Leeds: University of Leeds, pp. 208.

Traag, B. A., G. H. Kelemen & G. P. van Wezel, (2004) Transcription of the sporulation gene *ssgA* is activated by the *IclR*-type regulator SsgR in a *whi*-independent manner in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **53**: 985-1000.

Traag, B. A. & G. P. van Wezel, (2007) The SsgA-like proteins in actinomycetes: small proteins up to a big task. In: 14th Meeting of the International Symposia on the Biology of Actinomycetes (ISBA). Newcastle upon Tyne, ENGLAND, pp. 85-97.

Traag, B. A. & G. P. van Wezel, (2008) The SsgA-like proteins in actinomycetes: small proteins up to a big task. *Anton. Leeuw. Int. J. G.* **94**: 85-97.

Travers, A. A., (1980) Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. *J. Bacteriol.* **141**: 973-976.

Tuerk, C. & L. Gold, (1990) Systematic evolution of ligands by exponential enrichment - RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**: 505-510.

Uguru, G. C., K. E. Stephens, J. A. Stead, J. E. Towle, S. Baumberg & K. J. McDowall, (2005) Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. *Mol. Microbiol.* **58**: 131-150.

van Wezel, G. P. & K. J. McDowall, (2011) The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Nat. Prod. Rep.* **28**: 1311-1333.

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

Vertesy, L., E. Ehlers, H. Kogler, M. Kurz, J. Meiwes, G. Seibert, M. Vogel & P. Hammann, (2000) Friulimicins: Novel lipopeptide antibiotics with peptidoglycan synthesis inhibiting activity from Actinoplanes friuliensis sp nov II. Isolation and structural characterization. *J. Antibiot.* **53**: 816-827.

Vesper, O., S. Amitai, M. Belitsky, K. Byrgazov, A. C. Kaberdina, H. Engelberg-Kulka & I. Moll, (2011) Selective Translation of Leaderless mRNAs by Specialized Ribosomes Generated by MazF in *Escherichia coli*. *Cell* **147**: 147-157.

Vockenhuber, M. P., C. M. Sharma, M. G. Statt, D. Schmidt, Z. Xu, S. Dietrich, H. Liesegang, D. H. Mathews & B. Suess, (2011) Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*. *Rna Biol* **8**: 468-477.

Voeikova, T. A., (1999) Conjugative transfer of a plasmid from *Escherichia coli* to various strains of the order Actinomycetales. *Genetika* **35**: 1626-1633.

Vogtli, M., P. C. Chang & S. N. Cohen, (1994) *afsR2*: a previously undetected gene encoding a 63-amino-acid protein that stimulates antibiotic production in *Streptomyces lividans*. *Mol. Microbiol.* **14**: 643-653.

Vvedenskaya, I. O., J. S. Sharp, S. R. Goldman, P. N. Kanabar, J. Livny, S. L. Dove & B. E. Nickels, (2012) Growth phase-dependent control of transcription start site selection and gene expression by nanoRNAs. *Genes Dev.* **26**: 1498-1507.

Waksman, S. A. & A. T. Henrici, (1943) The nomenclature and classification of the actinomycetes. *J. Bacteriol.* **46**: 337-341.

Wang, L. Q., X. Y. Tian, J. Wang, H. H. Yang, K. Q. Fan, G. M. Xu, K. Q. Yang & H. R. Tan, (2009) Autoregulation of antibiotic biosynthesis by binding of the end product to an atypical response regulator. *Proc. Natl. Acad. Sci. USA* **106**: 8617-8622.

Wang, R., Y. Mast, J. Wang, W. Zhang, G. Zhao, W. Wohlleben, Y. Lu & W. Jiang, (2013) Identification of two-component system AfsQ1/Q2 regulon and its cross-regulation with GlnR in *Streptomyces coelicolor*. *Mol. Microbiol.* **87**: 30-48.

Wernersson, R., A. S. Juncker & H. B. Nielsen, (2007) Probe selection for DNA microarrays using OligoWiz. *Nat Protoc* **2**: 2677-2691.

White, J. & M. Bibb, (1997) *bldA* dependence of undecylprodigiosin production in *Streptomyces coelicolor* A3(2) involves a pathway-specific regulatory cascade. *J. Bacteriol.* **179**: 627-633.

Wietzorrek, A. & M. Bibb, (1997) A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. *Mol. Microbiol.* **25**: 1181-1184.

Willey, J., J. Schwedock & R. Losick, (1993) Multiple Extracellular Signals Govern the Production of a Morphogenetic Protein Involved in Aerial Mycelium Formation by *Streptomyces coelicolor*. *Genes Dev.* **7**: 895-903.

Willey, J. M. & A. A. Gaskell, (2011) Morphogenetic signaling molecules of the streptomycetes. *Chem. Rev.* **111**: 174-187.

Willey, J. M., A. Willems, S. Kodani & J. R. Nodwell, (2006) Morphogenetic surfactants and their role in the formation of aerial hyphae in *Streptomyces coelicolor*. *Mol. Microbiol.* **59**: 731-742.

Williamson, N. R., P. C. Fineran, F. J. Leeper & G. P. C. Salmond, (2006) The biosynthesis and regulation of bacterial prodiginines. *Nature Rev. Microbiol.* **4**: 887-899.

Wright, L. F. & D. A. Hopwood, (1976) Actinorhodin is a chromosomally-determined antibiotic in *Streptomyces coelicolar* A3(2). *J. Gen. Microbiol.* **96**: 289-297.

Xu, Y., A. Willems, C. Au-Yeung, K. Tahlan & J. R. Nodwell, (2012) A two-step mechanism for the activation of actinorhodin export and resistance in *Streptomyces coelicolor*. *Mbio* **3**: e00191-00112.

Yamazaki, H., A. Tomono, Y. Ohnishi & S. Horinouchi, (2004) DNA-binding specificity of AdpA, a transcriptional activator in the A-factor regulatory cascade in *Streptomyces griseus*. *Mol. Microbiol.* **53**: 555-572.

Yim, G., H. H. Wang & J. Davies, (2007) Antibiotics as signalling molecules. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **362**: 1195-1200.

Yu, Z., H. Zhu, F. Dang, W. Zhang, Z. Qin, S. Yang, H. Tan, Y. Lu & W. Jiang, (2012) Differential regulation of antibiotic biosynthesis by DraR-K, a novel two-component system in *Streptomyces coelicolor*. *Mol. Microbiol.* **85**: 535-556.

Zhang, Y. W., U. Werling & W. Edelmann, (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40**.

Zotchev, S. B., (2012) *Antibiotics, In: Natural Products in Chemical Biology. Natanya Civjan (Ed). John Wiley & Sons, Inc., Hoboken, New Jersey*.