Investigating the role of *rrf2* genes: *nsrR* and *rsrR* in *S. coelicolor* and *S. venezuelae*.

John Tyson Munnoch

A thesis submitted in fulfilment of the requirements for a degree of Doctor of Philosophy at the University of East Anglia.

School of Biological Sciences, University of East Anglia.

December 2016.

Publications arising from work in this thesis

Crack, J., Munnoch, J., Dodd, E., Knowles, F., Al Bassam, M., Kamali, S., Holland, A., Cramer, S., Hamilton, C., Johnson, M., Thomson, A., Hutchings, M., and Le Brun, N. (2015) NsrR from *Streptomyces coelicolor* is a Nitric Oxide-Sensing [4Fe-4S] Cluster Protein with a Specialized Regulatory Function. *J. Biol. Chem. 290, 12689–12704*

Crack, J. C., Svistunenko, D. A., Munnoch, J., Thomson, A. J., Hutchings, M. I., and Le Brun, N. E. (2016) Differentiated, promoter-specific response of [4Fe-4S] NsrR DNA-binding to reaction with nitric oxide. *J. Biol. Chem. 291, 8663–8672*

Munnoch, J. T., Pellicer-Martinez, M. T., Svistunenko, D. A., Crack, J. C., Le Brun, N. E., and Hutchings, M. I. (2016) Characterization of a putative NsrR homologue in *Streptomyces venezuelae* reveals a new member of the Rrf2 superfamily. *Sci. Rep. 6, 31597*

Munnoch, J. T., Widdick, D. A., Chandra, G., Sutcliffe, I. C., Palmer, T., and Hutchings, M. I. (2016) Cosmid based mutagenesis causes genetic instability in *Streptomyces coelicolor*, as shown by targeting of the lipoprotein signal peptidase gene. *Sci. Rep. 6, 29495*

Each research paper can be found in the appendices at the end of this document.

Acknowledgements

Contents

Figures

Tables

Abbreviations

1G	10% Glycerol
2G	20% Glycerol
4G	40% Glycerol
5'PPP	5' triphosphates
5'P	5'monophosphates
Tex	5'P-dependant terminator exonuclease
ACT	Actinorhodin
apr	Apramycin
bld	bald
bp	Base pair
CDA	calcium-dependent antibiotic
CCCP	carbonyl cyanidem-chlorophenyl-hydrazone
ChIP-seq	Chromatin-immunoprecipitation sequencing
CymR	Cysteine metabolism Regulator
dRNA-seq	Differential RNA sequencing
dH$_2$O	distilled water
EMSA	Electrophoretic mobility shift assay
ER	Endoplasmic reticulum
EHEC	Enterohemorrhagic *E. coli*
Fur	Fe uptake regulator
FNR	Fumerate and Nitrate Reduction regulator
GATC	GATC BIOTECH
GS	Glutamate synthase
hmpA1p	*hmpA1* promoter
hmpA2p	*hmpA2* promoter
H$_2$O$_2$	Hydrogen peroxide
HO$^{\cdot}$	Hydroxyl radical
hyg	Hygromycin

ISC	Iron Sulfur Cluster
IscR	Iron sulfur cluster Regulator
[Fe-S]	Iron-sulfur cluster
KO	Knockout
lsp	lipoprotein signal peptidase
LEE	Locus of enterocyte effacement
MIC	Minimum inhibitory concentration
MSH	Mycothiol
MSNO	Mycothiol, NO conjugate
NGS	Next generation sequencing
Nur	Ni uptake regulator
NO	Nitric oxide
NsrR	Nitric oxide-sensitive Regulator
Hmp	NO-detoxifying flavohemoglobin
nsrRp	*nsrR* promoter
nt	Nucleotide
OD	Optical density
oriT	Origin of transfer
OE	Overexpression
MSSM	Oxidesed mycothiol
RsrRox	Oxidised RsrR
pJMxxx	Plasmids from this work
PCR	Polymerase chain reaction
JMxxxx	Primers from this work
PCC	protein conducting channel
PMF	Proton motive force
RES	Reactive electrophilic species
RNS	Reactive nitrogen species
ROS	Reactive oxygen species

RsrR	Redox sensitive response regulator
RsrRred	Reduced RsrR
RT	Room temperature
rsrRp	*rsrR* promoter
SEM	Scanning electron microscopy
SNPs	single nucleotide polymorphisms
sRNA	small RNA
JTMxxx	Strains from this work
O_2^-	Superoxide anion
SOD	Superoxide dismutase
Sec	The general Secretary pathway
TGAC	The Genome Analysis Centre
TAP	Tobacco Acid Pyrophosphatase
TSS	Transcription start site
TMD	transmembrane domains
TAT	Twin Arginine Transport pathway
T1-6SS	Type I-VI secretion system
RED	Undicylprodigiosin
UTR	Untranslated region
whi	whi
wbl	*whiB*-like
WT	Wildtype

Abstract

This thesis furthers work on and identifies two Rrf2 regulators with distinct biochemistry and regulons involved in redox homeostasis in *S. coelicolor* and *S. venezuelae*.

Members of the Rrf2 superfamily of transcription factors are widespread in bacteria and have a range of known regulatory functions including sensing and responding to nitric oxide stress, iron limitation, cysteine availability and the iron sulfur cluster status of the cell. Here we report further study of the Rrf2 protein named NsrR and the identification and characterisation of a novel putative redox sensor RsrR in *S. coelicolor and S. venezuelae* respectively.

NsrR is a global regulator in a range of bacteria, controlling gene expression of a diverse regulon (~35 to >60 genes in *B. subtilis* and *E. coli,* respectively) in response to nitrogen stress through a [4Fe-4S] co-factor. We show through ChIP-seq and *in vitro* DNA-binding studies that NsrR controls only *hmpA1, hmpA2* and *nsrR* by binding to a 11-bp inverted repeat sequence upstream of each gene. Hmp is an NO-detoxifying flavohemoglobin, indicating that NsrR contains a specialised regulon responsible for NO detoxification. We also report an NsrR-independent, NO dependent delay of sporulation in *S. coelicolor* through an unknown route*.*

RsrR, encoded by a diverse group of actinomycetes, contains an oxygen stable, putative redox-sensing [2Fe-2S] cluster that cycles between an oxidised and reduced state. ChIP-seq suggests RsrR binds strongly to an 11-3-11bp inverted repeat to, at least, 16 target sites and more weakly to an 11bp half site at the remaining >100 targets while dRNA-seq indicates a distinctly different set of targets while suggesting the main function of RsrR is regulation of the *sven6562* (*nmrA*) gene, which is transcribed divergently from RsrR. NmrA is a putative LysR-type regulator containing a C-terminal DNA-binding domain and an N-terminal NmrA domain that we hypothesis senses redox poise by binding NAD(P)+ but not NAD(P)H.

This thesis is 238 pages and ~64,000 words in length.

1 Introduction

1.1 Actinobacteria – the good, the bad and the tasty

Actinobacteria, a GC rich phylum of Gram-positive bacteria, contains some of the most clinically and industrially important bacteria known to man. The best-studied order of actinobacteria is the Actinomycetales, commonly known as actinomycetes. Due to their broad enzymatic versatility in processing often difficult or recalcitrant chemicals, the production of important compounds for the food industry (*Corynaebacterium glutamicum*) (Udaka 2008) as well as their capacity to produce a spectrum of clinically important bioactive secondary metabolites (*Streptomyces sp.*) (Hopwood 2007), actinomycetes have been indispensible to industry for decades. However, contained within this order are also some of the most globally destructive human pathogens. These include the causative agents of tuberculosis, leprosy and diphtheria: *Mycobacterium tuberculosis, Mycobacterium leprae* and *Corynebacterium diphtheriae* respectively (Madigan & Martinko 2006).

1.1.1 Actinomycetes in the food industry: *Corynebacterium glutamicum*

The soil dwelling, aerobic, Gram-positive bacterium *C. glutamicum* has earned great fame in the biotechnological and industrial sectors since its isolation in the late 1950's by Shigezo Udaka. *C. glutamicum* is best known as a producer of the amino acid L-glutamate, which in the early 20th century was of great interest to the Japanese population as a taste enhancing agent (Udaka 2008). *C. glutamicum* is an efficient producer of relatively large quantities of L-glutamate and L-lysine by fermentation, unlike most current multimillion tonne per annum chemical processes which result in racemic mixtures of both the L and D stereoisomers (Kinshita 2005). Due to its use in the industrial sector *C. glutamicum* is one of the best-characterised actinomycetes.

1.1.2 Actinomycetes and antibiotics: the heralding of *Streptomyces*

Antibiotic, the name first coined by P. Vuillemin in 1889, was a general term used to describe the destruction of an organism by another (Hopwood 2007). Since the discovery of the first natural product antibiotic penicillin in 1928, natural product antibiotics have almost constantly been the focus of rigorous research. In 1943 the isolation and subsequent mass production of streptomycin (from *Streptomyces griseus*), the first effective treatment of tuberculosis, by Selman Waksman's laboratory heralded the modern era of antibiotic use in the clinical setting. For the next 10 years an increasing number of

antibiotics were isolated and brought to market during the "Golden age" of discovery, most of which were isolated from actinomycetes, particularly *Streptomyces* bacteria, which are ultimately responsible for more than 80% of antibiotics used today (Hopwood 2007; de Lima Procópio et al. 2012).

1.1.3 Actinomycete human pathogens: a scourge past and present

As has already been mentioned, *M. tuberculosis, M. leprae and C. diptheriae* are major human pathogens. In addition to these there are a range of opportunistic pathogens: *M. avium, M. chelonae, M. fulaceum, and M kansasii* (Madigan & Martinko 2006). Tuberculosis, symptomatic with severe coughing, fever and chest pains, was responsible for nearly 1.5 million deaths in 2013, second only to HIV/AIDS in mortality rate (Fogel 2015; Ahsan 2015). *M. tuberculosis*, with the advent of the BCG vaccine in 1908 and the discovery of streptomycin, went from being a disease deadlier than the plaque, to being nearly eradicated within the 20th century. However, with the onset of severe antibiotic resistance and the increase in immunocompromised individuals living with HIV, tuberculosis has returned as a major public health issue.

1.2 *Streptomyces*

An important genus within the actinobacteria, *Streptomyces*, are the focus of my research and are frequently studied because they are ubiquitous in soils, they are unsurpassed as secondary metabolite producers and they have a complex developmental life cycle (Figure 1.1). *Streptomyces* are saprophytic, obligate aerobic organisms, obtaining nutrients from the soil by the release of hydrolytic exo-enzymes that can break down insoluble and otherwise recalcitrant plant, insect and fungi material (Chater et al. 2010). Their life cycle, consisting of distinct developmental stages involving hyphal tip extension and polar growth (Flärdh & Buttner 2009), is similar in appearance to that of fungi, which originally led to their misclassification. *Streptomyces* contain both pathogenic and non-pathogenic strains. The most commonly studied pathogen because of its economical importance is *S. scabies,* which can infect the roots of many plants and is a common cause of potato scab disease (Widdick et al. 2011). The strains *S. sudanensis* and *S. somaliensis* are capable of causing human infections, symptomatic with severe and debilitating deep tissue and bone infections (Kirby et al. 2012).
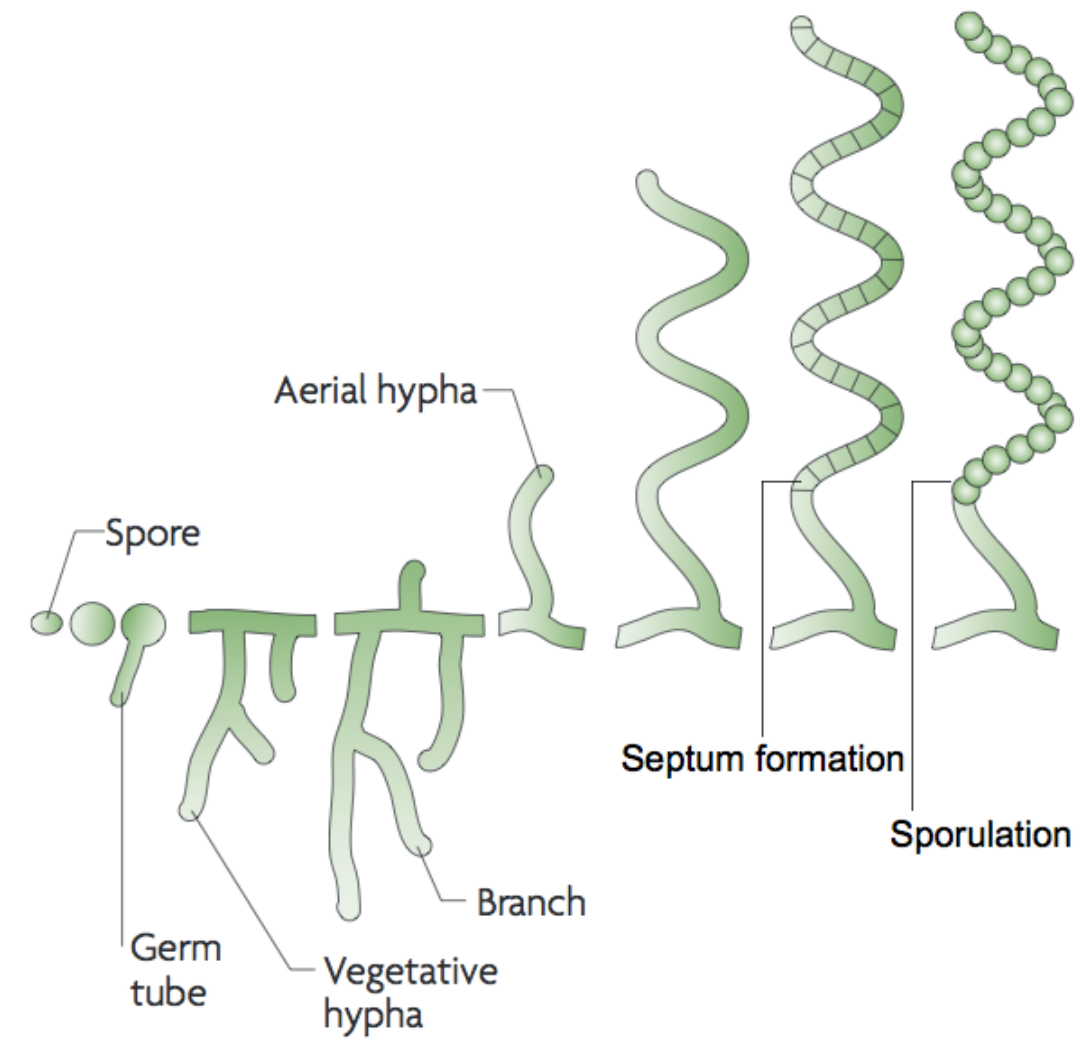
Figure 1-1 The *Streptomyces* life cycle. Modified from Flardh et al., (2009). Starting from a unigenomic spore, germ tubes form, growing into the environment seeking nutrients. From these germ tubes vegetative hyphae develop, branching repeatedly until the signal to switch towards reproduction is received. Upon onset of reproduction, aerial hyphae form, growing out of the soil environment, and begin placing down cell division septa in preparation for sporulation. Sporulation then occurs resulting in chains of unigenomic spores that are dispersed into the environment, providing an effective escape mechanism.

1.2.1 Model organisms: *S. coelicolor* and *S. venezuelae*

Model organisms are incredibly important for the progression of a given field. Particularly in microbial genetics, they usually consist of genetically tractable organisms with similar but also unique characteristics that make them an indispensible tool. This is the case for both model organisms used in this study and widely within the *Streptomyces* field: *S. coelicolor* and *S. venezuelae*. *S. coelicolor* and the subsequent modified laboratory strains were originally studied because of their pigmented antibiotics: actinorhodin (ACT) undicylprodigiosin (RED) and the calcium-dependent antibiotic (CDA) which are blue, red

19

and colourless respectively (Xu et al. 2012; Schäberle et al. 2014; Hojati et al. 2002). The production of these antibiotics has been shown to be up regulated in strains suffering stresses, genetic damage and/or instability, making them particular useful as phenotypic indicators when carrying out mutational studies. The genome sequence of *S. coelicolor* was completed relatively early; over a decade ago (Bentley et al. 2002), providing an unparalleled model organism for its time. *S. venezuelae,* on the other hand, was developed as a model *Streptomycetes* more recently because of its rapid progression through development, its ability to sporulate to near completion in liquid media (Glazebrook et al. 1990), a relatively rare ability in other *Streptomyces* strains, as well as the recent endeavour by the John Innes Centre to provide the genome sequence and a wealth of microarray expression data spanning the developmental life cycle (Pullan et al. 2011).

1.2.2   *Streptomyces* differentiation: *whi* and *bld,* a midlife crisis

Streptomyces growth and differentiation is a complex topic. Unlike the canonical method of bacterial cell growth and cell division by binary fission, mycelial growth of *Streptomyces* bacteria by apical tip extension allows for a unique opportunity to study the usually essential steps of cell division, which is dispensable in these complex organisms. As illustrated in Figure 1.1, there are several key stages of the *Streptomyces* life cycle: Spore germination, vegetative growth, the switch to and growth of aerial hyphae, septum formation and sporulation (Flärdh & Buttner 2009). Each growth stage (vegetative, aerial and sporulation) has a distinct phenotypic appearance under standard laboratory conditions: vegetative growth results in clear, shiny colonies; aerial hyphae result in a white, almost furry appearance; while the onset of sporulation results in cell division and the production of an associated spore pigment (grey and green for *S. coelicolor* and *S. venezuelae* respectively). Two major classes of genes, known as the bald (*bld*) and white (*whi*) genes, are global, master regulators essential for the switch from vegetative to aerial hyphal growth and aerial growth to sporulation respectively (Elliot et al. 2008; Flärdh & Buttner 2009; McCormick & Flardh 2012). Mutations within *bld* genes result in an inability to form aerial hyphae and therefore "bald" looking colonies (Chater 1993) while mutations in the *whi* genes result in an inability to complete sporulation or produce characteristic spore pigment and therefore white colonies (Ryding et al. 1998).

1.2.2.1   *bld* gene cascade

Various studies of the *bld* genes have implicated them in a complex cascade to switch to aerial hyphal growth (Chater 2001; Claessen et al. 2006) as illustrated in Figure 1.2. One of the most important effects of *bld* gene regulation is the production and release of SapB,

a surfactant that reduces surface tension above the mycelium (Willey et al. 1991). SapB, in combination with the hydrophobic sheath proteins known as Rodlins (Claessen et al. 2002; Claessen et al. 2004) and Chaplins (Claessen et al. 2003), allow the aerial mycelium to break through the weakened surface tension and grow into the air. The *bld* genes have been placed in a hierarchy of early to late acting genes. This hierarchy is based on the ability to express a later acting *bld* gene to complement a *bld* mutant phenotype. In order, the early to late acting *bld* genes are: *bldJ, K, A, H, N, G, C, DM.* Each will be discussed below.
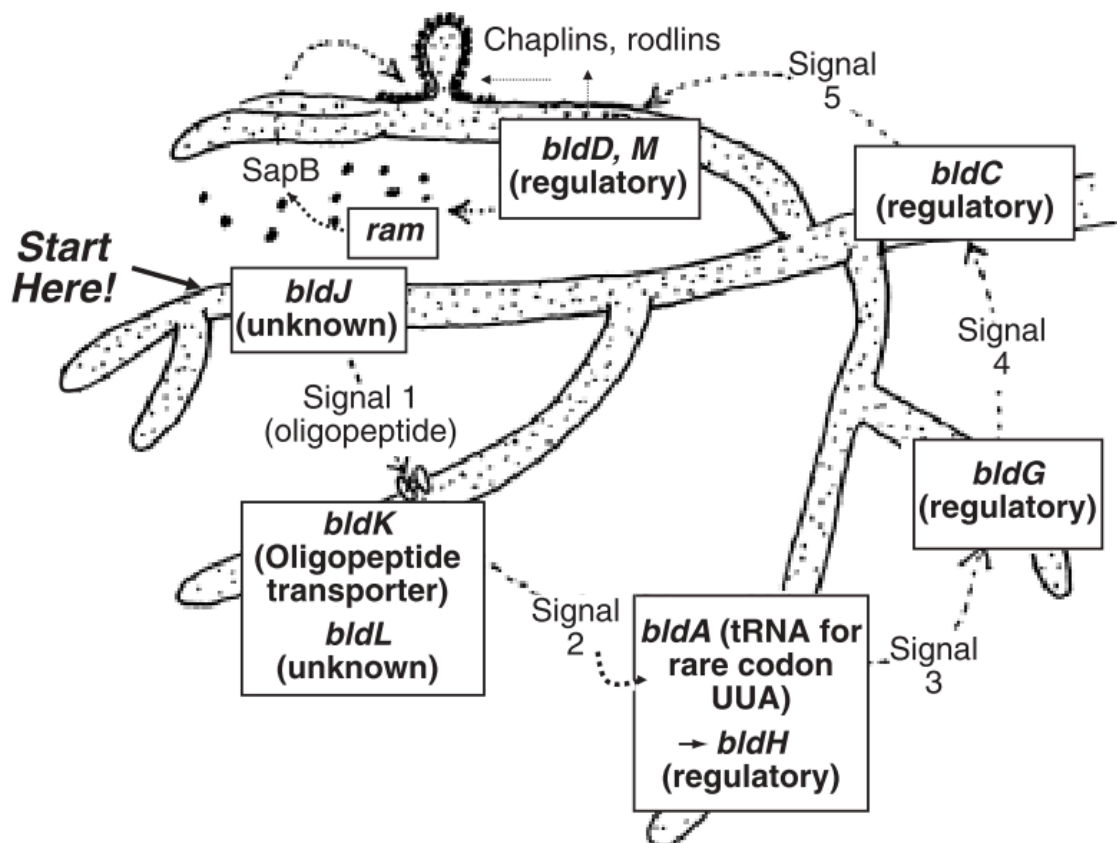
Figure 1-2 Extracellular signalling cascade for signalling and switch to aerial hyphal growth dependent on the *bld* genes (Keith F. Chater & Chandra 2006).

bldK encodes an oligopeptide importer that is thought to import a *bldJ*-dependent extracellular signalling molecule (Nodwell & Losick 1998; Nodwell et al. 1996). The oligopeptide produced by *bldJ* has not been fully characterised but is thought to provide a signal in response to environmental cues to begin sporulation. The link between *bldK* and *bldA* is currently unknown, as is the case with links between most of the members in the cascade, but this could be by direct signalling by the *bldJ* oligopeptide. *bldA* encodes a leucine tRNA which recognises the rare codon UUA that is found in ~2-3% of genes in the *S. coelicolor* genome, however the *bldA* phenotype is likely due to the loss of the UUA

dependent repressor, AdpA/BldH (Takano et al. 2003). The BldH regulon, although not fully defined, contains: SsgA, which alongside SsgB regulates cell division by FtsZ localisation *ramR* (Yamazaki et al. 2000; Ueda et al. 2005; Xu et al. 2010), involved in SapB synthesis (Kodani et al. 2004), and *bldN* (Yamazaki et al. 2000), a sigma factor essential for aerial growth. A recent study has further defined the regulon showing the *bldH* dependent changes in expression at stationary phase as well as direct interaction studies of BldH with, among others, the *ramR* promoter (Guyet et al. 2014).

bldG and a downstream anti-sigma factor, referred to as ORF3, are both up regulated in early aerial development (Bignell et al. 2000) and appears to be under the control of σ^H, an alternative stress response sigma factor (Sevcikova et al. 2010). How they are linked and their effect on aerial development is unknown. Both *bldG* and *bldH* are necessary for *bldN* transcription (Bibb et al. 2000). *bldC* function has not yet been linked to its phenotype, however, expression of it can rescue a *bldG* mutation.

bldD and *bldM* are master regulators of development and form the final stage where known later stage *bld* genes can complement SapB production. *bldD* is a cyclic-di-GMP binding transcriptional regulator that represses target genes in its dimeric holo-form, (Tschowri et al. 2014) which has previously been shown to repress genes involved in sporulation (den Hengst et al. 2010). The BldD regulon has at least 167 genes, including 42 regulatory proteins; among these are: *bld genes* (*bldA, bldC, bldH, bldM, bldN*), *whi* genes, *(whiG whiB)* and cell division genes *(ftsZ, ssgA, ssgB and sffA)*, to list but a few (den Hengst et al. 2010; McCormick 2009). The absence of cyclic-di-GMP appears to be a signal for the initiation of sporulation caused by a reduced level of the BldD target CdgB, a diguanylate cyclase responsible for cyclic-di-GMP production (Tran et al. 2011; Den Hengst et al. 2010).

bldM and *whiI* each encode two orphan response regulators of the NarL/FixJ subfamily of two component system response regulators (Al-Bassam et al. 2014). Recent work has shown that BldM homodimers regulate early sporulation, while BldM/WhiI heterodimers appear to regulate late sporulation, specifically and exactly correlating to the WhiI regulon where it is the heterodimer instead of either form of homodimer that regulates the "WhiI" regulon.   *BldN* regulates *bldM* while WhiG regulates *whiI*.   This functional heterodimerisation is an interesting regulatory function which has been suggested to play a timing and fine-tuning role during complex and sensitive cellular developmental events. Following *bldD* and *bldM* activation, we begin to see the *whi* genes truly come into play.

1.2.2.2 *whi* genes

Mutations in the *whi* genes inhibit the progression of sporulation. Early acting *whi* genes including *whiA, B, G, H, I* and *J* (Chater 1972), when disrupted, fail to develop sporulation septa, while late acting *whi* genes, *whiD, whiE* and *sigF,* achieve septation but fail to phenotypically complete sporulation. Each of the early acting *whi* genes is ultimately required for production of σH, a spore specific sigma factor. WhiE is responsible for synthesis of the polyketide spore pigment and FtsZ expression in aerial hyphae is essential for septum formation. In terms of the hierarchy of these genes, the order of progression generally has been elucidated to be, *whiG, whiH, whiA/B,* then *whiI* (Figure 1.3). Little is known about *whiJ* but it, along with its flanking genes, at least in *S. coelicolor,* putatively represent a possible regulatory system for repressing sporulation (Aínsa et al. 2010).
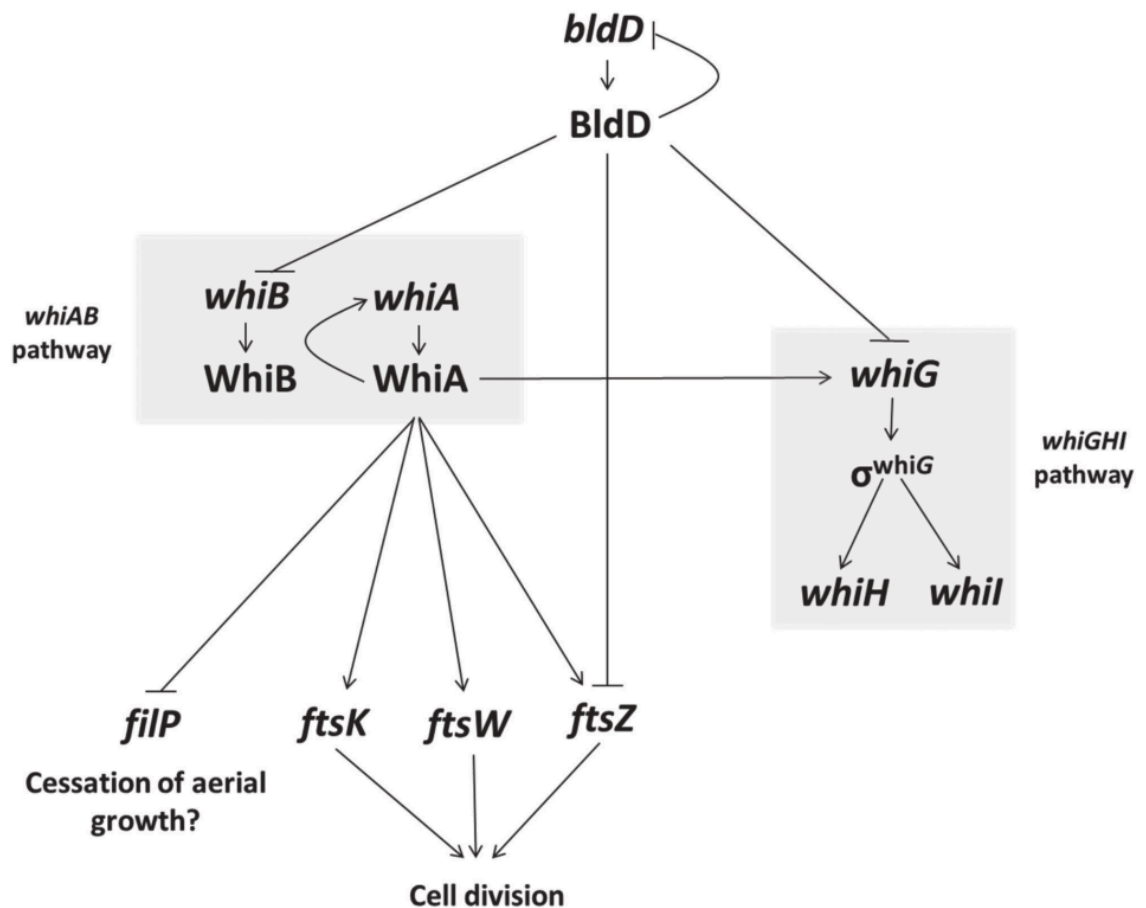
Figure 1-3 A model of the *whi* gene regulatory network (Bush et al. 2013).

whiG is thought to encode an essential sigma factor for initiation of sporulation (Chater et al. 1989). The *whiG* regulon has not been fully elucidated, however, three of its direct targets include *whiH, whiI* and *whiA* (Ryding et al. 1998; Aínsa et al. 2000; Kaiser & Stoddard 2011). *whiH* encodes a transcriptional regulator of the GntR family (Ryding et al.

1998), but little is known about its overall function as the only known target is *whiH* itself (Ryding et al. 1998). Phenotypically, *whiH* mutants show loosely coiled aerial hyphae that slightly fragment with an absence of spore septa (Chater 1972). *whiI* mutants have similar phenotypes to the *whiH* mutants, however, they form spore septa that are abnormally spaced, causing the formation of different sized spores (Chater 1972). Before the recent work describing the *bldM/whiI* heterodimers (Al-Bassam et al. 2014), described previously, little was known about the *whiI* regulon. Significant steps have been made recently in the study of *whiA* (Buttner 2013). Shown to be an unusual transcription factor, *whiA,* similar to eukaryotic homing endonucleases, regulates ~240 transcriptional units, acting in almost equal parts as an activator and repressor. The majority of WhiA target binding sites contain a short, asymmetric repeat sequence (GACAC), and the targets include *ftsZ, ftsW, ftsK* and *whiG.* WhiA and WhiB have been shown to effect the expression profile of each other in a currently unknown mechanism (Jakimowicz et al. 2006) and their respective mutants have a similar phenotype, with long, tightly coiled aerial hyphae, lacking spore septa (Chater 1972; Buttner 2013). *whiH and bldM/whiI* and *whiA/B* appear to represent independent pathways (Chater 2001; Buttner 2013; Al-Bassam et al. 2014). *whiB,* the first member of the Wbl (WhiB-like) protein family (discussed additionally in section 1.3.3.2.1), along with WhiD, all contain appropriately positioned cysteine residues responsible for the coordination of a [4Fe-4S] cluster. The *whiB* phenotype is indistinguishable from the *whiA* phenotype, while *whiD* results in fully formed spore septa, but the spores are thin walled, lacking spore pigment and prone to lysis. Little is known about their functional roles within a *Streptomyces* cell, though they are thought to be DNA-binding proteins. A greater body of work has been presented on the Wbl proteins from *C. glutamicum, M. smegmatis* and *M. tuberculosis*. The Wbl proteins will be further discussed in section 1.3.3.3.1. FtsZ is responsible for the dispensable, but not irrelevant, role of cell division in mycelial actinomycetes (McCormick 2009) is also a *whi* protein, as septum formation in spore chains as well as periodic compartmentalisation of hyphae is absent in an *ftsZ* mutant. Interestingly, the early stage *whi* phenotypes can be rescued by constitutive expression of FtsZ (Willemse et al. 2012) indicating that, like the *bld* cascade leads to SapB expression, the *whi* genes are ultimately essential for the correct spacial and temporal expression of FtsZ.

Many other genes have key roles in *Streptomyces* development but the *bld* and *whi* genes contain the more historically well-characterised and are key phenotypic markers for genetic studies. Among the plethora of regulators within the cell, including the *bld* and *whi* genes, we are particularly interested in those associated with redox stress.

1.3   Soil environment and stress

Streptomyces bacteria are ubiquitous within the soil, a compound they produce, geosmin, first described in 1891 by Berthelot et al, is responsible for the earthy smell, following rainstorms (J. Jiang et al. 2007; Gerber & Lechevalier 1965) also known as petrichor. The complex nutrient availability of soil spawns a competitive environment for microorganisms. When considering resources as a whole, the soil is a rich environment but when considering their relative accessibility to microbes, it can be a poor nutrient environment, as much is locked within complex sources that are typically difficult to metabolise. For those organisms, including *Streptomyces*, that can access these resources, they are often set upon by opportunists and must defend their resources, often using antibiotics to achieve this. Ultimately, this competition drives the constant evolutionary battle of antibiotic production and resistance within the soil environment between producers and their co-habitants.

In combination and as part of this competitive stress, microbes are constantly under a barrage of changing conditions from temperature to pH but also from both oxidative and nitrosative stress from internal processes and external exposure as a result of the environment and other microbes. In the next section we will be focusing on redox stress, the sources and types of redox-active compounds and the methods bacteria have developed to sense and respond to them, with a particular focus on *Streptomyces* [Fe-S] cluster containing transcription factors and specifically the Rrf2 family of proteins.

1.3.1   Redox stress: ROS/RNS/RES

It is important that cells manage their internal redox state. Reactive oxygen species (ROS), reactive nitrogen species (RNS) and reactive electrophilic species (RES) can all compromise this balance in a range of ways (Antelmann & Helmann 2011).   The production of primary and secondary reactive species is summarised in Figure 1.4 showing the physiological and pathological routes.

Hydrogen peroxide (H_2O_2), peroxides, superoxide anion (O_2^-), hypohalous acids and the hydroxyl radical ($HO^.$) are all ROS capable of damaging cellular components (Imlay 2003), although there is no evidence for ROS directly causing cell death (Imlay 2015). Endogenous sources of ROS arise from the incomplete reduction of molecular oxygen (Figure 1.5) by autoxidation of reducing enzymes (Messner & Imlay 2002), while evidence has been presented that they are not simply formed as aerobic respiratory intermediates (Minghetti & Gennis 1988). Exogenous sources of these ROS are formed during the oxidative burst produced both by human and plant cells during the immune response

(Slauch 2011; Yoshioka et al. 2008) and environmental factors including pollution and ionising radiation.
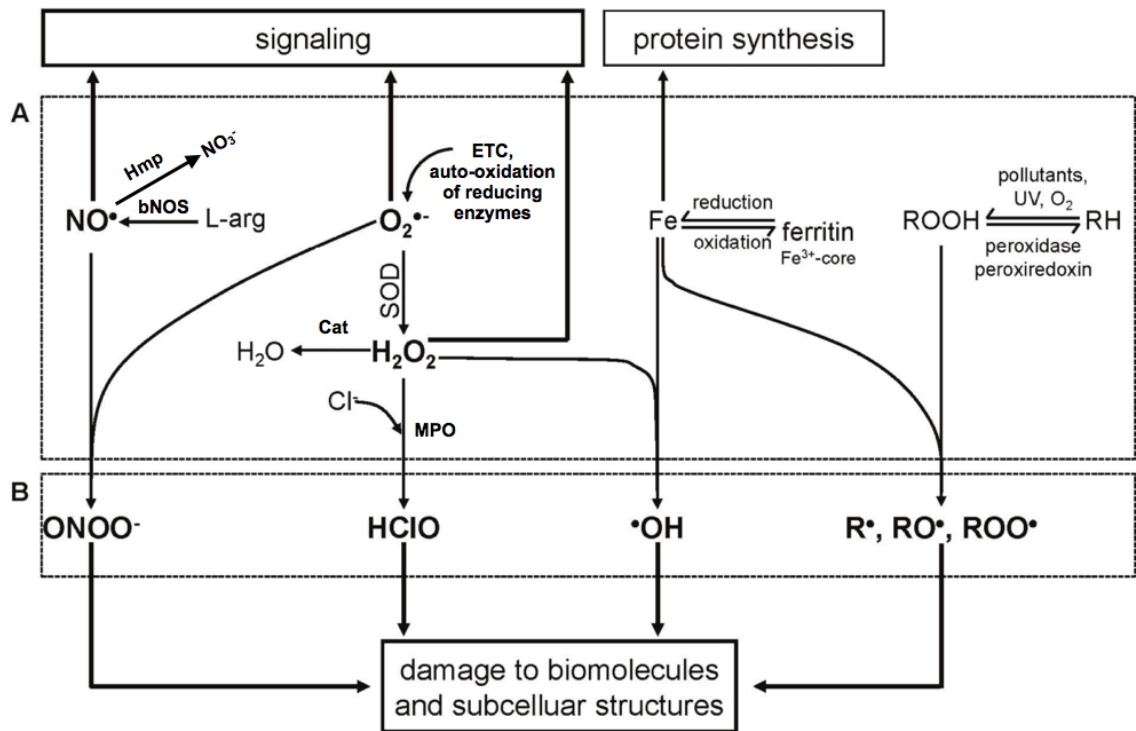
Figure 1-4 The physiological and pathological routes of reactive species, modified from Weidinger & Kozlov (2015). A) The production of primary reactive species (NO⁻, O_2^-, Fe and ROOH) the common routes of detoxification (Hmp, SOD-Catalase, -, peroxidase-peroxiredoxins) and the routes of formation by interaction with another primary species are shown here. (B) the secondary reactive species responsible for damage to biomolecules and subcellular structures. Abbreviations NO, nitric oxide; O_2•-, superoxide; Fe, iron; ROOH, lipid peroxide; H_2O_2 hydrogen peroxide; RH, non-oxidised lipid; R•, RO•, ROO•, lipid radicals; bNOS, bacterial nitric oxide synthase; L-arg, L-arginine; ONOO⁻, peroxynitrite; ETC, electron transport chain; SOD, superoxide dismutase; Cat, catalase; H_2O, water; Cl⁻, chloride ion; MPO, myeloperoxidase; HClO, hypochlorous acid; •OH, hydroxyl radical.
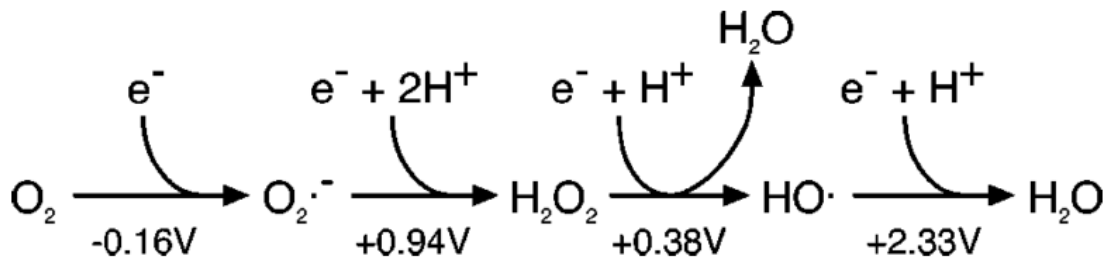
Figure 1-5 Complete electron reduction of molecular oxygen to water. Reducing potentials (V) of each reaction step are shown below along with the balanced in/output of electrons and molecules.

RNS, similarly to ROS, come in a range of forms; including nitric oxide (NO), peroxynitrite ($ONOO^-$) and S-nitrosothiols (RSNOs), illustrated in Figure 1.6. These compounds tend to be formed from NO, produced as a by-product of nitrate metabolism or NO synthases, and interact with ROS forming secondary compounds including dinitrogen trioxide (N_2O_3), $ONOO^-$ and nitrogen dioxide (NO_2^-) which ultimately lead to the S-nitrosylation of proteins. Denitrifying bacteria in the soil are the most abundant and common source of nitrosative stress to themselves and others (Tucker et al. 2008).
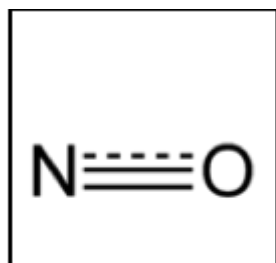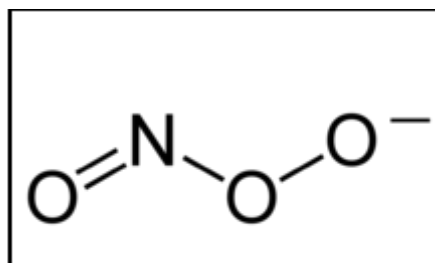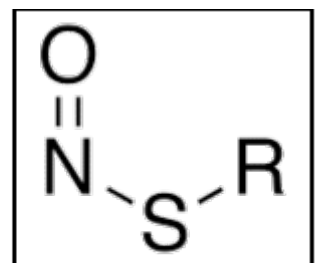
Figure 1-6 Chemical structures of NO, peroxynitrite and S-nitrosothiol.

NO is a highly diffusible, lipophilic radical (Crack et al. 2012) that at low concentrations has been shown to function as a signalling molecule, while at high concentrations it causes widespread cytotoxicity due to its reactivity with, and nitrosylation of, cellular constituents including amino acids, DNA and protein metallocofactors, particularly [Fe-S]. The majority of NO in the cell arises from nitrate reduction by the *nar* genes and bNOS (bacterial nitric oxide synthases) that produce NO for a range of biological purposes. The most well characterised NOS genes are in eukaryotes but at least one bNOS has been studied, in the case of *S. scabies txtD,* which is responsible for the nitrosylation of its

phytotoxin thaxtomin using L-tyrosine as a substrate (Barry et al. 2012; Johnson et al. 2008). Direct evidence for NO damage/mutagenesis in *E. coli* as a result of a NO build-up due to nitrate/nitrite metabolism has been provided; indicating it is the switch to aerobic growth and the introduction of oxygen which results in secondary nitrosative compounds that cause the damage (Weiss 2006).

$ONOO^-$ is a structural isomer of nitrate and can be formed from the interaction of NO/NO_2^- with $^{\cdot}O_2^-/H_2O_2$ respectively, with the former being ascribed to the most likely route of endogenous production (Pacher et al. 2007). As both an oxidant and a nitrating agent, peroxynitrite can damage an array of biomolecules. Both NO and $ONOO^-$ often interact with thiol groups to form S-nitrothiols, ultimately leading to S-nitrosylation of proteins.

RES, including quinones, aldehydes and epoxides, are compounds containing electron deficient carbon centres that are often produced during primary and secondary metabolism from the interaction of reaction intermediates with ROS and RNS.

These reactive species can cause significant damage within the cell, damaging proteins, nucleic acids, lipids and metal cofactors, including [Fe-S] cluster-containing proteins (Figure 1.4). Often the response to these stressors relies on detoxification proteins under the control of key sensory, transcriptional regulators that utilise metal cofactors or thiol chemistry to sense the reactive species directly.   The more common detoxification proteins and sensors associated with *Streptomyces* will be discussed below.

1.3.2   Common free radical and peroxide scavenging enzymes

Among the free radical and peroxide scavenging enzymes, there is a range of functional groups: the families of peroxidases, peroxiredoxins, catalases, superoxide dismutases (SOD's), hemo/flavohemoglobin deoxygenises and NO reductases (Gardner 2012). We will be discussing primarily the peroxidase/catalases, SOD enzymes and flavohemoglobin deoxygenases. Table 1.1 contains a summary of the SOD, catalase and peroxidase, and putative NO-detoxifying flavohemoglobin proteins, in which we will be focusing for this section, found in *S. coelicolor* and *S. venezuelae.*

1.3.2.1   Peroxidase and catalase enzymes

Peroxidase enzymes have been found in all aerobic organisms and typically carry out the reaction ROOR to ROH and ROH.   Many of these, including members of its own subfamily, catalase, prefer the substrate H_2O_2, while many have a higher activity for organic hydroperoxides like lipid peroxides (the result of redox toxicity on lipids). Catalase enzymes are the most common, well-studied and most reliably defined as true H_2O_2

scavengers within the cell (Mishra & Imlay 2012). Catalase converts H_2O_2 into H_2O and O_2 via an essential haem Fe moiety (Arita et al. 2006; Isobe et al. 2006; Halliwell & Gutteridge 2007). This family of enzymes has a very high specific activity for small molecules, which enter through a ~50 Å wide channel to access the haem pocket (Chelikani et al. 2005). This activity rate is possibly due to the constrained haem binding pocket (Candelaresi et al. 2013). 2D-IR spectroscopy and crystallography data has implicated the conformation of the distal histidine, coordinating the haem and a defined H_2O molecule in providing the constrained conditions that allows such rapid H_2O_2 turn over. In addition to H_2O_2, evidence from a *S. cerevisiae* catalase has implicated catalase enzymes having the capacity to process ONOO- as well (Sahoo et al. 2009).

Enzyme type

sco gene number	Sven gene number	Annotation

Catalase

sco gene number	Sven gene number	Annotation
sco0379		Catalase
sco0560	sven7337	Catalase or Peroxidase
sco0666	sven7254	Catalase
sco6204	sven6086	Catalase
sco7590	sven0140	Catalase
	sven4860	Catalase

Peroxidase

sco gene number	Sven gene number	Annotation
	sven0529	Catalase or Peroxidase
sco4444	sven0797	Glutathione peroxidase
sco2276	sven1956	Ferrous iron transport peroxidase EfeB
sco2901	sven2650	Thiol peroxidase, Bcp-type
sco3963	sven3736	Dyp-type peroxidase family protein
sco5031	sven4699	Alkylhydroperoxidase protein D
sco0465	sven6923	Non-haem chloroperoxidase
	sven7339	Thiol peroxidase, Tpx-type

Superoxide dismutase (SOD)

sco gene number	Sven gene number	Annotation
sco0999	sven2415	sodF - Fe containing SOD
sco2633	sven4944	sodN - Ni containing SOD
sco5254		sodF - Fe/Zn containing SOD

NO-detoxifying flavohemoglobin (Hmp)

sco gene number	Sven gene number	Annotation
sco7428		Flavohemoglobin (class I)
sco7094		Flavohemoglobin (class II)
sco0103	sven7234	Flavohemoglobin (class III)
	sven3696	Flavohemoglobin (class III)

Table 1-1 A summary of the catalase, peroxidase, superoxide dismutase and NO detoxifying flavohemoglobin genes from *S. coelicolor* and *S. venezuelae.*

It is common to find multiple different functional copies of these enzymes within a genome. Although the reason for so many hasn't been confirmed a variety of hypotheses have been suggested, such as allowing for different the reaction types (reductive or disproportioning), the reaction rate in regards to ROS concentration, the reliance on different metal centres and also substrate specificity allowing for finer tuning of responses. Various forms of catalase and other enzymes have been implicated in H_2O_2 scavenging with varying degrees of evidence, these include Monofunctional and bifunctional catalases, Manganese catalase, Alkyl hydroperoxide reductase, bacterioferritin comigratory protein, thiol peroxidase, glutathione peroxidase, organic hydroperoxide reductase, cytochrome c peroxidase, rubreythrin and reverse rubreythrin (Mishra & Imlay 2012).

1.3.2.2 Superoxide dismutase – SOD enzymes

Superoxide dismutase (SOD) enzymes catalyse the conversion of O_2^- to H_2O_2 and molecular O_2. This reaction is reliant on bound metal ions of Cu/Zn in eukaryotes and some bacteria, Fe/Mn and in bacteria, mitochondria and chloroplasts and Ni in some bacteria (Fridovich & Fridovich 1997; Perry et al. 2010; Miller 2004; Miller 2012). The Ni-containing SOD enzymes were first discovered in *Streptomyces* (Youn et al. 1996). Cells typically contain multiple types of SOD enzymes, as is the case with the catalase and peroxidases, with different metal centres for conditional usage. *sodF,* an Fe containing SOD in *S. coelicolor,* and *sodN*, a Ni-containing enzyme, have been shown to be differentially expressed based on Ni content in the cell by regulation of Nur (Ni uptake regulator), a Fur (Fe uptake regulator) family Ni sensor (Kim et al. 2014). The paper has shown that expression of *sodN* is controlled post transcriptionally and *sodF* by direct DNA-binding and repression by the Nur protein; this mechanism is illustrated in Figure 1.7. SodN appears to be the housekeeping SOD enzyme whereas SodF appears to be repressed until Nur senses Ni-limiting conditions. A sequence downstream of *sodF* included in the 3'UTR (untranslated region) of the *sodF* transcript matches exactly part the 5'UTR of *sodN*. In the absence of Ni or low concentrations, Nur loses DNA-binding and *sodF* mRNA is produced and in part, post transcriptionally modified by an unknown cleavage process resulting in a 3' portion of the transcript (containing the 3'UTR), containing the antisense sequence matching the *sodN* mRNA (5'UTR). These mRNA interact, which inhibits *sodN* translation and destabilises the *sodN* mRNA.
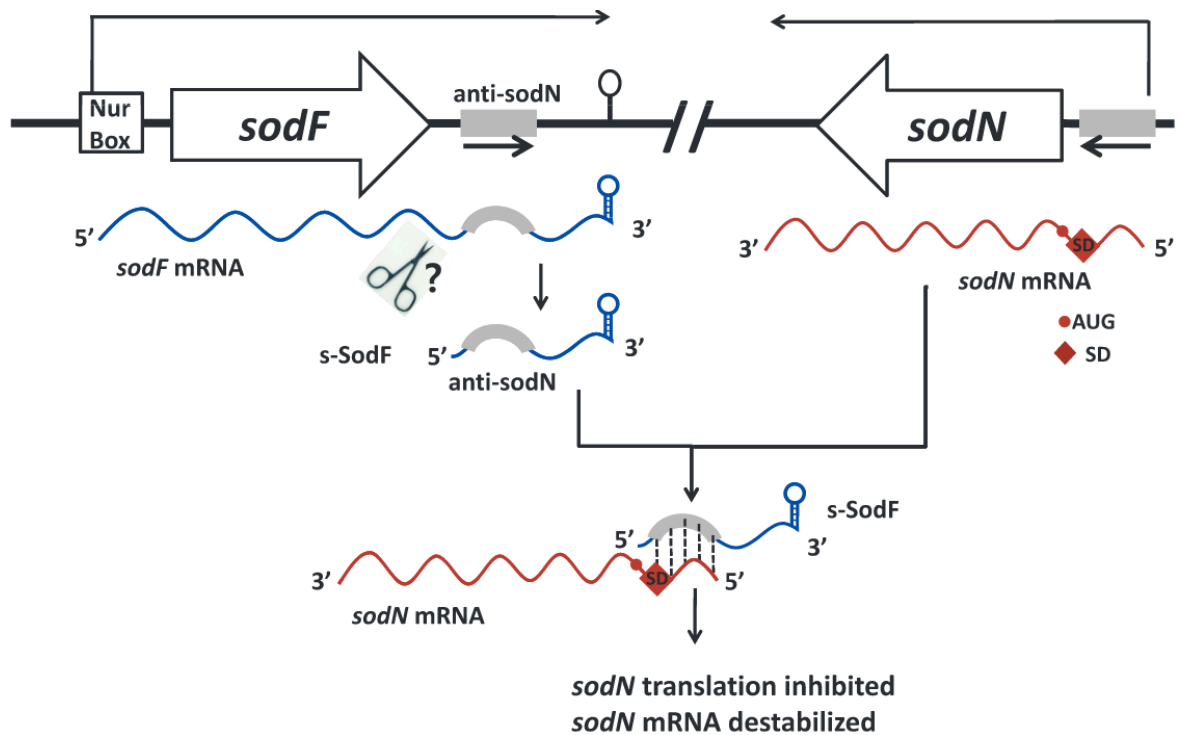
Figure 1-7 The regulation of *sodF* and *sodN* from *S. coelicolor* (Kim et al. 2014). Under nickel-limited conditions, Nur fails to repress *sodF* expression, resulting in transctiption, and processing of the mRNA. The mRNA contains 3'UTR, antisense to a sequence matching the 5'UTR of the *sodN* transcript. The processed *sodF* mRNA binds the *sodN* mRNA and results in translational inhibition and destabilization of the *sodN* mRNA.

1.3.2.3   NO-detoxifying flavohemoglobin – Hmp enzymes

The Hmp NO-detoxifying flavohemoglobin enzymes, which utilise NAD(P)H as an electron donor, have been better characterised over the last decade. This subgroup of the haemoglobin family carry out the conversion of NO to nitrate aerobically (Figure 1.8) or N_2O anaerobically. This is facilitated by the multidomain structure of the protein where the N-terminal globin domain carries out the reaction and the C-terminal reductase domain recharges the haem by oxidation for the next cycle (Forrester & Foster 2012). This reaction led to them being characterised as the major NO scavengers within the cell (Poole & Hughes 2000; Gardner et al. 2006). *E. coli hmp* has been shown to be regulated by a concerted effort of Fnr (Cruz-ramos et al. 2002), MetR (Membrillo-Hernández et al. 1998), and NsrR (Bodenmiller & Spiro 2006b) while in *B. subtilis* ResD, NsrR and Fur are the major regulators of *hmpA* (Henares et al. 2014).

A

$$NO + O_2 + Fe^{2+} \longrightarrow NO_3^- + Fe^{3+}$$

B
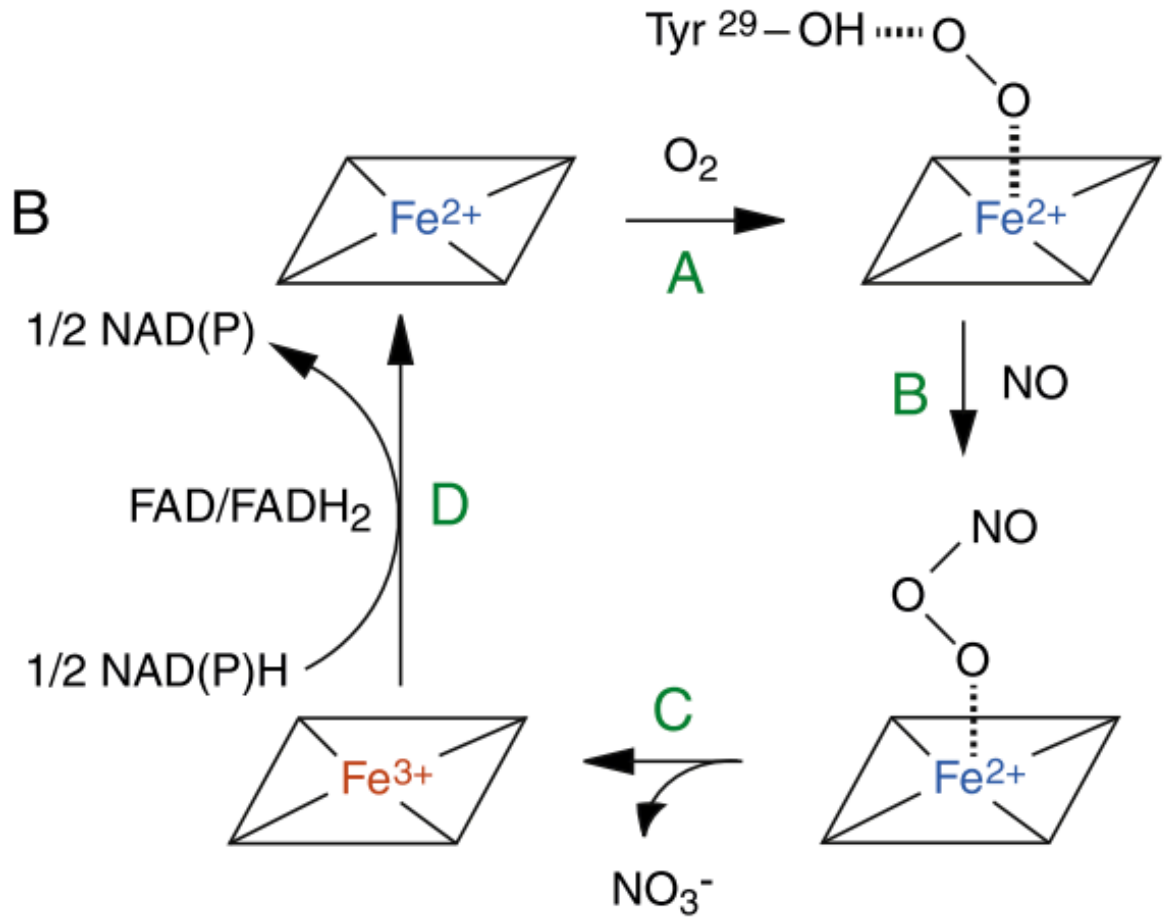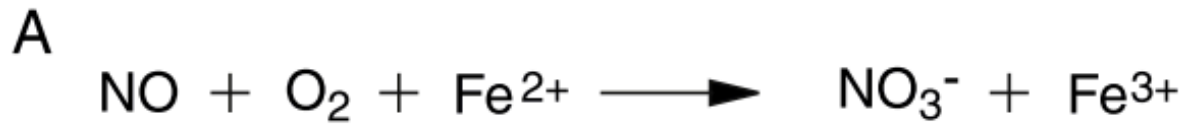
Figure 1-8 The deoxygenation reaction carried out by Hmp (Forrester & Foster 2012). (A) The deoxygenation reaction carried out by NO detoxifying flavohemoglobins. (B) A graphical representation of the chemical cyclic deoxygenation reaction, showing: step A, the addition of O_2 to the system, binding the Fe^{2+} and a conserved tyrosine residue. The inclusion of NO into the system, in step B, results in the formation of a transient peroxynitrite intermediate bond to the Fe^{2+}. Step C, the isomerisation of ONOO- and release of NO_3^- (nitrate) and the resulting oxidation to form the ferric haem. Step D, reduction of the Fe^{3+} to Fe^{2+} haem by the C-terminal reductase domain (containing FAD) driven by NADPH or NADH donated electrons with each donor able to convert 2 NO and 2 O_2 to 2 NO_3^-.

1.3.3 [Fe-S] clusters and assembly

Many of the redox sensors discussed further in this chapter utilise [Fe-S] clusters as part of their sensory mechanism, as such the following section contains a brief summary of the types of [Fe-S] clusters found in nature including important examples which facilitated our understanding of them and the major pathways responsible for their synthesis.

1.3.3.1 Types of [Fe-S] clusters

[Fe-S] clusters, the most ancient and versatile inorganic cofactors, come in various forms (Beinert 2000; Johnson et al. 2005; Meyer 2008) and were first reported in 1962 (Mortenson et al. 1962). The two most common, particularly in regards to transcription factors, are the rhombic [2Fe-2S] and cubic [4Fe-4S] forms. In addition to these there are the [3Fe-4S] contained in certain ferrodoxins, [8Fe-7S] clusters associated with nitrogenase enzymes, [1Fe-4S] clusters of certain rubridoxins, as well as the various oxidation and spin states of these cofactors (Meyer 2008). Within proteins, cysteine residues tend to be the primary ligands for cluster coordination but histidine, arginine, glutamine, serine, glutamate, aspartate or the amide backbone are also implicated in binding (Meyer 2008; Miller & Auerbuch 2015).

These clusters are not formed spontaneously from their substituents within the cell, as individually they would be toxic, instead there are dedicated biogenesis systems that utilise Fe^{2+} and L-cysteine to generate [Fe-S] clusters.

1.3.3.2 [Fe-S] cluster biogenesis

[Fe-S] cluster biogenesis is an essential process in both bacteria and eukaryotes due to the presence of many essential [Fe-S] cluster proteins. [Fe-S] cluster proteins and their biogenesis have been broadly reviewed from a variety of perspectives in recent years due to the biological processes involving [Fe-S] proteins. Some of these include respiration, central metabolism and their involvement in mammalian genetic disease (Beilschmidt & Puccio 2014), [Fe-S] cluster biogenesis (Roche et al. 2013; Boyd et al. 2015; Py & Barras 2015; Blanc et al. 2015; Wayne Outten 2014), DNA replication and repair (Fuss et al. 2015), gene regulation (Mettert & Kiley 2014) RNA modification (Kimura & Suzuki 2014) and the combinatory role of these in bacterial pathogenesis of mammalian tissue (Miller & Auerbuch 2015).

As a whole, [Fe-S] cluster biogenesis can be subdivided into three systems; NIF, ISC and SUF, each will be discussed in turn. The NIF system, unlike the ISC and SUF pathways is responsible for the assembly of clusters destined for nitrogenase enzymes. The ISC

and SUF pathways are involved in general house-keeping [Fe-S]/protein-cluster assembly but in some reports in strains, including *E. coli*, where both are present, SUF is involved primarily in stress response [Fe-S] cluster assembly (Santos et al. 2015). Although not as universally conserved as the other pathways, NIF and its original discovery (Streicher et al. 1971) facilitated the identification of ISC and ultimately SUF. Due to the relative importance of the ISC and SUF they will be the focus of the next sections. Each system contains at least 3 major domains (Figure 1.9): A cysteine disulfurase to transfer a sulfur from L-cysteine, a scaffold receptacle/platform to receive and build [Fe-S] clusters from the two substrates and a carrier to transport the cluster to the final accepting apo-target (Roche et al. 2013). In addition to the core biosynthesis machinery, we will also be looking into the key regulators involved in maintaining [Fe-S] cluster homeostasis under various conditions. Most studies of the ISC and SUF pathways were carried out either in *E. coli* or mammalian mitochondria (ISC) and as a result the majority of the hypotheses made are based on data from these organisms. The bacterial systems will be further focused on below.
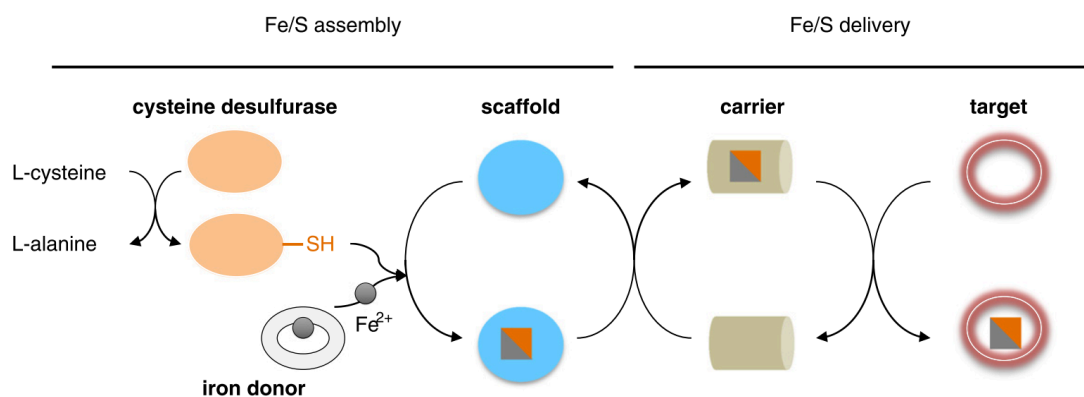
Figure 1-9 The general principles of [Fe-S] cluster biogenesis (Roche et al. 2013). Firstly, a cysteine disulfurase transfers sulfur from L-cysteine, forming L-alanine, to a scaffold protein along with Fe^{2+} from a currently unconfirmed iron donor. The scaffold protein receives and builds the [Fe-S] cluster, which is subsequently transferred to a carrier protein to finally be deposited with the terminal apo-target protein.

1.3.3.3 ISC pathway

[Fe-S] cluster assembly, carried out by the Iron Sulfur Cluster (ISC) assembly machinery is reliant on a 5-protein complex (Figure 1.10) consisting of IscS – which provides the sulfur, Fdx – which is speculated to be involved in persulfide reduction and formation of [Fe-S] cluster (Lange et al. 2000), IscU – the scaffold protein for building [Fe-S] and HscA/B – playing either a stabilising role for the scaffold/cluster and/or a role in

release/transfer of the cluster to transporters/apo-targets (Blanc et al. 2015). Additionally the major regulator of the pathway, IscR, is discussed in section 1.4.1.2.
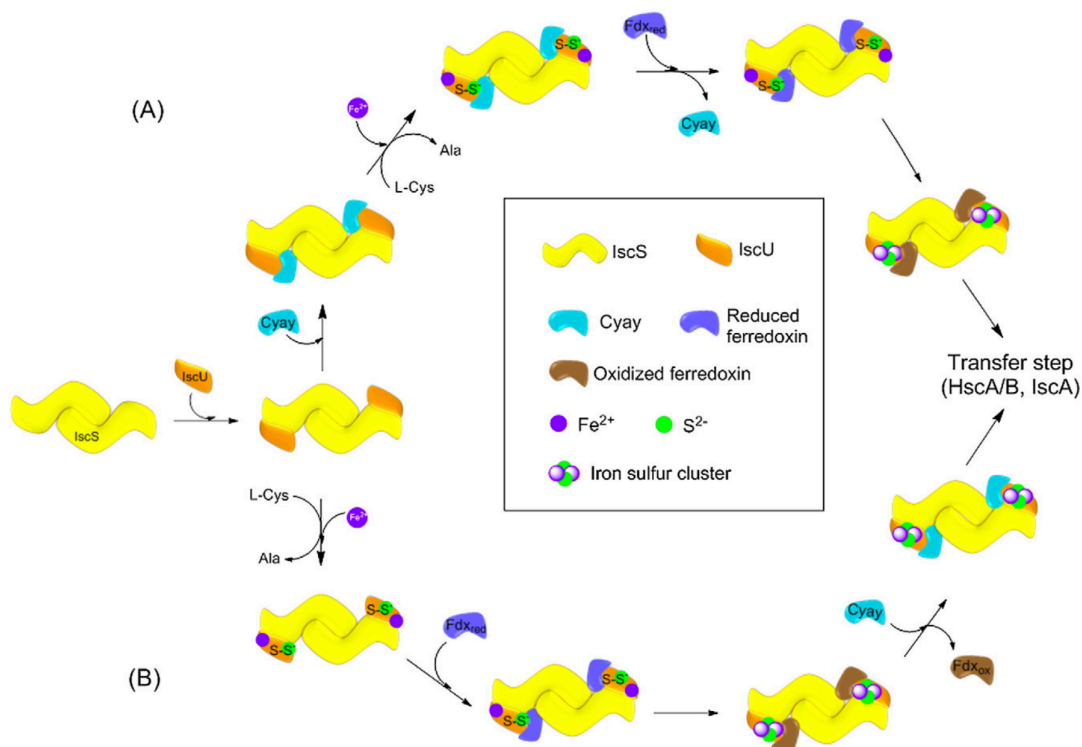
Figure 1-10 Proposed models of [Fe-S] cluster assembly via the ISC pathway (Blanc et al. 2015). (A) An IscS dimer binds two monomers of IscU and CyaY (ferrataxin) binds to the IscS-U complex. IscS produces sulfide from L-cysteine and is transferred from IscS to IscU. Fdx then displaces CyaY which carries out persulfide reduction on IscU facilitating formation of [Fe-S] cluster ready for transfer to apo-targets by HscA/B. (B) Fdx interacts with the IscS-U heterotetramer before CyaY carrying out persulfate reduction, stabilizing the complex and is dislodged later by CyaY in preparation for the transfer step.

IscS, the cysteine disulfurase enzyme (45 kDa), converts L-cysteine to L-alanine releasing a sulfur group utilised in cluster assembly. This reaction is dependant on the presence of a pyrodixal-phosphate (PLP) molecule (Prischi et al. 2010). A single PLP molecule is bound per monomer, while IscS exists as a homodimer in solution. The crystal structure of IscS differs from the analogous structures of both NifS and SufS. All three proteins contain their active site cysteine on a loop that interacts with the PLP and bound cysteine substrate . In the case of IscS this loop is large and spans a greater distance to reach the other reaction units, whereas in NifS and SufS this loop is much shorter and the PLP/Cysteine substrate resides much closer to the active cysteine residue (Cupp-Vickery et al. 2003; Shi et al. 2010; Blanc et al. 2015). This difference has been hypothesised to

be due to IscS having a broader substrate range (sulfurs are utilised for other sulferated cofacters, not only [Fe-S]) and the longer loop may facilitate interaction with differential targets (Blanc et al. 2015).

IscU, the scaffolding protein for [Fe-S] cluster assembly (Agar et al. 2000; Raulfs et al. 2008), acts as a platform to build the [Fe-S] clusters from substrate materials before transfer to the recipient protein. Apo-IscU has been shown to exist as a dimer when free in solution (Agar et al. 2000), in a disordered state, which upon binding IscS, and transfer of the persulfate group, it becomes ordered (Jin et al. 2009), suggesting the switch is involved with cluster transfer (Blanc et al. 2015). Interaction of IscS-IscU has been demonstrated and shown to facilitate transfer of persulfate to IscU (Urbina et al. 2001; Smith et al. 2001; Cupp-Vickery et al. 2003; Smith et al. 2005). Although significant steps have been made in identifying the source of Fe^{2+} (major candidates include IscX/CyaY), it is still unclear whether the candidate proteins are true Fe donors or if they regulate transfer of Fe to the complex and/or inhibit [Fe-S] formation. Due to this uncertainty it is unclear at which stage Fdx reduces the persulfate to form the cluster but its role is clear and has been shown to be essential for [Fe-S] maturation (Tokumoto & Takahashi 2001).

Various studies have been carried out to determine how these proteins interact, along with a comprehensive review of the topic by Blanc et al 2015 with two major models proposed (Figure 1.10).

Following cluster assembly the system must be able to passage the final construct to the appropriate apo-target. Two major hypotheses have built up suggesting how IscU would interact with two chaperoning molecules, HscA/B, to achieve this. The first consists of a linked cluster release/ATP hydrolysis step (Kim et al. 2012) and the second consists of linked ATP-dependant conformational changes in HscA (T/R form) and isoforms of IscU, with the IscU forms having differential cluster binding affinities (Bonomi et al. 2011). Each hypothesis is summarised in Figure 1.11. The role of HscA/B in [Fe-S] cluster maturation has been controversial, with reports of contradictory data from *in vivo* and *in vitro* work (Blanc et al. 2015). HscA has been shown to bind to a LPPVK motif of IscU (Hoff et al. 2003), to the co-chaperone, HscB which facilitates the transfer/interaction of IscU with HscA and regulates the known ATPase activity of HscA (Füzéry et al. 2008). Both models are possible based on the data available but many assumptions are made based on IscU-apo-target interactions in the absence of the chaperones. Ultimately, experiments showing multi-protein interactions and dynamics are essential for elucidating the interactions.
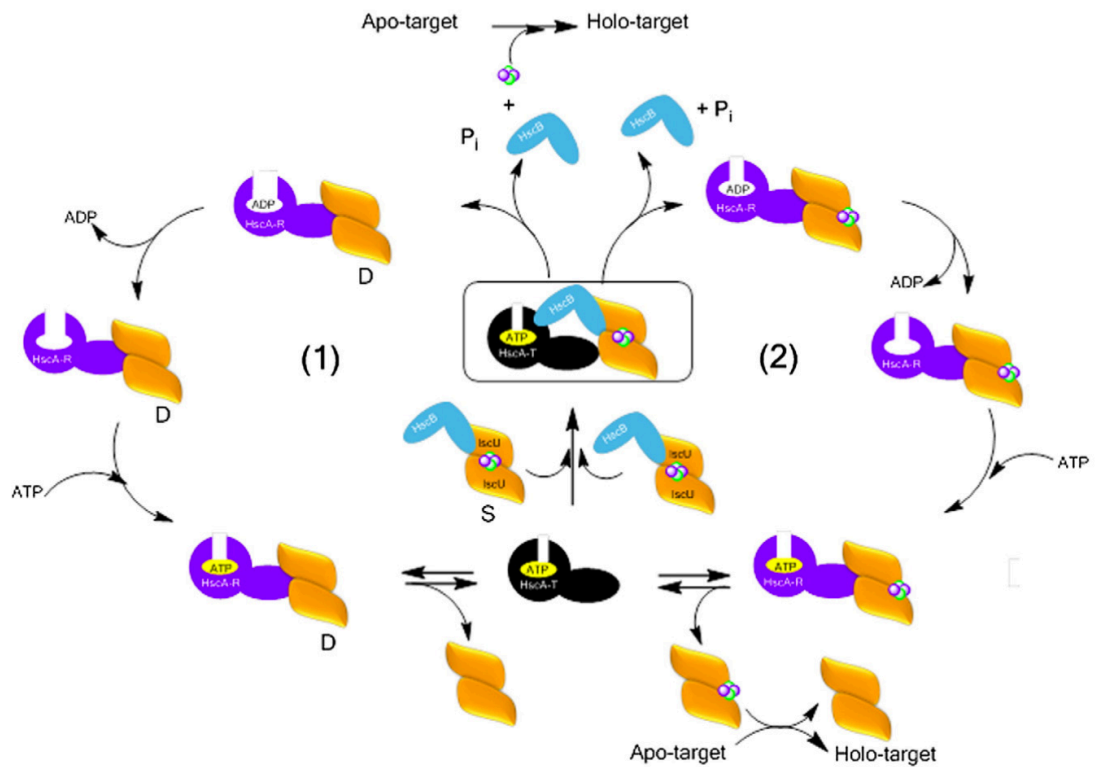
Figure 1-11 The two proposed models of [Fe-S] cluster transfer from IscU to apo-targets (Blanc et al. 2015). Model (1) shows the progression of IscU containing [Fe-S] cluster, binding HscB, the complex of HscA-B-IscU forming, the ATP dependent transfer of the [Fe-S] cluster from IscU to the apo-target protein, while dissociating from HscB and ultimately the recycling of HscA/IscU by binding ATP, recharging HscA and releasing IscU for the next round of cluster assembly. Model (2) shows the ATP dependent release of HscB and the conformational change of IscU reducing the cluster binding efficiency. Following the release of HscB/ATP hydrolysis by HscA, fresh ATP is bound and the subsequent conformational change associated with this mechanism causes the release of IscU and IscU interaction with an apo-target protein, releasing the [Fe-S] cluster to this acceptor.

The final major contributor to the function of the ISC systems is, IscR, the core ISC regulator protein. This will be discussed later in contribution to the regulation of both the ISC and SUF pathway.

1.3.3.4 SUF pathway

As with the ISC system we are still at the stage of having several hypotheses rather than a completed story, as is the nature of science. SUF follows the same major stages of [Fe-S] development as seen by ISC. Two major SUF complexes form during biogenesis:

SufSE and SufBCD. The hierarchy of the system can be seen in Figure 1.12. In addition to these complexes, SufA plays a major role in [Fe-S] cluster transport to apo-targets.
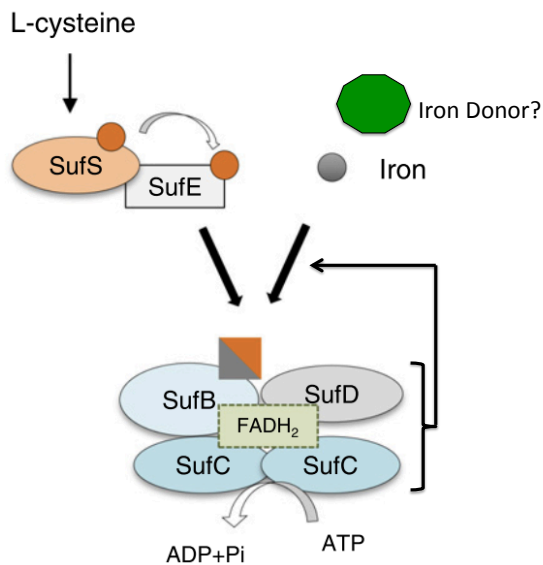
Figure 1-12 The SUF pathway. Modified from Roche et al. (2013). L-cysteine is metabolised to L-alanine by SufS releasing a sulfide that is transferred from conserved cysteines on SufS to SufE and then to SufB. SufBC$_2$D acquires Fe from an unknown Fe donor in a SufC ATPase and SufD dependant mechanism.

SufR, the major SUF regulator, primarily studied in cyanobacteria, is a dimeric [4Fe-4S]-containing protein that primarily represses the SUF pathway (Wang et al. 2004; Shen et al. 2007). Similar to NsrR and FNR it binds its cluster using 3 conserved cysteines, however, the 4th ligand is currently unknown but is not the 4th present cysteine (Shen et al. 2007). Both the apo and holo [4Fe-4S] forms of SufR can bind DNA; the cluster containing protein binds much tighter and is suggested to be the major regulatory form. This suggestion implies that in the absence of sufficient [4Fe-4S] SufR derepresses the SUF pathway, facilitating the up regulation of cluster synthesis.

SufS is the cysteine disulfurase as previously described. SufE, the sulfur transfer shuttle, associates with SufS, stimulating its activity and accepts the resulting sulfide via a conserved cysteine residue (Dai & Outten 2012). In addition, SufSE interacts with SufBC$_2$D (via SufE-SufB interaction) promoting cluster assembly rate/passage the sulfur to SufB (Dai & Outten 2012; Selbach et al. 2013; Boyd et al. 2015; Layer et al. 2007).

From an evolutionary perspective, SufBC forms the core of the Suf system (Boyd et al. 2015). SufB is thought act as the molecular scaffold for [Fe-S] assembly (Boyd et al. 2015), similar to IscU, primarily due to its ability to produce and pass [Fe-S] clusters, from minerals provided, into Fdx (Chahal & Outten 2012). SufC, always encoded along side

SufB, has been shown to have ATPase activity linked and enhanced by the presence of either SufB and/or SufD (Eccleston et al. 2006; Petrovic et al. 2008). SufD, a paralogues of SufB acquired through a duplication event, has been shown to be essential for Fe acquisition, alongside the SufC ATPase activity, and passage of Fe to SufB (Saini et al. 2010). The resulting function of SufC to date has not concretely been demonstrated and is unlikely to play a similar function to that of the ISC system ATPase, HscA. Addition of ATP, depending on cluster state and subunit coordination, has been shown to inhibit [Fe-S] transfer (Chahal & Outten 2012). Various combinations of the SufBCD complex have been isolated over the years. These different forms include a $SufBC_2D$ and $SufB_2C_2$. It has been suggest that the $SufBC_2D$ form is an early stage/cluster assembly complex utilizing SufCD to incorporate Fe into the SufB cluster assembly, whereas the $SufB_2C_2$ complex is involved in the stepwise reduction of two 2Fe-2S clusters to form a 4Fe-4S cluster (Wayne Outten 2014). However, there currently is no substantial evidence for the biological significance of a $SufB_2C_2$.complex.

Iron acquisition/donation to these systems has been speculative over the years but recently disruption of IscA/SufA or depletion of Fe resulted in accumulation of red SufS (sulfide bound) form (Yang et al. 2015). It has been suggested that Fe/S transfer to SufB/IscU is a connected process and that limitation of Fe (either by removal of Fe chaperones/acquisition proteins or Fe chelator) results in an accumulation of sulfide on IscS/SufS and ultimately a stop to [Fe-S] cluster assembly. The history of IscA/SufA has been filled with reports of contradictory evidence either acting as a Fe mononuclear binding protein or an [Fe-S] transporter and debates weather data is an artefacts of *in vitro* studies (Blanc et al. 2015).

1.3.3.5 [Fe-S] cluster assembly in *Streptomyces*

The ISC system is conserved primarily in Gram-negative bacteria but has some conservation within Gram-positives, while SUF tends to be more widely conserved. From an evolutionary perspective, SUF is thought to play a more important role. Case in point, *Streptomyces* species only have the SUF pathway (*S. coelicolor* genes: *sco1926-1919*, and *S. venezualae* genes: *sven1557-50*) as summarised in Table 1.2.

sco	Sven	Gene name	Annotation
sco1919	*sven1550*	PaaD	Contains domain of unknown function
sco1920	*sven1551*	SufE2	Sulfur transfer shuttle
sco1921	*sven1552*	SufS	Cysteine disulfurase
sco1922	*sven1553*	SufC	ATPase activity, facilitate cluster transfer
sco1923	*sven1554*	Fdx	Cluster transfer
sco1924	*sven1555*	SufD	Molecular scaffold
sco1925	*sven1556*	SufB	Molecular scaffold
sco1926	*sven1557*	SufR	SUF regulator

Table 1-2 A list of the SUF Pathway genes in *S. coelicolor* and *S. venezuelae.*

1.3.4   Redox sensors: *Streptomyces* and beyond

A list of the key redox sensors discussed below, their presence in *Streptomyces*, both containing [Fe-S] or other metal centres and those with active thiol groups and the role they play in redox stress can be found in Table 1.3. Due to the reactivity of the redox chemicals it can be difficult to define a sensor from an interaction partner. Generally, the below the information is provided for direct sensors of the redox chemicals. Although forming an incomplete list, the examples below illustrate the major sensor types with potentially applicable characteristics to our studied regulators.

1.3.4.1   Non [Fe-S] containing redox sensors

1.3.4.1.1   OxyR – Peroxide sensor

Originally discovered in *Salmonella typhimurium*, OxyR is a member of the LysR family of transcriptional regulators and regulates a large peroxide-inducible regulon (Antelmann & Helmann 2011). OxyR subunits form a tetramer, which upon oxidation of a conserved cysteine (C199), forms a disulfide bond to another conserved cysteine (C208), activating it and, subsequently, its regulon (Lee et al. 2004). The OxyR regulon in *E. coli* and *Salmonella enterica* includes among other targets, catalase, which converts H_2O_2 to water, glutaredoxins and glutathione reductases for homeostasis of the low molecular thiol pool, Fur, for Fe homeostasis and control of deleterious Fenton chemistry, regulation of the SUF pathway and alkylhydroperoxide reductase (Imlay 2008; Antelmann & Helmann

2011; Santos et al. 2015).   OxyR in *S. coelicolor* has been shown to induce the alkylhydroperoxide reductase system but not the catalase genes under H_2O_2 stress conditions, suggesting a more specialised role (Hahn et al. 2002).

Sco genes	Sven genes	Protein	Senses	Method of sensing
sco5033	sven4701	OxyR	Peroxide	Redox sensitive cysteine residues, intramolecular disulfide bond formation, loses DNA-binding
sco5216	sven4870	SigR	Thiol pool	Redox sensitive cysteine residues in anti-sigma factor RsrA, disulfide bond formation, release of Zn(II) ion, release of sigma factor
sco2987	sven2241	OhrR	Organic hydroperoxide	Redox sensitive cysteine residues, forms cyclic sulfenamide, losses DNA-binding
sco3320	sven3182	Rex	Redox poise	NAD(H) binding
See wbl section	See wbl section	Wbl	O_2/NO??	Unknown/[4Fe-4S]
sco7427	sven6563?	NsrR	NO	Loss of the [4Fe-4S]
sco1697	sven1301	SoxR	ROS/RNS stress	Oxidation/nitrosylation of the [2Fe-2S]
sco0561	sven4114	Fur/IdeR/DtxR	Fe concentration	Regulates increases/decrease of Fe uptake based on Fe content by direct Fe binding.

Table 1-3 A summary of the key *Streptomyces* redox sensors discussed in this chapter.

1.3.4.1.2 Rex – NAD(H)/NAD$^+$ redox poise sensor

First described in *S. coelicolor*, the Rex regulator senses redox poise within the cell by binding to NAD+, which modulates Rex DNA-binding (Brekasis & Paget 2003). Structural studies of the Rex dimer bound to NADH and NAD$^+$ have been carried out in *Thermus aquaticus* (McLaughlin et al. 2010; Sickmier et al. 2005), *T. thermophiles* (Nakamura et al. 2007), *B. subtilis* (Wang et al. 2011; Pagels et al. 2010) and *Streptococcus aglactiae* (3KET Protein DATA bank (PDB) entry). Rex forms a dimer, with each monomer containing a N-terminal DNA-binding winged helix-turn-helix domain and a C-terminal Rossman fold that is involved in dinucleotide binding. The Rex regulon in *S. coelicolor* includes the *cydABCD,* the *cytochrome bd* terminal oxidase operon, *hemACD,* the haem biosynthesis genes and *nuoA-N,* the membrane bound proton-translocating NADH dehydrogenase operon (Brekasis & Paget 2003). Rex detects redox poise, as the pool of reducing resources (NADH) is depleted, NADH inhibition of Rex DNA-binding is relieved allowing the activation of target genes (Brekasis & Paget 2003). A highly conserved Rex binding sequence, generalised to TTGTGAANNNNTTCACAA, has been reported across the Rex containing species, initially investigated in *B. subtilis* (McLaughlin et al. 2010) and *S. aureus* (Pagels et al. 2010), then further investigated across 119 strains bioinformatically that was then validated experimentally in the *Thermotogales* (Ravcheev et al. 2012).

1.3.4.2 [Fe-S] containing sensors

1.3.4.2.1 Wbl – a range of functions, including O$_2$/NO sensing

The WhiB-like (Wbl) proteins, named because of similarity to the *S. coelicolor* WhiB protein are a family of [4Fe-4S] containing proteins with a range of biological roles including morphogenesis, cell division, virulence, metabolism and antibiotic resistance (Ventura et al. 2007; Keith F Chater & Chandra 2006). Found only in Actinobacteria, they contain a highly conserved group of cysteines (C-X$_n$-C-X$_2$-C-X$_5$-C) essential for ligation of a [Fe-S] cluster. The best-characterised Wbl protein, WhiB3 from *M. tuberculosis,* which is essential for virulence (Saini et al. 2012), is a transcriptional regulator, binding DNA in its apo form. Holo-WhiB3, lacking DNA-binding capacity, when exposed to high oxygen or NO undergoes a stepwise reaction: [4Fe-4S], [4Fe-4S]$^{1+}$, [4Fe-4S]$^{2+}$, [3Fe-4S]$^{1+}$ and eventually loses its cluster (Singh et al. 2007; Singh et al. 2009). Work on WhiB1 and *S. coelicolor* WhiD has shown that the [4Fe-4S] cluster interacts with 8 NO molecules incredibly quickly in a

multiphasic reaction (Crack et al. 2011). Upon loss of cluster and acquisition of DNA-binding capacity, apo-WhiB3 now has four exposed cysteine thiol residues which, if additionally exposed to thiol oxidants, forms 2 pairs of disulfide bonds increasing the strength of the protein/DNA interaction (Singh et al. 2009). Mycobacterial WhiB1-7 stimuli include detergent exposure, acid, heat, nutrient concentration, ethanol, oxygen, NO and iron (Salerno et al. 2009; Geiman et al. 2006; Morris et al. 2005; Burian et al. 2012; Chandra & Chater 2014). The *S. coelicolor* chromosome encodes 11 Wbl proteins, five of which are highly conserved in actinobacteria: WhiB and WhiD, involved in sporulation, WblC, important in multi-drug resistance and WblA, involved in developmental transitions and the little studied WblE which is essential (Fowler-Goldsworthy et al. 2011). A list of the *S. coelicolor*, *S. venezuelae* and *M. tuberculosis* Wbl proteins can be found in Table 1.4. *wblA* in *C. glutamicum, whcA,* negatively influences the oxidative stress response (Choi et al. 2009). The Wbl proteins have been suggested to interact with the major antioxidant thiol, MSH (mycothiol), in actinobacteria (Soliveri et al. 2000). Some Wbl proteins have been found in non-actinobacteria, which suggests this is not strictly the case (Chandra & Chater 2014). *wblC* mutants in *M. tuberculosis* and *S. lividans* (Morris et al. 2005) suffer the same sensitivity to antibiotics as mycothiol deficient mutants (Rawat et al. 2002; Dosanjh et al. 2008; Liu et al. 2013) and MSH was shown to control MSSM (Morris et al. 2005; Burian et al. 2012). Some Wbl proteins have been shown to have specific DNA-binding activity (Rybniker et al. 2010; Smith et al. 2010; Stapleton et al. 2012) and this activity is dependent on the presence/absence of their [4Fe-4S] clusters. The interaction of these with NO has been shown to enhance the interaction with DNA (Singh et al. 2007; Smith et al. 2010; Crack et al. 2011; Crack et al. 2013; Stapleton et al. 2012). Phylogenetic investigations of the *wbl* genes have implicated them, based on co-conservation, with two NO-related genes, corresponding to *sco0741* and *sco4179* (Chandra & Chater 2014). *sco0741,* a *M. smegmatis* ortholog, has been shown to rapidly convert MSNO (a mycothiol, NO conjugate) into MSH sulphonamide *in vitro,* which *M smegmatis* converts to MSSM (oxidised MSH) and nitrate *in vivo* (Vogt et al. 2003). The second gene, *sco4179,* has significant similarity to the nitrobindins; haem containing proteins that bind NO under anoxic conditions, and a potential role has been suggested for the Wbl/NO/nitrobindin interaction, summarised in Figure 1.13 from an evolutionary perspective (Chandra & Chater 2014). Chandra and Chater, 2014, suggest that the interaction is based on the Wbl binding NO (which affects down stream gene expression), while the nitrobindins denitrosylate the Wbl protein by passaging the NO

onto MSH, forming MSNO, which is reduced to MSH *via* a MSNO reductase, ultimately regenerating the MSH and Wbl.

Gene	*S. coelicolor*	*S. venezualae*	*M. tuberculosis*
wblA	*sco3579*	*sven3349*	*whiB4*
whiB	*sco3034*	*sven2776*	*whiB2*
wblC	*sco5190*	*sven4842*	*whiB7*
whiD	*sco4767*	*sven4452,*	*whiB3*
wblE	*sco5240*	*sven4905*	*whiB1*
wblH	*sco6715*	*sven6397*	*whiB2*
wblI	*sco5046*	*sven4715*	-
wblJ	*sco7106*	-	*whiB1*
wblK	*sco7306*	-	*whiB1*
wblL	*sco6965a*	-	*whiB1*
wblM	*sco6922*	*sven5661*	-
wblN	*scp1.95*	-	*whiB1*
wblO	*scp1.115*	-	-
wblP	*scp1.161*	-	-

Table 1-4 The *wbl* genes identified in *S. coelicolor, S. venezualae* and *M. tuberculosis.*
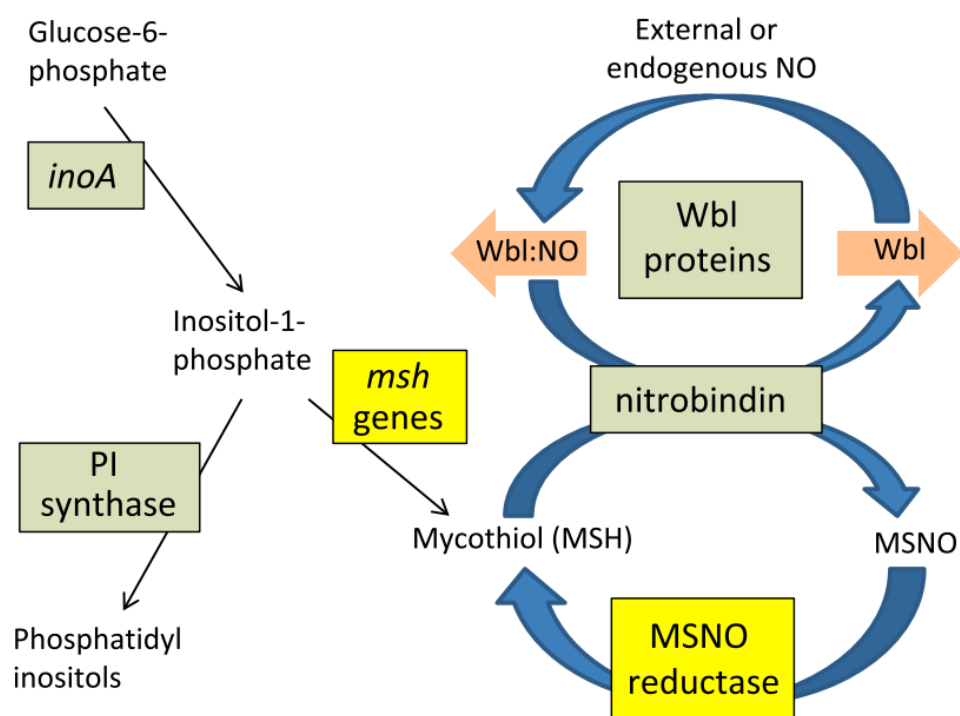
Figure 1-13 A hypothetical scheme indicating the physiological inclusion of NO to the cell via the Wbl proteins (Chandra & Chater 2014). It is suggested that the green boxes contain proteins/genes associated with early actinobacteria and the yellow contain those acquired later in evolution based on improved functionality. Initially it is thought they made phosphoinositol-containing phospholipids and Wbl proteins to respond to NO, the pink arrows corresponding to down stream gene regulation depending on the state of the Wbl protein. The predicted nitrobindin protein is thought to denitrosylate the Wbl:NO utilising MSH forming Wbl and MSNO. MSNO is thought to be reduced by MSNO reductase forming MSH and nitrate improving the regeneration of both MSH and as a result Wbl proteins.

1.3.4.2.2 FNR – Oxygen/NO sensor

The Fumerate and Nitrate Reduction (FNR) regulator, in its holo-form is best characterised in *E. coli,* is a $[4Fe-4S]^{2+}$ containing dimeric global transcriptional regulator controlling gene expression in response to oxygen (Lambden & Guest 1976). Produced constitutively under aerobic conditions in its $[4Fe-4S]^{2+}$ form, the FNR cluster is oxidised to a $[2Fe-2S]^{2+}$ (Khoroshilova et al. 1997), via a $[3Fe-4S]^{1+}$ intermediate (Crack et al. 2007), which is eventually lost from extended oxygen exposure (Green et al. 1996). This results in the apo-FNR, the inactive form,

accumulating within the cell. Under conditions where oxygen is depleted, FNR is reconstituted with a [4Fe-4S] by the ISC pathway, it then dimerises and activates its regulon (Kiley & Beinert 1998; Constantinidou et al. 2006). The FNR regulon is involved in anaerobic survival and respiration, including regulating the expression of the nitrate and nitrite reductases, potent sources of endogenous NO. FNR has also been shown to sense NO as well as oxygen (Pullan et al. 2007) and, like the Wbl proteins, has been shown bind 8 molecules of NO (Crack et al. 2013). FNR is characterised as a virulence factor because of its essentiality in many intracellular human pathogens. Many pathogens, including *Neisseria meningitides* (Bartolini et al. 2006)*, Shigella flexneri* (Marteyn et al. 2012; Schroeder & Hilbi 2008)*, Salmonella enterica* (Rollenhagen & Bumann 2006; Fink et al. 2007) and Uropathogenic *E. coli* (Barbieri et al. 2014; Ronald 2003)*,* undergo the switch of aerobic to/from anaerobic growth during a cycle of infection and rely on FNR for the production of other virulence factors (including type three secretion systems) and the switch to growth using alternative electron acceptors e.g. denitrificaiton (Miller & Auerbuch 2015). *Streptomyces* are obligate aerobic organisms but many have been shown, with key focus on our model organisms, to be able to survive extend periods under anaerobic conditions with little to no deleterious effects (van Keulen et al. 2007). There is no currently reported oxygen sensor for either of our model organisms, but CRP/FNR family members are present however, these lack the key cysteine residues for cluster coordination.

1.3.4.2.3 NsrR – NO sensor

NsrR has been included as a [Fe-S] cluster protein but will be discussed below in section 1.4.1.1.

1.3.4.2.4 SoxR – Oxidative/nitrosative stress sensor

SoxR is a dimeric transcription factor conserved within proteobacteria and actinobacteria containing one redox active [2Fe-2S] per monomer, activated by oxidation or nitrosylation of the cluster (Lee et al. 2015). The target regulon has been shown to differ depending on the strain analysed, with those known in *S. coelicolor* being an NADPH-dependant. These include flavin reductase, quinone reductase, an ABC transporter, a monoxygenase and a hypothetical protein, which are each oriented to deal with extracellular redox-active molecules including actinorhodin (Shin et al. 2011) and a sixth being a putative oxidoreductase (Naseer et al. 2014). Recent work has shown that strain specific SoxR responds at differing

sensitivities to ROS (Lee et al. 2015). *S. coelicolor* (ScoSoxR) and *E. coli* (EcoSoxR) proteins respond differentially to ROS stresses and mutational studies can interchange their relative sensitivities. The difference in sensitivity arises from key residues (ScoSoxR: R127/P131, EcoSoxR: L126/V130) surrounding the [2Fe-2S] binding site. Mutations converting these residues within the ScoSoxR (R127L/P131V) or EcoSoxR (L126R/V130P) confer the opposing proteins standard sensitivity e.g., EcoSoxR acquires the ScoSoxR sensitivity and *vice versa*.

In addition to the already mentioned regulators, and including NsrR, I will now discuss the Rrf2 family of proteins, which will be the focus of this thesis.

1.4   Rrf2 proteins: action, detection and production.

1.4.1   Rrf2 Family

The Rrf2 super-family of transcriptional regulators (PF02082, Pfam database) encompasses a variety of functionally diverse regulators, found widely throughout the bacterial kingdom, with key members of the family being involved in nitric oxide-sensing/detoxification (NsrR) (Crack et al. 2015), [Fe-S] cluster biogenesis (IscR) (Santos et al. 2015), cysteine metabolism (CymR) (Shepard et al. 2011), iron uptake (RirA) (Hibbing & Fuqua 2011) and photosynthesis (Slr0846/MppG)(Midorikawa et al. 2012; Midorikawa et al. 2009; Imam et al. 2014). The first reported *rrf2* gene was identified in *Desulfovibrio vulgaris* subsp. *Vulgaris* Hildenborough as a regulator of the cytochrome redox complex encoded by the *hmc* operon (Keon et al. 1997). The main body of the work within this thesis is investigating specific members of the Rrf2 family in *Streptomyces* species.

Although difficult to characterise due to low sequence homology, analysis of the Rrf2 family highlighted key characteristics including an N-terminus winged helix-turn-helix DNA-binding domain, the proteins have a molecular weight between 12-18 kDa (Tucker et al. 2010), a dimerization helix and anywhere from 0-5 cysteines. For cysteine-containing Rrf2 proteins, three C-terminal cys residues are more highly conserved than the others and are important ligands for iron sulfur cluster [Fe-S] binding or in thiol chemistry. The use of [Fe-S] clusters to sense the environment, in particular sensing the type of atmosphere, specific nutrient availability and redox (ROS/RNS/ROS) conditions, is becoming an increasingly observed theme.

To provide a detailed explanation of this family of proteins, I will first consider each of the characterised Rrf2 proteins listed above, the steps involved in their biosynthesis

i.e. the different types of [Fe-S] clusters, [Fe-S] cluster biogenesis and the different systems involved therein.

1.4.1.1 NsrR

Originally described in *Nitrosomonas euopaea* as a novel nitrite-sensing transcriptional repressor (Beaumont et al., 2004), NsrR is an NO sensing, dimeric transcriptional regulator containing a [4Fe-4S] cluster as shown for purified *S. coelicolor* and *Bacillus subtilis* (Crack et al. 2015; Yukl et al. 2008; Vincent M. Isabella et al. 2009). The cluster is coordinated by 3 cysteine residues, as described in the *N. gonorrhoeae* work and the 4th ligand, as has been shown for *S. coelicolor* NsrR, is likely a glutamic acid (Crack et al. 2015). In *E. coli,* NsrR has been shown to sense NO and regulate at least 60 genes (Partridge et al. 2009, Vine 2011), including *hmp* (Stevanin et al. 2007). Unlike NorR, NsrR appears to control a global response to NO, rather than control a specific detoxification response (D'Autréaux et al. 2005). Various different studies in different backgrounds (*E. coli, Salmonella enterica* and *B. subtilis* primarily) are elucidating ever expanding NsrR regulons by investigating its role in the *S. enterica* stress response to acidified nitrite and nitrosative stress (Mühlig et al. 2014; Karlinsey et al. 2012; McLean et al. 2010), expression of the locus of enterocyte effacement (LEE) in Enterohemorrhagic *E. coli* (EHEC) (Branchu et al. 2014), *B. subtilis* global transcriptional regulation of class I and II genes (Kommineni et al. 2012), coregulation of ResD, Fur and NsrR target genes (Henares et al. 2014), and the regulation of anaerobic metabolism in *B. subtilis* (Härtig & Jahn 2012). Some of the most highly positively regulated ResD targets are *nasDEF* and *hmp,* each are NsrR repressed targets while ResD acts as an antirepressor of Fur for *ykuN* and a corepressor with NsrR. The specificity of NsrR as an NO sensor, rather than additionally acting as an Fe sensor (through its [4Fe-4S]), was investigated by comparison of the IscR and NsrR [Fe-S] maturation process showing that, unlike IscR, NsrR can be maturated by both the ISC and SUF pathways (see section 1.4.2) (Vinella et al. 2013). Additionally, as mentioned, *B. subtilis* NsrR has both class I (NO sensitive) and class II (NO insensitive) targets (Kommineni et al. 2012). In all *B. subtilis* NsrR repressed targets tested, the release of NsrR repression was not enough to activate expression because activation by ResD was also required (Kommineni et al. 2012). Holo-NsrR functionally binds the promoters of class I targets, hence the cluster is the main target for NO, while class II targets are insensitive as they are bound more weakly by apo-NsrR instead. NsrR in *S. coelicolor* has been shown by our collaborative group, and as part of this work, to

regulate only three targets, *nsrR* and two *hmp* genes (Crack et al. 2015) more similar to NorR in *E. coli*. The information we know about the *S. coelicolor* regulon, and the protein itself, is illustrated in Figure 1.14. Recent work in *E. coli* his indicated that although *nsrR* has a high expression rate, is not efficiently translated, resulting in a highly transcribed gene but a low abundance of protein (Chhabra & Spiro 2015).
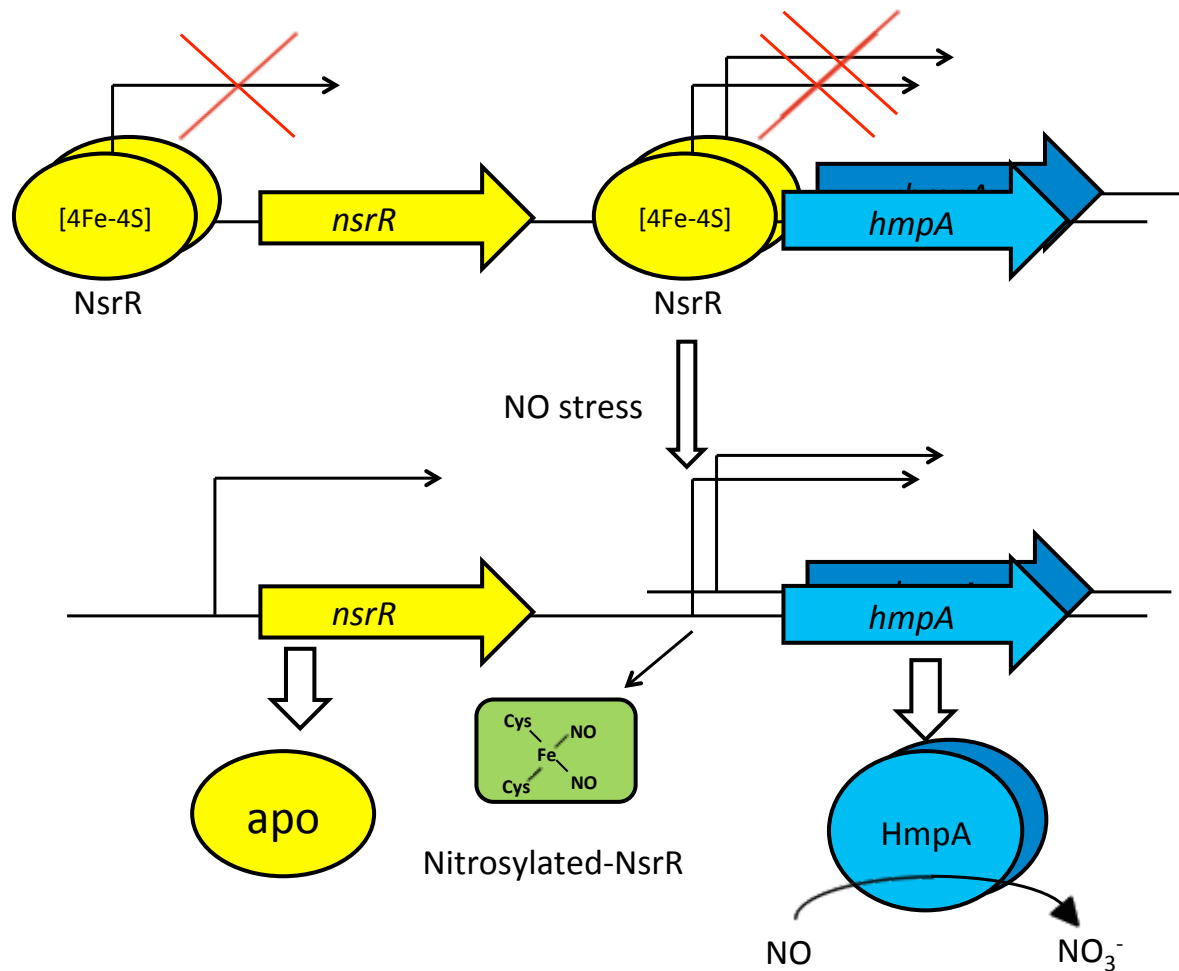
Figure 1-14   A schematic representation of NsrR, its targets and regulation mechanism, modified from Tucker et al. (2010) and updated with information from Crack et al. (2015). NsrR is produced and binds upstream of 3 target genes, *nsrR, hmpA1 and hmpA2*. Nitrosylation of the cluster putatively by NO gas releases DNA-binding and allows the expression of the NO detoxifying *hmp* genes.

1.4.1.2 IscR

IscR, Iron sulfur cluster Regulator, is the first gene encoded on the *icsRSUA-hscAB-fdx* operon and is one of the major regulators of both the ISC and SUF system (Schwartz et al. 2001). When isolated anaerobically, IscR contains a 2Fe-2S cluster,

which facilitates repression of the *isc* operon (Schwartz et al. 2001). This repression is dependent on a functional ISC system (based on work with *iscS* and *hscA* disruptions) for cluster incorporation (Schwartz et al. 2001; Nesbit et al. 2009). Mutational studies of IscR have highlighted that the 3 conserved cysteines (C92, C98, C104) and histidine (H107), ligate the 2Fe-2S cluster necessary for repression (Yeo et al. 2006; Nesbit et al. 2009). Interestingly, apo/holo-IscR regulates a different subset of genes, modulating expression and production of [Fe-S] clusters under either stress or non-stress conditions respectively (Nesbit et al. 2009; Giel et al. 2013; Santos et al. 2015).

1.4.1.3 CymR

CymR, Cysteine metabolism Regulator, unlike many other Rrf2 proteins studied to date, has a different, and ultimately more complex regulatory mechanism (Ji et al. 2012). CymR, In *B. subtilis* and *S. aureus*, is a global regulator of cysteine metabolism and, under Cysteine rich conditions, in coordination with CysK, represses a large group of genes involved in cysteine uptake and synthesis from a range of sulphur sources (Even et al. 2006; Tanous et al. 2008; Soutourina et al. 2009). CysK, an O-acetylserine lyase, binds to CymR, facilitating DNA-binding. Under conditions of cysteine starvation, there is a build up of O-acetylserine, CysK senses this and disassociates from CymR, with a loss of CymR DNA-binding, depressing the target regulon. More recently, it has been reported that upon induction of oxidative stress and subsequent thiol stress CymR loses DNA-binding capacity by the oxidation then thiolation of its sole cysteine residue, leading to derepresses its target regulon (Ji et al. 2012). These two mechanisms play an important role in sensing the different sides of possible cysteine depletion within the cell. The first mechanism, sensing direct cysteine depletion through O-acetylserine levels, which are restricted to low concentrations while cysteine is present, and the second, sensing oxidative and thiol stress which can cause a depletion of available cysteine within the cell. Disruption of the *cymR* gene leads to increased intracellular cysteine concentrations and hydrogen sulphide formation (Hullo et al. 2010; Soutourina et al. 2010).

1.4.1.4 RirA

RirA, Rhizobial iron regulator, is a transcriptional regulator that, in concert with Irr, are the major iron sensors in the alpha proteobacteria (Johnston et al. 2007) replacing the widely considered core Fe sensor in proteobacteria, Fur, which tends to

be a manganese sensor in this group of bacteria. RirA has been proposed to sense [Fe-S] concentration, and Irr (a Fur family regulator), haem, rather than the environmental Fe concentration directly (Johnston et al. 2007). With 3 conserved cysteine residues, RirA is thought to coordinate an [Fe-S] cluster but this has not been experimentally confirmed (Crack et al. 2012).

1.5   Thesis outline for chapter 3 and 4

The aims of this work were to document and the further investigation of the *S. coelicolor nsrR* gene (Chapter 3) building on work carried out by Tucker et al. (2008 and 2010) and the newly studied *rrf2* gene in *S. venezualae, rsrR* (Chapter 4)*, identified through work based on *nsrR.*

We aimed to investigate the regulon of *nsrR* and *rsrR* to determine their function within the cell. To do so, ChIP-Seq and ChIP and dRNA-seq respectively, were used to investigate the regulons. *In vitro* biochemistry, using purified protein, was carried out to validate the data. It was hypothesised that both regulators would be involved in nitric oxide sensing/detoxification and each would have 10s of regulated target genes.

2 Materials and Methods

2.1 Strains culture conditions and storage

2.1.1 Strains, plasmids, primers, media and culture conditions

Details of all plasmids, bacterial strains, primers, media and antibiotics used in this study are listed in Tables 2.1-5 respectively. General nomenclature is denoted as: plasmids – pJMxxx, strains – JTMxxx and primers JMxxxx, with the "x" corresponding to a specific reference number. Liquid cultures of *E. coli* were routinely grown in 10-50 ml of LB broth at 250 rpm at 37°C unless otherwise stated. Liquid cultures of *S. coelicolor* were grown at 30°C, shaking at 250 rpm with specific media recipes given in later sections if applicable. Typically, 10-50 ml of liquid culture was grown. Cultures grown on solid media were grown at the same temperatures listed above, unless otherwise stated. Where necessary, cultures were supplemented with antibiotics at concentrations listed in Table 2.5. Solid cultures of *Streptomyces* strains used for biomass isolation were grown on top of sterile cellophane discs, covering the media, which allow the organisms to grow while facilitating simple harvesting of the mycelium.

2.1.2 Preparation of *Streptomyces* spores

S. coelicolor or *S. venezuelae* were grown on MS or MYM-tap, respectively. Spores from a single colony were streaked out on agar plates, either using a sterile cotton bud or loop, to form a confluent lawn. Agar plates were typically incubated at 30°C for 5 days for *S. coelicolor* and 2-3 days for *S. venezualae*. Spores were harvested by pipetting 1 ml of 20% glycerol (2G) onto the plate and sloughing off the spores with a sterile cotton bud. A further 1 ml of 2G was added to the agar surface and spores were transferred to a sterile 15 ml falcon tube by pipetting. Spores were briefly vortexed then filtered through sterile cotton wool, using a syringe, into a fresh tube and finally transferred to a sterile centrifuge tube and stored at -20°C.

2.1.3 Glycerol stocks

Glycerol stocks of *E. coli* were produced by taking 1-10 mls of *E. coli* overnight culture, pelleting cells and resuspending it in 1 ml of fresh, sterile, 1:1 LB 40% glycerol (4G) mix. Glycerol stocks where stored at -20°C.

Plasmids	Genotype/description	Resistance	Source
pIJ773	*aac(3)IV oriT bla* (contains apramycin (apr) resistance cassette)	AprR	Gust et al., 2002
pIJ10700	contains hygromycin resistance cassette, FRT oriT-hyg FRT MkII	HygR	Gust et al., 2002
pIJ790	*araC-Parab, Y, ß, exo, cat, repA1001ts, oriR101*	CmR	Gust et al., 2002
pUZ8002	RK2 derivative with a mutation in oriT		Kieser et al., 2000
pSET152	*oriT, lacZα, aac(3)IV, lacO, ColE1, intΦC31, att*	AprR	Kieser et al., 2000
pMS82	*ori,* pUC18, *hyg, oriT,* RK2, int ΦBT1	HygR	Gregory et al., 2003
pIJ10257	*oriT*, ΦBT1 *attB-int*, Hyg[r], *ermEp**, pMS81 backbone	HygR	Hong et al., 2005
pGEM-T-Eazy	*bla, lacZα*	AmpR	Promega
pGUS	*gusA* containing 2.0-kb BamHI fragment cloned into BamHI site of pSET152, KpnI cloned *tipA* promoter	AprR	Myronovskyi et al, 2011
pGS-21a	Genscript overexpression and purification vector (SD0121)	AmpR	Genscript
St1A2	*Supercos-1*-cosmid *with (39829 bp) fragment containing (sco6811-08 and scr6809)*	KanR/AmpR	Redenbach et al., 1996
St4A10	*Supercos-1*-cosmid *with (43147 bp) fragment containing (sco2074 - lsp)*	KanR/AmpR	Redenbach et al., 1996
St5C11	*Supercos-1*-cosmid with (insert size) fragment containing *(sco7427-8, nsrR* and *hmpA1)*	KanR/AmpR	Redenbach et al., 1996
St3A4.2.A04	*Supercos-1*-cosmid with (insert size) fragment containing *(a* transposon disrupted *sco7094 - hmpA2).*	KanR/AmpR/AprR	Fernández-Martínez et al. 2011
SV-5-F05	*Supercos-1*-cosmid with (insert size) fragment containing *(sven6563, RsrR)*	KanR/AmpR	Pullan et al, 2011
pJM009	St1A2 containing *sco6811::apr oriT*	KanR/AmpR/AprR	This work
pJM010	St1A2 containing *sco6808::apr oriT*	KanR/AmpR/AprR	This work
pJM011	St1A2 containing *sco6811-08::apr oriT*	KanR/AmpR/AprR	This work
pJM012	St1A2 containing *scr6809::apr oriT*	KanR/AmpR/AprR	This work
pJM013	St4A10 containing *sco2074::apr oriT*	KanR/AmpR/AprR	This work
pJM014	St4A10 containing *bla::hyg oriT*	KanR/HygR	This work

pJM015	pMS82 containing full length SCO2074 and promoter (300 bp of upstream DNA)	HygR	This work
pJM016	*lsp* suicide vector, pGEM-T-Eazy, 411 bp fragment of the *lsp* gene with a *Bam*HI site. The *aac(3)IV* containing *Bam*HI fragment from a pIJ773 was sub cloned in.	AmpR/AprR	This work
pJM017	pMS82, KpnI/HindIII insert containing the pMC500 MCS and terminators with *scr6809*	HygR	This work
pJM018	St5C11 containing *sco7427::apr oriT*	KanR/AmpR/AprR	This work
pJM019	St5C11 containing *sco7428::apr oriT*	KanR/AmpR/AprR	This work
pJM020	St5C11 containing *sco7427-8::apr oriT*	KanR/AmpR/AprR	This work
pJM021	St3A4 containing *sco7094::hyg oriT*	KanR/AmpR/HygR	This work
pJM022	pIJ10257 containing NdeI/HindIII *sco7428 (hmpA1)*	HygR	This work
pJM023	pGUS KpnI/XbaI cloned 500 bp upstream of *sco7427*	AprR	Knowles., 2014
pJM024	pGUS KpnI/XbaI cloned 500 bp upstream of *sco7428*	AprR	Knowles., 2014
pJM025	pGUS KpnI/XbaI cloned 500 bp upstream of *sco7094*	AprR	Knowles., 2014
pJM026	SV-5-F05 containing *sven6563::apr oriT*	KanR/AmpR/AprR	This work
pJM027	pMS82, *rsrR* gene plus 300 bp upstream DNA with a c-terminal synthetic linker and 3x FLAG tag	HygR	This work
pJM028	pGS-21a, full length *rsrR* cloned NdeI/XhoI	AmpR	This work
pJM029	pJM028 with a c-terminal 6xHis tag	AmpR	This work
pJM030	pJM028 with a c-terminal synthetic linker as with (FLAG), 2xFLAG tag and a 6xHis tag, cloned NdeI/XhoI	AmpR	This work
pJM031	pJM030, containing C91A, C108A, C112A mutations	AmpR	This work

Table 2-1 DNA plasmids and constructs used throughout this study.

Strain				
E. coli	Genotype/description	Plasmid	Resistance	Source
TOP10	F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G			Invitrogen
BW25113	F⁻, DE(*araD-araB*)567, *lac*Z4787(del)::*rrnB*-3, LAM⁻, *rph*-1, DE(*rhaD-rhaB*)568, *hsdR*514	pIJ790	CmR	Datsenko & Wanner, 2000
BL21	F⁻ ompT gal dcm lon hsdS$_B$(r$_B$⁻ m$_B$⁻) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])			Studier & Moffat, 1986
ET12567	*dam- dcm- hsdM-*	pUZ8002	CmR/TetR	MacNeil et al., 1992
Streptomyces	Genotype/description	Plasmid	Resistance	Source
M145	S. coelicolor wild type strain, SCP1-, SCP2-			Hopwood et al, 1985
BJT1000	M145 *lsp::apr*		AprR	Thompson et al., 2010
BJT1001	M145 *lspflp*			Thompson et al., 2010
BJT1004	BJT1000 + Sco *lsp* cis	pJM014		Thompson et al., 2010
Sven	S. venezuelae ATCC 10712 WT strain			Pullan et al, 2011
ΔnsrR	M145, *nsrRflp*			Knowles., 2014
JTM004	*sco6811::apr*	pJM009	AprR	This work
JTM005	*sco6808::apr*	pJM010	AprR	This work
JTM006	*sco6811-08::apr*	pJM011	AprR	This work
JTM007	*scr6809::apr*	pJM012	AprR	This work
JTM008	*lsp::apr*	pJM013	AprR	This work
JTM009	M145 + St4A10 *bla::hyg*	pJM014	KanR/HygR	This work
JTM010	M145 + *lsp trans* complementation	pJM015	HygR	This work
JTM011	*sco6811::apr*	pJM009	AprR	This work
JTM012	*sco6808::apr*	pJM010	AprR	This work
JTM013	*sco6811-08::apr*	pJM011	AprR	This work

JTM014	*scr6809::apr*		pJM012	AprR	This work
JTM015	*lsp::apr*		pJM013	AprR	This work
JTM016	M145 + St4A10 *bla::hyg*		pJM014	KanR/HygR	This work
JTM017	BJT1001 + *lsp* trans complementation		pJM015	HygR	This work
JTM018	M145 + *lsp* suicide vector		pJM016	AmpR/AprR	This work
JTM019	M145 + *scr*6809 overexpression strain		pJM017	HygR	This work
JTM020	BJT1001 + *scr*6809 overexpression strain		pJM017	HygR	This work
JTM021	*nsrR::apr*		pJM018	AprR	This work
JTM022	*hmpA1::apr*		pJM019	AprR	This work
JTM023	*hmpA2::hyg*		pJM020	AprR	This work
JTM024	*nsrR/hmpA1::apr*		pJM021	HygR	This work
JTM025	JTM023, *hmpA1::apr*		pJM019	HygR/AprR	This work
JTM026	JTM023, nsrR/hmpA1::apr		pJM021	HygR/AprR	This work
JTM027	pIJ10257 *sco7428 (hmpA1), ermE** overexpression strain		pJM022	HygR	This work
JTM034	Sven, *rsrR::apr*		pJM026	KanR/AmpR/AprR	This work
JTM035	JTM035, *rsrR* 3x FLAG		pJM027	HygR	This work

Table 2-1 Strains used throughout this study.

Primer	Description	Sequence
JM0001	*nsrR (sco7427)* forward disruption primer (Redirect)	GGCGAACCTAGCATGCGCATTTGATAGCGTCCTGGTGTGATTCCGGGGATCCGTCGACC
JM0002	*nsrR (sco7427)* reverse disruption primer (Redirect)	GGCCACAGCTCGACGGCCTCAGGGGGCCGCCGCCCGTCATGTAGGCTGGAGCTGCTTC
JM0003	*nsrR (sco7427)* forward test primer	CGCCGTCTTCCGCTGCCATC
JM0004	*nsrR (sco7427)* reverse test primer	GTCCCGCAGCAGCTCCGGGT
JM0005	*nsrRp* DNaseI footprinting forward probe (187 bp)	GTGCCCGTACGCCGCCCCGT
JM0006	*nsrRp* DNaseI footprinting reverse probe (187 bp)	GTCTCTCACGACCGCCAGGCGC
JM0007	*hmpA1p* DNaseI footprinting forward probe (187 bp)	GCACCGACCGGCCCGGTTCTG
JM0008	hmpA1p DNaseI footprinting reverse probe (187 bp)	GCCGCTCCGACGGCGGGGA
JM0009	hmpA2p DNaseI footprinting forward probe (187 bp)	GACATCGCGTCACAAGTCCGGCCG
JM0010	hmpA2p DNaseI footprinting reverse probe (187 bp)	GTGCCGAGCGAGGCTCCGACG
JM0011	6fam labelled forward primers for *hmpA2*	ACCCGGTCTCCGGCTTACC
JM0012	6fam labelled reverse primers for *hmpA2*	ACGGGACGCTCCTCGAACA
JM0013	6fam labelled forward primers for *nsrR*	TCTTCCGCTGCCATCAGG
JM0014	6fam labelled reverse primers for *nsrR*	ACCTCGGCCACCTCTCG
JM0015	6fam labelled forward primers for *hmpA1*	ACACTCGACCCACTGACC
JM0016	6fam labelled reverse primers for *hmpA1*	TGGGCGTCGAAGAGCTTG
JM0017	UniFam sequence labelled with 6'Fam	AT CAT GAT GGG G
JM0018	UniFam nested *sco7428* For primer sequence	ATC ATG ATG GGG GCA CTC GAC CCA CTG ACC
JM0019	UniFam nested *sco7428* Rev primer sequence	ATC ATG ATG GGG TGG GCG TCG AAG AGC TTG
JM0020	UniFam nested *sco7428* UniFor primer sequence	ATC ATG ATG GGG GAG GTC TAC TGG CTG ATG GC
JM0021	UniFam nested *sco7428* UniRev primer sequence	ATC ATG ATG GGG GAC GCG CTT GAC GGT GAT
JM0022	UniFam nested *sco6108p* For primer sequence	AT CAT GAT GGG G TTCATTCACATACCACGCAGGT
JM0023	UniFam nested *sco6108p* Rev primer sequence	AT CAT GAT GGG G AGGCTCGCCGGAAATCGC
JM0024	UniFam nested *sco3773p* For primer sequence	AT CAT GAT GGG G AGGTCACGGTCACGGTC

JM0025	UniFam nested *sco3773p* Rev primer sequence	AT CAT GAT GGG G ACGCCCAGCTCCTTCTC
JM0026	UniFam nested *sco1447p* For primer sequence	AT CAT GAT GGG G TACCGGGCGTGTCTCAC
JM0027	UniFam nested *sco1447p* Rev primer sequence	AT CAT GAT GGG G AGGCGTTCGACGAGGTC
JM0028	UniFam nested *sco0166p* For primer sequence	AT CAT GAT GGG G A GAGGCCCCGTGGGAGTCCACG
JM0029	UniFam nested *sco0166p* Rev primer sequence	AT CAT GAT GGG G A GCCTGCCATGCGGCCTCCTGC
JM0030	UniFam nested *sco0447p* For primer sequence	AT CAT GAT GGG G A GTCCGACGAAGACGCCCGCGG
JM0031	UniFam nested *sco0447p* Rev primer sequence	AT CAT GAT GGG G A GCTCATGCCTCCAGGCTAATGTCCACAG
JM0032	UniFam nested *sco0622p* For primer sequence	AT CAT GAT GGG G A GAGGAGACGCTGGCCCGACTGG
JM0033	UniFam nested *sco0622p* Rev primer sequence	AT CAT GAT GGG G A GAGCGGCACGTGTGGCCCGCTT
JM0034	UniFam nested *sco1343p* For primer sequence	AT CAT GAT GGG G A GCGGAGCAGGCCCTGGACCC
JM0035	UniFam nested *sco1343p* Rev primer sequence	AT CAT GAT GGG G A GCACCTCGCGCCAGGATTCGGG
JM0036	UniFam nested *sco1434p* For primer sequence	AT CAT GAT GGG G A GCCCGCTCCCACCTCGCGGA
JM0037	UniFam nested *sco1434p* Rev primer sequence	AT CAT GAT GGG G A GTGCCCCGTCGGTATGCCACC
JM0038	UniFam nested *sco1570p* For primer sequence	AT CAT GAT GGG G A GCGCTCATAAGTCTTTCCCATGTCTAACGATTATG
JM0039	UniFam nested *sco1570p* Rev primer sequence	AT CAT GAT GGG G A GGCCGCCCCAGAGCCGGAC
JM0040	UniFam nested *sco1663p* For primer sequence	AT CAT GAT GGG G A GTGTAGACCTCCGATGCCACACCGG
JM0041	UniFam nested *sco1663p* Rev primer sequence	AT CAT GAT GGG G A GCCGGTCGCGGTGTCGTGGATC
JM0042	UniFam nested *sco2494p* For primer sequence	AT CAT GAT GGG G A GGCGTACGAGAGTTCATGCGTTCAGG
JM0043	UniFam nested *sco2494p* Rev primer sequence	AT CAT GAT GGG G A GACTCCTCGAGGACGCGGTGGC
JM0044	UniFam nested *sco2610p* For primer sequence	AT CAT GAT GGG G A GACGCCTCGCCCCGTAGATGACG
JM0045	UniFam nested *sco2610p* Rev primer sequence	AT CAT GAT GGG G A GGCCGTGCCTTCCTCGTCGG
JM0046	UniFam nested *sco3485p* For primer sequence	AT CAT GAT GGG G A GAATCACGTCCGCGCGTCACAGATAACC
JM0047	UniFam nested *sco3485p* Rev primer sequence	AT CAT GAT GGG G A GAGCCGCGCTCGCGCCGGAA
JM0048	UniFam nested *sco4908p* For primer sequence	AT CAT GAT GGG G A GCGGCAGGCTTCTCGTCACAGGC
JM0049	UniFam nested *sco4908p* Rev primer sequence	AT CAT GAT GGG G A GTGCTGACGCCCGGTGCCGC
JM0050	UniFam nested *sco5085p* For primer sequence	AT CAT GAT GGG G A GCCTCGACCACTGCCTCTCGG

JM0051	UniFam nested *sco5085p* Rev primer sequence	AT CAT GAT GGG G A GTTGAATCTCATCTGCGCCCCCGTCG
JM0052	UniFam nested *sco6535p* For primer sequence	AT CAT GAT GGG G A GACCCGCTCCGCGCACCGAC
JM0053	UniFam nested *sco6535p* Rev primer sequence	AT CAT GAT GGG G A GCCCGGCTCGGTGGGCGGT
JM0054	UniFam nested *sco7168p* For primer sequence	AT CAT GAT GGG G A GCAGCGTTGATGGATCTTCTCCGGGC
JM0055	UniFam nested *sco7168p* Rev primer sequence	AT CAT GAT GGG G A GGCTCACCGACGCCCGCCATG
JM0056	UniFam nested *sco7459p* For primer sequence	AT CAT GAT GGG G A GCCCGGCACCGCGGATGAGTAC
JM0057	UniFam nested *sco7459p* Rev primer sequence	AT CAT GAT GGG G A GTCGGCCTGGGAGCGGGCTG
JM0058	UniFam nested *sco7705p* For primer sequence	AT CAT GAT GGG G A GGCATCACCCGCACGCCCGTC
JM0059	UniFam nested *sco7705p* Rev primer sequence	AT CAT GAT GGG G A GAGGGGTGCTCATGGTGGCCTCC
JM0060	UniFam nested *sco6535p* For primer sequence	AT CAT GAT GGG G A GGGGGGTGGCCCGCAAGCCCT
JM0061	UniFam nested *sco6535p* Rev primer sequence	AT CAT GAT GGG G A CGCGCGCAGCAGCGCCGC
JM0062	M13_Fwd sequence labelled with 6'Fam	CTAAAACGACGGCCAGT
JM0063	M13_Rev sequence labelled with 6'Fam	CAGGAAACAGCTATGAC
JM0064	M13Fam nested *hmpA1p* AUC sequence	GTAAAACGACGGCCAGTCTAAAACACGAATATCATCTACCAATTAAGGTCATAGCTGTTTCCTG
JM0065	M13Fam nested Predicted *nsrRp* sequence	GTAAAACGACGGCCAGTAAGGCGAACCTAGCATGCGCATTGTCATAGCTGTTTCCTG
JM0066	M13Fam nested Predicted *hmpA1p* sequence	GTAAAACGACGGCCAGTAACACGAATATCATCTACCAATTGTCATAGCTGTTTCCTG
JM0067	M13Fam nested Predicted *hmpA2p* sequence	GTAAAACGACGGCCAGTAACAAGCATCTGAGATCCCAGTTGTCATAGCTGTTTCCTG
JM0068	M13Fam nested *hmpA1p* with 2 bp removed from each end of consensus sequence	GTAAAACGACGGCCAGTCACGAATATCATCTACCAGGTCATAGCTGTTTCCTG
JM0069	M13Fam nested *hmpA1p E. coli* sequence	GTAAAACGACGGCCAGTATAAGATGCATTTGAGATACATCAAGTCATAGCTGTTTCCTG
JM0070	M13Fam nested *hmpA1p B. subtilis* sequence	GTAAAACGACGGCCAGTAAGATCATGTATTTTAAAGATATATTTTAGTCATAGCTGTTTCCTG
JM0071	M13Fam nested containing the most conseved *nsrR* binding site bases with preference for GC rich sequence	GTAAAACGACGGCCAGTAACGCGCATCTGAGATGCGCGTTGTCATAGCTGTTTCCTG
JM0072	JC101 with centre base being replaced T-C	GTAAAACGACGGCCAGTAACGCGCATCCGAGATGCGCGTTGTCATAGCTGTTTCCTG
JM0073	M13Fam nested *sco0103p* M13 For primer sequence	CTAAAACGACGGCCAGTAGGCCCTGACGTCAAACCGG
JM0074	M13Fam nested *sco0103p* M13 Rev primer sequence	CAGGAAACAGCTATGACCGTCCGCTCCAGCCCTGTTC

JM0075	M13Fam nested *sco0715p* M13 For primer sequence	CTAAAACGACGGCCAGTTCTCCAGTGATCGGGGCCTGACA
JM0076	M13Fam nested *sco0715p* M13 Rev primer sequence	CAGGAAACAGCTATGACAAGTCGGCCTCGTCGATCCGGAT
JM0077	M13Fam nested *sco2358p* M13 For primer sequence	CTAAAACGACGGCCAGTTACTTCGTCGTCTGCGAGGCGGT
JM0078	M13Fam nested *sco2358p* M13 Rev primer sequence	CAGGAAACAGCTATGACCCTTCAGGAGCAGCTCGGCATTGC
JM0079	M13Fam nested *sco3359p* M13 For primer sequence	CTAAAACGACGGCCAGTGTTCGTCGAGGACGACGATGTCATCCG
JM0080	M13Fam nested *sco3359p* M13 Rev primer sequence	CAGGAAACAGCTATGACCTGCATTGTGCACGACCAGGCTCA
JM0081	P1 sequence for amplification of disruption cassettes	ATTCCGGGGATCCGTCGACC
JM0082	P2 sequence for amplification of disruption cassettes	TGTAGGCTGGAGCTGCTTC
JM0083	*sco6808* forward disruption primer (Redirect)	GTCTATGGTTGACGGGTGACTGTCATAGATCTGCAGATGATTCCGGGGATCCGTCGACC
JM0084	*sco6808* reverse disruption primer (Redirect)	GTCATCTTCCGAACGGAGATGGAGGGAGATCCGGAATCATGTAGGCTGGAGCTGCTTC
JM0085	*sco6808* forward test primer	CGGAGGCCGCCTGTCCTAGC
JM0086	*sco6808* reverse test primer	AACGCGCACTCGCTGCGGTC
JM0087	*sco6811* forward disruption primer (Redirect)	GGCGGCATCGACAAACATCGAAGCCGAGGAGTCATCGTGATTCCGGGGATCCGTCGACC
JM0088	*sco6811* forward disruption primer (Redirect)	GGCTCCCGCGGCCGCAACGGCTGCGGGTGGGGACGATCATGTAGGCTGGAGCTGCTTC
JM0089	*sco6811* forward test primer	AACGAGCTTGAGCGGCGGCTCG
JM0090	*sco6811* reverse test primer	GCTCTTCCTTGCCCAGCGGCTG
JM0091	*scr6809* forward disruption primer (Redirect)	TCCGACATCTGCAGATCTATGACAGTCACCCGTCAACCAATTCCGGGGATCCGTCGACC
JM0092	*scr6809* reverse disruption primer (Redirect)	TGGTACACGGCACCGACTCCGGCTGCCAGAAAGCCATAGTGTAGGCTGGAGCTGCTTC
JM0093	*scr6809* forward test primer	CAGACGCAGGCCTCGCCATC
JM0094	*scr6809* reverse test primer	CCCATCGCTACGGCCGCCT
JM0093	*bla (bla::hyg)* gene forward disruption primer (Redirect) for supercos-1	AATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTCAGGCGCCGGGGGCGGTG
JM0096	*bla (bla::hyg)* gene forward disruption primer (Redirect) for supercos-1	CCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTAAGTTCCCGCCAGCCTCGCA
JM0097	*bla (bla::hyg)* gene forward test primer (Redirect) for supercos-1	AAGCAGCAGATTACGCGCAG

JM0098	*bla (bla::hyg)* gene forward test primer (Redirect) for supercos-1	GTGCGCGGAACCCCTATTTG
JM0099	*lsp (Sco2074)* forward disruption primer (Redirect)	TCGTGCTCAGTCAAGGACCTAGGCTGAGGGACTCACGTGATTCCGGGGATCCGTCGACC
JM0100	*lsp (Sco2074)* forward disruption primer (Redirect)	GACAACCAGTCCCTGTGGACAGCCGGACCGGAGGGGTCATGTAGGCTGGAGCTGCTTC
JM0101	*hmpA1 (sco7428)* forward disruption primer (Redirect)	AAAACACGAATATCATCTACCAATTAAGGAGTCGCTGTGATTCCGGGGATCCGTCGACC
JM0102	*hmpA1 (sco7428)* reverse disruption primer (Redirect)	GAGTTTTGCGGCGGGCGCCGTCGGCCGCCACCGGCGCTATGTAGGCTGGAGCTGCTTC
JM0103	*hmpA1 (sco7428)* forward test primer	GCGGCGTGCGCTGCGCGACG
JM0104	*hmpA1 (sco7428)* reverse test primer	GCGGGCGGCTCGGTGTCCTG
JM0105	*hmpA2 (sco7094)* forward disruption primer (Redirect)	GAAAACAAGCATCTGAGATCCCAGTTCGGAGTAGGCATGATTCCGGGGATCCGTCGACC
JM0106	hmpA2 (*sco7094*) reverse disruption primer (Redirect)	GACGCCCGCCCGCTCACTGCCGGGCGGCCGGCGCTGCTATGTAGGCTGGAGCTGCTTC
JM0107	hmpA2 (*sco7094*) forward test primer	CCAAGCCCAGGGAGTCGTCC
JM0108	hmpA2 (*sco7094*) reverse test primer	CCACCTGCGCAAGGTCTTCG
JM0109	*rsrR (sven6563)* forward disruption primer (Redirect)	CCAGTCCCCTCCCCCACGGACCTGCTGCGTCGCACCATGATTCCGGGGATCCGTCGACC
JM0110	*rsrR (sven6563)* reverse disruption primer (Redirect)	CACCGAACAGCCAAGCCCCCCTCAGCAAGCCCTCCCTCATGTAGGCTGGAGCTGCTTC
JM0111	*rsrR (sven6563) forward* test primer	ACGCGGCGACCACGTCGTGG
JM0112	*rsrR (sven6563)* reverse test primer	GCCCGTACGGTAGACCGCCG
JM0113	pMS82 cloning forward test primer	GCAACAGTGCCGTTGATCGTGCTATG
JM0114	pMS82 cloning reverse test primer	GCCAGTGGTATTTATGTCAACACCGCC
JM0115	Forward primer amplifies a 411 bp fragment of the lsp gene, adding a bamHI site upstream.	GGATCCCTGTTCGCGGTCGCCCTGTTCGCGTACCT
JM0116	Reverse primer amplifies a 411 bp fragment of the lsp gene	GATGCCGCCGCACACGATCGCCGAGTCGG
JM0117	M13Fam nested *sven1847* for primer sequence	CTAAAACGACGGCCAGTTCTCCTCGCCCGCCCCGTCG
JM0118	M13Fam nested *sven1847* rev primer sequence	CAGGAAACAGCTATGACCCGTCCGGCGCCCCGGGTGG
JM0119	M13Fam nested *sven3827* for primer sequence	CTAAAACGACGGCCAGTCTCGCCCACTCGCCGTACCG
JM0120	M13Fam nested *sven3827* rev primer sequence	CAGGAAACAGCTATGACCATCACGAGATCGCCCGCCT
JM0129	M13Fam nested *sven6563* for primer sequence	CTAAAACGACGGCCAGTTCGTCGAAGGTCGGGGAGTT

JM0130	M13Fam nested *sven6563* rev primer sequence	CAGGAAACAGCTATGACCGTGCAGCTCAGCGAGCCGG
JM0131	M13Fam nested *sven0247* for primer sequence	CTAAAACGACGGCCAGTTCGTCATGATCGTGTGGCGGCTGCG
JM0132	M13Fam nested *sven0247* rev primer sequence	CAGGAAACAGCTATGACAGCACCAGCCGCTCGTCGAACGCGG
JM0133	M13Fam nested *sven0519*for primer sequence	CTAAAACGACGGCCAGTAGACGATGATCAACGTGAAGGTGTCCG
JM0134	M13Fam nested *sven0519* rev primer sequence	CAGGAAACAGCTATGACAAGGTCGCGACGCACACCATGATCAT
JM0135	*sven6563/2* for DNaseI footprinting forward probe	TGCCGCCGATGACGGTGATCTTCATGGTGT
JM0136	*sven6563/2* rev DNaseI footprinting reverse probe	TCAGGACCACACAGCAGTGGAGCGCCCA
JM0141	M13Fam nested *sven6562/3* Site 1-4 primer sequence	CTAAAACGACGGCCAGTCAAACTCGGATACCCGATGTCCGAGATAATACTCGGATAGTCTGTGTC CGAGTCAAGTCATAGCTGTTTCCTG
JM0142	M13Fam nested *sven6562/3* Site 1-2 primer sequence	CTAAAACGACGGCCAGTGCAAACTCGGATACCCGATGTCCGAGATAATGTCATAGCTGTTTCCTG
JM0143	M13Fam nested *sven6562/3* Site 3-4 primer sequence	CTAAAACGACGGCCAGTTAATACTCGGATAGTCTGTGTCCGAGTCAAAGTCATAGCTGTTTCCTG
JM0144	M13Fam nested *sven6562/3* Site 1 primer sequence	CTAAAACGACGGCCAGTGCAAACTCGGATACCCGGTCATAGCTGTTTCCTG
JM0145	M13Fam nested *sven6562/3* Site 2 primer sequence	CTAAAACGACGGCCAGTCCGATGTCCGAGATAATGTCATAGCTGTTTCCTG
JM0146	M13Fam nested *sven6562/3* Site 3 primer sequence	CTAAAACGACGGCCAGTTAATACTCGGATAGTCTGTCATAGCTGTTTCCTG
JM0147	M13Fam nested *sven6562/3* Site 4 primer sequence	CTAAAACGACGGCCAGTTCTGTGTCCGAGTCAAAGTCATAGCTGTTTCCTG

Table 2-2 Primers used thoughtout this study.

Media	Composition	Weight, %v/v, %w/v
Lysogenic Broth (LB)	Tryptone	10 g
	Yeast extract	5 g
	NaCl	5 g
	dH$_2$O	To 1000 ml
For solid media	Agar	15 g
YT	Tryptone	20 g
	Yeast extact	10 g
	dH$_2$O	To 1000 ml
MS	Mannitol	20 g
	Soya flour	20 g
	dH$_2$O	To 1000 ml
For solid media	Agar	20 g
TSB	TSB powder (Oxoid	30 g
	dH$_2$O	To 1000 ml
Yeast extract, malt extract media	Yeast extract	3 g
	Peptone	5 g
	Malt extarct	3 g
	Glucose	10 g
	dH$_2$O	To 1000 ml
MYM-tap	Maltose	4 g
	Yeast Extract	4 g
	Malt Extract	10 g
	dH$_2$O/Tap water	500/500 ml
	R2 trace elements	2 ml*
For solid media	Agar	20 g
Minimal media	L-asparagine	0.5 g
	K$_2$HPO$_4$	0.5 g
	MgSO$_4$.7H$_2$O	0.2 g
	FeSO$_4$.7H$_2$O	0.01 g
	Glucose	10 g*
	dH2O	To 1000 ml
For solid media	Agar	20 g
SMM	PEG 6000 (BDH 6.1% w/v in	5% w/v
	MgSO$_4$. 7H$_2$O	5 mM
	TES buffer	25 mM
	NaH$_2$PO$_4$+K$_2$HPO$_4$	1 mM
	Glucose	1% w/v
	Antifoam 289 (sigmaA 5551)	0.2% w/v
	Trace elements	1 ml*
	dH2O	To 1000 ml
	Glycine (20%)	25 ml
Trace elements for SMM (filter Sterilised)	ZnSO$_4$.7H$_2$O	0.1 g
	FeSO$_4$.7H$_2$O	0.1 g
	MnCl$_2$.4H$_2$O	0.1 g
	CaCl$_2$.6H$_2$O	0.1 g
	NaCL	0.1 g

Table Error! No text of specified style in document.**-1** Growth media used throughout this study.

* = Added after autoclaving

Antibiotics	Concentration		
	Stock (mg/ml)	For growth media (µg/ml)	For overlays (mg/ml)
Ampicillin	100	100	
Apramycin	50	50	1.25
Chloramphenicol	25	25	
Hygromycin	25	25	0.625
Kanamycin	50	50	
Nalidixic acid	25	25	0.5

Table 2-5 Concentrations of antibiotics used throughout this study.

2.2 Centrifugation

All centrifuge steps were carried out at 21,130 x g / 15,000 rpm in a bench top Centrifuge 5424 (eppendorf) for microcentrifuge tubes or at 3,345 x g / 4,000 rpm in an Accupsin 1R with a Ch. 007379 rotor (Fisher scientific) for flacon tubes (1-50 ml volumes) unless otherwise stated.

2.3 Genetic Manipulation

2.3.1 DNA/RNA preparation

2.3.1.1 Plasmid Preparations

Plasmid DNA was prepared using Qiaprep Spin Miniprep kits (Qiagen) from 5-10 ml overnight cultures as per manufacturer's instructions. Plasmids were eluted from the column routinely using 50 µl of autoclaved distilled water (dH$_2$O) unless otherwise stated.

2.3.1.2 Cosmid Preparations

Cosmid DNA was prepared either by phenol extraction (Gust et al. 2002) or by Wizard® Plus SV Miniprep DNA Purification System as per the manufacturer's instructions.

Phenol extractions were carried out using 1-3 ml of *E. coli* overnight culture. Cells were sequentially pelleted using a bench top centrifuge at 21,130 x g for 30 s in 1.5 ml micro-centrifuge tubes. Pellets were resuspended in 100 µl of solution I (50 mM Tris-HCl, pH8; 10 mM EDTA) followed by addition of 200 µl of solution II (200 mM NaOH; 1% SDS) and mixed by inversion. Immediately, 150 µl of solution III (3 M potassium acetate, pH 5.5) was added and the solution was mixed by inversion. Each sample was centrifuged for 5 min at room temperature (RT). The supernatant was immediately extracted with 400 µl of

a phenol/chloroform and vortexed for 2 min, followed by a further centrifugation for 5 min. The upper phase was transferred to a fresh tube and the DNA precipitated by adding 800 μl of 100% ethanol and storing at -20°C for 10 min. Tubes were then centrifuged as above, the pellets washed in 70% ethanol and centrifuged again. The sample tubes were left open at RT to dry the pellet and resuspended in 50 μl dH$_2$O. A sterile pipette tip was dipped into DNase-free RNase (Sigma-Aldrich) then into the DNA sample tube, which was then left at RT for 15 min to remove RNA, and finally stored at -20°C.

2.3.1.3 Chromosomal DNA preparation from *Streptomyces*

Streptomyces cultures were grown overnight in a mixture of 50% TSB / 50% YEME media at 30°C shaking at 250 rpm. Cells were isolated by centrifugation at 3,345 x g for 5 min. The supernatant was discarded and the cells resuspended in 500 μl Solution I (50 mM Tris/HCl, pH8; 10 mM EDTA) and transferred to a micro-centrifuge tube. Aliquots of 10 μl of filter-sterilised lysozyme solution (30 mg/ml) and 5 μl DNase-free RNase (10 mg/ml) were added to the samples and incubated for 1 h at 30°C followed by addition of 5 μl of 20% SDS, with samples mixed in by inversion. Approximately 500 μl (one volume) of 1:1 phenol-chloroform was added and samples were mixed thoroughly by vortexing for 1 min and then centrifuged at 15,000 rpm for 5 min. The upper aqueous phase (containing the DNA) was removed and transferred to a fresh microfuge tube. The phenol chloroform step was repeated until the upper phase was clear (i.e. protein free). The clear aqueous layer was transferred to a fresh microfuge tube. Then 1 ml of 100% ethanol was added to the tube and it was mixed by inversion followed by centrifugation as above. Ethanol was removed and the DNA pellet was washed in 70% ethanol and centrifuged for a further 2 min at 21,130 x g. The DNA pellet was dried at RT ensuring all ethanol was removed and the pellet was finally resuspended in 50 μl sterile dH$_2$O and stored at 4°C.

2.3.1.4 RNA isolation from *Streptomyces*

Mycelium was harvested at experimentally appropriate time points and immediately transferred to 2 ml round bottom tubes, flash frozen in liquid N$_2$ then stored at -80°C or used immediately. All apparatus used was treated with RNaseZAP (Sigma) to remove RNases for a minimum of 1 h before use. RNaseZAP treated mortar and pestles were used, the pestle being placed and cooled on a mixture of dry ice and liquid N$_2$ with liquid N$_2$ being poured into the bowl and over the mortar. Once the bowl had cooled the mycelium samples were added directly to the liquid N$_2$ and thoroughly crushed leaving a fine powder of mycelium. Grindings were transferred to a pre-cooled 50 ml Falcon tube and stored on dry ice. Directly to the tube, 2 ml of TRI reagent (Sigma) was added to the grindings and mixed. Samples are then thawed while vortexing intermittently at room

temperature for 5-10 min until the solution cleared.   To 1 ml of TRI reagent resuspension, 200 µl of chloroform was added and vortexed for 15 s at room temperature then centrifuged for 10 min at 21,130 x g. The upper, aqueous phase (clear colourless layer) was removed into a new 2 ml tube. The remainder of the isolation protocol follows the RNeazy Mini Kit (Qiagen) instructions carrying out both on and off column DNase treatments. On column treatments were carried out following the first RW1 column wash. DNaseI (Qiagen) was added (10 µl enzyme, 70 µl RDD buffer) to the column and stored at RT for 1 h. The column was washed again with RW1 then treated as described in the manufacturer's instructions. Once eluted from the column, samples were treated using TURBO DNA-free Kit (Ambion) following manufacturer's instructions to remove residual DNA contamination.

2.3.2   Analysis and construct design

2.3.2.1   DNA/RNA analysis

DNA (plasmid, cosmid and genomic) and RNA concentrations were routinely quantified using a NanoDrop ND2000c Spectrometer (Thermo Scientific). Samples sent for Next-generation sequencing (NGS) were analysed further for improved accuracy: ChIP-seq samples were analysed by PicoGreen quantitation (Ahn et al. 1996) and RNA/RNA-seq samples were analysed using an Experion Automated Electrophoresis platform (Bio-Rad) using RNA StdSens chips (Bio-Rad), both following the manufacturer's instructions.

2.3.2.2   Polymerase chain reaction (PCR)

GoTaq (Promega) or BioTaq polymerase (Bioline) were used for colony PCR and Phusion or Q5 polymerase was used for high fidelity PCR and cloning in each case following the manufacturer's instructions unless otherwise stated. The primers utilised throughout are listed in table 2.3.

2.3.2.3   General restriction digest

Reactions using a single enzyme were routinely 50 µl total volume and consisted of up to 44 µl dH$_2$O, 1 µl of the enzyme, 5 µl of appropriate buffer (Roche) and 1 µg of DNA to be digested. Digestions were carried out in a 37°C water bath for 1-3 h. Reactions requiring two enzymes with appropriately functional buffers differed with the addition of 1 µl of second enzyme at the expensive of 1 µl of water. Any specific additions required for enzymes will be mentioned when applicable.

2.3.2.4 DNA cloning

Ligations were carried out following instructions of manufacturers for T4 ligase, (New England Biolabs inc.) or ligase independent methods such as using the TOPO TA Cloning Kit (Invitrogen). A molar ratio of 3:1 insert to vector was typically used. Ligations were carried out for either 1 h at 37°C or overnight at RT.

2.3.2.5 Preparation of CaCl$_2$ competent cells

A modified version of the CaCl$_2$ protocol by Cohen *et al.* (1972), as described below, was used. All centrifugation steps were at 3,345 x g, 4°C for 10 min to pellet cells.   An overnight culture (~16 h) of each *E. coli* strain was grown in 10 ml of sterile LB broth from frozen cells stored at -80°C, grown at 37°C, 250 rpm. Aliquots (100 µl) of these were then added to 500 ml of sterile LB.   The cells were then grown at 37°C, 250 rpm, until an optical density (OD) 600 nm of 0.3-0.4 was reached (1 cm path length). At all steps the reagents and equipment were maintained at ~4°C beyond this point. The samples were centrifuged; the bacterial pellet was resuspended and washed gently in 125 ml of ice-cold sterile 100 mM MgCl$_2$. The suspension was centrifuged and the pellet resuspended in 25 ml of ice-cold sterile 100mM CaCl$_2$ before a further 225 ml was added. The suspension was left on ice for ~25 min. Finally, the suspensions were centrifuged and resuspended in 10 ml ice-cold sterile 100 mM CaCl$_2$, 20% (v/v) glycerol, dispensed into 100 µl aliquots and stored at -80°C.

2.3.2.6 Transformation of CaCl$_2$ competent cells

The process was carried out by first gently thawing a 100 µl aliquot of pre-prepared competent cells. Two 50 µl aliquot were used, one as a negative control and one for the sample reaction. Aliquots (1 µl) of plasmid DNA were added to the sample reactions and 1 µl of dH$_2$O to the control reaction. Both were left on ice (4°C) for at least 30 min then heat shocked for 90 s then immediately cooled at 4°C for 2 min. Aliquots (450 µl) of LB was added to each reaction and incubated at 37°C for 1 h at 250 rpm then plated on LB agar (containing appropriate antibiotics when applicable) at 37°C for ~16 h unless otherwise stated.

2.3.2.7 Preparation and transformation of electrocompetent cells

Cells were grown as 10 ml overnight cultures in LB (containing appropriate antibiotics) to an OD of 0.4-0.6 and pelleted by centrifugation in a 15 ml Falcon tube at 4000 rpm for 5 min at 4°C. Cells were resuspended and pelleted in 1 ml ice-cold 10% (v/v) glycerol (1G) 6 times and either flash frozen using liquid nitrogen and stored at -80°C for future use or

used immediately for transformation. DNA (cosmid 2 µg, plasmid 0.1-1 µg) was added to ~50 µl of cells immediately before electroporation. Cells and DNA were transferred to an ice-cold electroporation cuvette and electroporated using the BioRad® Electroporator set to 200 Ω, 25 µF and 2.5 kV. The electroporated cells were diluted in YT and transferred from the cuvette to a micro-centrifuge tube, incubated for 1 h shaking (250 rpm) at 37°C before plating onto LB containing appropriate antibiotic selection.

2.3.2.8 Agarose gel electrophoresis

DNA fragment size was determined using agarose gels. Gels consisted of 0.9% (w/v) agarose in 1x TBE buffer (90 mM Tris-HCl, 90 mM Boric Acid, 2 mM EDTA) with 1.5 µl ethidium bromide/50 ml of TBE. Samples were mixed with 0.25 volumes of DNA-loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene-cyanol blue, 40% (w/v) sucrose in water) and the gels were run in 1x TBE buffer at 120V for approximately 45 min. A 1Kb DNA ladder (Invitrogen) was run alongside the samples, and the DNA was visualised by exposure to UV light.

2.3.2.9 DNA extraction from agarose gel

Gel fragments containing DNA bands of interest were excised using a sterile scalpel and extracted using a Qiaquick Gel Extraction Kit (Qiagen), as per manufacturer's instructions. DNA was eluted in 50 µl autoclaved dH_2O.

2.3.2.10 DNA isolation from PCR reactions

Following PCR, if a single band was produced, as determined by agarose gel electrophoresis, and no gel purification was required, then the product was purified using the QIAquick PCR Purification Kit following the manufacturer's instructions, otherwise, bands were excised using a scalpel and the QiAquick Gel Extraction kit protocol used.

2.3.3 Constructing gene knockouts: Redirect methodology

2.3.3.1 Generating a knockout PCR product

Antibiotic resistance cassettes from either pIJ773 or pIJ10700-1 containing either apramycin (*apr*) or hygromycin (*hyg*) resistance genes respectively and an origin of transfer (*oriT*), were amplified by PCR using primers specific for the disruption of the gene of interest (See Table 2.4). Details of plasmids used and how to design appropriate primers to leave in frame gene knockouts (KO) were reported by Gust et al. (2002). The forward primers consisted of 39 nucleotides (nt) up stream of the gene of interested ending in ATG, the translation start codon, with the 20 nt P1 sequence (see Table 2.3)

corresponding to the 5' end of the antibiotic resistance cassette. The reverse primer had 39 nt of antisense sequence ending TGA, the translational stop codon of each gene plus a 19 nt sequence of P2 corresponding to the end of the antibiotic resistance cassette. The PCR cycling conditions consisted of 94°C for 2 min (initial denaturation step), followed by 10 cycles of 94 ^{o}C for 45 s (denaturation), 50 ^{o}C for 45 se (primer annealing) and 72°C for 90 s (extension), followed by 15 cycles with 94 ^{o}C for 45 s (denaturation) 55 ^{o}C for 45 s (primer annealing) and 72 ^{o}C for 90 s (extension) and final a single step of 72 ^{o}C for 5 min (final extension). PCR reactions typically were carried out using Taq (GoTaq or BioTaq). PCR products were checked by agarose gel electrophoresis; gel was extracted then stored at -20 ^{o}C until use.

2.3.3.2 Introducing cosmids into *E. coli*

Cosmids containing the wild-type gene to be targeted were obtained from either the John Innes Centre (JIC, Norwich) or Paul Dyson (University of Swansea). An aliquot (~50 µl) of *E. coli* BW25113/pIJ790 electro-competant cells were transformed with ~2 µg of cosmid DNA. The cells were grown in 1 ml of LB for 1 h and plated onto LB agar plates containing ampicillin and kanamycin to select for the incoming cosmid and chloramphenicol to select for the λRED recombinant plasmid pIJ790. The plates were incubated at 30°C over night. Cosmid identity was assessed routinely using *Bam*HI digests (1 µl *Bam*HI, 2 µl Roche Buffer B, 17 µl cosmid DNA) incubated at 37°C for 1 h and separated on a 0.9% agarose gel compared to *in silico* results obtained using ApE- A plasmid Editor v2.0.37 and sequence data from http://strepdb.streptomyces.org.uk. Additionally, cosmids were confirmed by positive amplification of the gene of interest. *E. coli* BW25113/pIJ790 (50 µl aliquot) was transformed as previously described using 2 µg of isolated wildtype cosmid DNA grown for 1 h at 30°C with shaking at 250 rpm and plated as described. A single colony was selected, picked and transferred to a 10 ml vial of LB containing antibiotic selection and grown at 30°C overnight and stored as a glycerol stock as previously described.

2.3.3.3 PCR-Targeting the *Streptomyces* cosmids.

From an overnight culture of *E. coli* BW25113/pIJ790 containing the target cosmid, 100 µl was inoculated and grown in 10 ml LB broth containing the appropriate antibiotic selection and 100 µl 1M L-arabinose at 30°C for 4 hours. The arabinose is essential as it induces the λRED genes on pIJ790 facilitating transformation with linear DNA.

These cells were then made electrocompetent (as described in section 2.3.2.7) and a 50 µl aliquot was electroporated with 2 µl of the KO PCR product with flanking regions homologous to the gene of interest to cause an in frame deletion of the chosen gene.

These cells were then incubated for 1 hour shaking at 37°C and ultimately plated on to LB plates containing kanamycin (selection for the cosmid) and apramycin (selection for the gene deletion). Each plate was incubated overnight at 37°C to promote loss of the temperature-sensitive plasmid, pIJ790. Single colonies where then picked and grown for 16 h at 37°C.

2.3.3.4 Checking the mutagenised cosmid

The PCR targeted cosmids were isolated from overnight cultures as described and checked by PCR for the gene disruption. Primers specific for the flanking wildtype region, or for the disruption cassette (P1 and P2), were used in combination, as well as checking by restriction digest (where appropriate), using appropriate enzymes, to check the disruption was successful.

2.3.3.5 Conjugating the mutant cosmids into *Streptomyces*

S. coelicolor contains a methylation-sensing restriction system and as such it is essential to passage disruption cosmids through a non-methylating (*dam- dcm*) *E. coli* strain ET12567 before introduction into *Streptomyces*. This is not required for *S. venezuelae* but the ET12567 strain containing the driver plasmid pUZ8002 is still used for all conjugations.

ET12567 / pUZ8002 (Table 2.2) was transformed by electroporation with 2 µg of cosmid DNA and subsequently plated onto LB agar containing chloramphenicol (to maintain *dam* mutation) and apramycin (to select for the incoming cosmid). Plates were incubated overnight at 37°C and single colonies were selected and grown in 10 ml LB broth at 37°C overnight in the presence of the antibiotics described previously, in addition to kanamycin and ampicillin.

A sample inoculum from the overnight cultures (500 µl) was diluted in 10 ml of fresh LB broth containing antibiotics for selection was grown shaking at 37°C until cells reached an OD_{600} of 0.4-0.6. Cultures were centrifuged at 21,130 x g for 5 min and the resulting pellet washed in fresh LB twice, to remove the selection antibiotics (potentially harmful to non-resistant recipient *Streptomyces* species). Cell pellets were resuspended in 1 ml of LB broth.

Washed *E. coli* cells (500 µl) were mixed with 490 µl of LB broth containing *Streptomyces* >1 x 10^7 *Streptomyces* spores. The mixture was centrifuged briefly and the supernatant removed. The resulting pellet was resuspended in ~100 µl of LB and a dilution series

made so as to avoid a lawn of growth, before incubation overnight for 16-20 hours at 30°C.

Following this incubation period, each conjugation plate was overlaid with 1 ml of sterile dH$_2$O containing 0.5 mg naladixic acid, selective bactericidal antibiotic for *E. coli* and either 1.25 mg of apramycin or 0.625 mg hygromycin, as appropriate, to select for insertion of the incoming cosmid by recombination. The overlay solution was distributed over the surface gently using a sterile spreader and incubated at 30°C for 4 days or until colonies appeared. Colonies were then spotted onto a master plate containing a grid and replicated, using velvets, onto LB agar containing kanamycin, then apramycin, then no selective antibiotic (LB only) and incubated at 30°C for 2 days.

Double crossover (homologous recombination) events were selected using apramycin (or hygromycin) resistance and kanamycin sensitivity as the marked antibiotic cassette has now replaced the target gene in the chromosomal DNA and the cosmid has been lost. These colonies were then picked from the master plate and purified for single colonies and replica plated a further two times to ensure that kanamycin resistance was completely absent. Spore stocks were prepared for these double cross-over exconjugants and stored at -20°C.

2.3.4 Protein Methods

2.3.4.1 Protein preparation, gel electrophoresis and analysis

2.3.4.1.1 Cell lysis (*Streptomyces*)

An overnight culture of the desired *Streptomyces* strain was grown shaking at 250 rpm at 30°C in 10-50 ml of TSB/YEME (*S. coelicolor*) or MYM-Tap media (*S. venezuelae*) containing springs for aeration. Mycelium was harvested by centrifugation at 3,345 x g for 5 min in a bench top centrifuge. Media was discarded and the resulting pellet was washed in 1 ml of TCB (100 mM Tris, 50 mM NaCl, pH8) and transferred to a microcentrifuge tube before centrifuging again and resuspending in fresh TCB ranging from 1-500 µl TCB (or enough to resuspend the pellet) followed by addition of 50x EDTA free protease inhibitor (Roche).

The sample was then sonicated, on ice, at 50 kHz for 10 s followed by 50 s on ice and repeated 5 times to fully rupture all cells.

All protein samples were prepared in this manner unless otherwise stated.

2.3.4.1.2 Isolation of cytoplasmic and membrane protein fractions (*Streptomyces*)

Cells were ultracentrifuged (Optima™ MAX-E ultracentrifuge) at 627,000 x g for 50 min. The resulting supernatant was removed (which contained the total cytoplasmic fraction of the cellular protein), the pellet (containing membrane bound proteins) was resuspended in 200 µl of TCB containing protease inhibitor as above and both fractions stored at -20°C.

2.3.4.2 Bradford assay

The Bradford assay method was adapted from Bradford (1976). Dye reagent was prepared by diluting 1:5 concentrated (5x) Bradford reagent (BioRad, UK) using deionised water (dH$_2$O). Using a 1 mg/ml stock of Bovine Serum Albumin (BSA) as a standard, dilutions were produced to 0, 1, 2, 4, 8, 12 and 16 µg/ml concentrations in dH$_2$O to a final volume of 20 µl volume. Dilute Bradford reagent (980 µl) was added to 20 µl of dilute protein standards. Samples were produced in the same way at suitable concentrations. Each was left for ~5 min then measured at 595nm comparing sample absorbance to those of the known standards in triplicate using a 96 well plate.

2.3.4.3 Iron assay

Quantification of Fe^{2+} content was carried out as described by (Crack *et al*., 2006). All samples and standards had 100 µl of 21.7% HNO$_3$ added and were incubated for 30 min at 95°C. The reactions were cooled and centrifuged briefly before adding 0.6ml of 7.5% (w/v) ammonium acetate and 100 µl of freshly made 12.5% (w/v) ascorbic acid, samples were mixed by inversion. To each, 100 µl of fresh 10 mM Ferene was added and again mixed by inversion. The reaction was left at RT for 30 min then had its absorbance measured at 593nm. With no Fe^{2+} present the reaction is yellow, whilst with increasing concentrations the reaction becomes increasingly blue. A standard curve was produced (at concentrations between 0 and ~200 µM in a 100 µl volume at 20 µM increments) using an Fe^{2+} standard solution prepared by diluting a stock solution to 200.133 µM Ferric ion in 0.5M nitric acid and subsequently processed by the same method as samples.

2.3.4.4 SDS-PAGE and analysis

All reagents and buffers for the sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) gels were prepared as described by Laemmli (1970) with the following modifications. The resolving gel was produced to a final concentration of 12.45% (0.375 M Tris-HCl (pH8.8), 0.1% SDS and 41.5% for total buffer volume being 30% (v/v) acrylamide solution), a stacking gel final concentration of 6% (2.5 ml 4x stacking buffer, 20% (v/v) of 30% (v/v) acrylamide solution) and both gel types were polymerised

chemically using 100 µl of 10% (w/v) ammonium persulphate (APS) and 10 µl Tetramethylethylenediamine (TEMED) as supplied. The running buffer was diluted 1:10 from a 10x stock solution (final concentrations 0.025 M Tris-HCl, 0.192 M glycine, 1% (w/v) SDS). Sample buffer concentrations were 0.0625 M Tris-HCl (pH 6.8), 2% (w/v) SDS, 10% (v/v) glycerol, 5% (v/v) β-mercaptoethanol and 0.001% (w/v) Bromophenol blue.

The BioRad Mini-PROTEAN®II Electrophoresis Cell was used and assembled as described in the instruction manual supplied. Glass plates were cleaned using ethanol and allowed to dry. The gel plates and clamps were assembled as described. The glass was marked 5 mm below the bottom of the well comb to mark the divide between the gels. The resolving gel was prepared to a final volume of 10 ml (suitable for two gels) adding the APS and TEMED to chemically polymerise the gel. The mixture was swirled to mix and poured between gel plates to the level mark then layered with 100% ethanol and left to polymerise for ~30 min. The stacking gel was then prepared (suitable for four gels) and added once the resolving gel had set (the ethanol was removed first and the surface washed with dH$_2$O) and the comb inserted and left for ~30 min to set.

Samples were prepared as described above and sample buffer mixed with protein samples in a 1:1 ratio before use. Before resolution of the proteins the samples were completely denatured by immersing each in boiling water for 1.5 min. Electrophoresis was carried out at 200V for ~45 min or until the loading dye front was ~1 cm from the bottom of the gel.

2.3.4.5 Coomassie Blue staining

To visualise resolved protein bands with the naked eye, gels were stained with Instant Blue (Expedeon) following the manufacturer's instructions.

2.3.4.6 Western blot analysis

Protein samples were loaded in equal concentrations onto an SDS-PAGE gel and separated as above using prestained SDS-PAGE standards (BIO-RAD PageRuler™ Prestained Protein Ladder, 10 to 180 kDa) as a size marker. Proteins were transferred to a nitrocellulose membrane (BIO-RAD Immuno-Blot® PVDF Membrane), pre-soaked in methanol, then transfer buffer using a BIO-RAD trans-blot SD semi-dry transfer cell, set up according to the manufacturer's instructions and run at 10V for 1 h. Following transfer, the nitrocellulose membrane was blocked, using Qiagen blocking solution (0.25 g of blocking reagent dissolved per 50 ml of Tris buffered saline (50 mM Tris, 0.85% (w/v) NaCl, pH 7.4) + 0.1% (v/v) Tween 20) overnight or for 1 h at RT.

The membranes were incubated with 1/5000 of either Qiagen Penta-His HRP conjugate antibodies or α-FLAG tag HRP conjugated antibodies dilution in the Qiagen blocking solution for 1 hour at RT and subsequently washed 3 times in blocking solution minus antibodies. Membranes were developed using the ECL system (GE Healthcare), exposed to X-ray film for between 5 min and 30 min and developed using an Xograph automatic X-ray film processor.

2.3.4.7   Protein purification

2.3.4.7.1   Purification of *S. coelicolor* NsrR from *E. coli*

Dr Jason Crack or Dr Erin Dodd purified all NsrR protein used in this study.

Native NsrR was overproduced in aerobically grown *E. coli* (BL21 DE3) cultures harbouring pNsrR, as previously described (Tucker et al. 2008). Cell pellets were washed with lysis buffer (50 mM Tris-HCl, 50 mM NaCl, 5% (v/v) glycerol, pH 7.1), transferred to the anaerobic cabinet and stored at -10°C in an anaerobic freezer until required (Belle Technology). Unless otherwise stated, all subsequent purification steps were performed under anaerobic conditions inside an anaerobic cabinet (O_2 < 4 ppm). Cell pellets were resuspended in lysis buffer with the addition of lysozyme (0.4 mg/ml), DNase I (1.3 µg/ml), 2 mM PMSF and 1.3% (v/v) ethanol. The cell suspension was thoroughly homogenised by syringe, removed from the anaerobic cabinet, sonicated twice while on ice, before being returned to the anaerobic cabinet. The cell suspension was transferred to O-ring sealed centrifuge tubes (Nalgene) and centrifuged outside of the cabinet at 627,000 × g for 45 min at 1 °C. The supernatant was passed through a HiTrap DEAE column (2 × 5 ml; GE Healthcare); the eluent was immediately loaded onto a HiTrap heparin column (3 × 5 ml; GE Healthcare), and washed with lysis buffer until $A_{280\,nm}$ ≤ 0.1. The heparin column was then washed with buffer A (50 mM Tris-HCl, 50 mM NaCl, 5% (v/v) glycerol, pH 8.0) and bound proteins were eluted (1 ml/min) using a linear gradient (20 ml) from 10% to 100% (v/v) buffer B (50 mM Tris, 2 M NaCl, 5% (v/v) glycerol, pH 8.0). Fractions (1 ml) containing NsrR, were pooled, diluted 10-fold with lysis buffer, transferred to O-ring sealed centrifuge tubes (Nalgene) and centrifuged outside of the cabinet at 627,000 × g for 30 min at 1°C. The supernatant was passed through a HiTrap DEAE column (5 ml) and immediately loaded onto a HiTrap heparin column (3 × 1 ml). The heparin column was then washed with buffer A containing 3% (v/v) buffer B and eluted using a linear gradient (2 ml) from 3% to 100% (v/v) buffer B. Fractions (1 ml) containing NsrR, were pooled and stored in an anaerobic freezer until needed. Where necessary, gel filtration was carried out under anaerobic conditions using a Sephacryl S-100HR 16/50 column (GE

Healthcare), equilibrated in buffer C, 50 mM Tris, 100 mM NaCl, 5% (v/v) glycerol, pH 8, and a flow rate of 1 ml min^{-1}.

Protein concentrations were determined using the method of Smith et al. (1985) with bovine serum albumin as the standard. The iron and sulfide content of proteins were determined as previously described (Crack et al. 2006) and the [4Fe-4S]$^{2+}$ cluster concentration was determined using $\varepsilon_{406\,nm}$ = 13.30 (±0.19) mM^{-1} cm^{-1}.

2.3.4.7.2 Purification of *S. venezuelae* RsrR from *E. coli*

All purifications of RsrR were carried out anaerobically unless otherwise stated. Protein purification was carried out using a separation funnel, 1 ml Propylene tubes (Qiagen) and Ni-NTA agarose beads. *E. coli* (BL21 or BL21 pLysS) cells containing an RsrR protein expression construct (pJM030), were grown aerobically over night and then sub-cultured 1:100 in fresh LB. The LB was supplemented with 200 µM ferric ammonium citrate and 50 µM (final concentration) to improve [Fe-S] production/incorporation. Cells were grown for 2-3 h until an OD$_{600}$ of 0.5-0.6 was reached. Protein expression was induced by adding IPTG, to a final concentration of 0.5 mM, and cells were grown for another 3 hours before being allowed to rest overnight at room temperature. Cells were harvested the following morning and collected in a 50 ml falcon tube.

Lysis was carried out using a French press (French Pressure Cell Press, Thermo Scientific, 40K cell) at 1000 psi or by freeze thawing pLysE containing cells. Cell debris was removed by centrifuging at 4°C for 30 min at 6010 x g. The supernatant was transferred to a fresh container and 50-100 µl of Ni^{2+} NTA agarose beads (Qiagen) were added per litre of LB used for culturing. Samples were incubated at 4°C mixing slowly by inversion on a vertical rotor for 1 h. The beads were then removed from the sample by centrifugation and immediately transferred to an anaerobic cabinet (O$_2$ < 4 ppm). These where then added to 1 ml propylene tubes (Qiagen) and washed with 20 ml of buffer A (50 mM Tris-HCl, 50 mM NaCl, 5% (v/v) glycerol, pH 8.0). Following washes with buffer A the sample is eluted from the beads using 1-2 ml of buffer B (buffer A + 500 mM imidazole). The eluted protein was buffer exchanged using (PD-10 Desalting Columns, GE Healthcare) following manufacturer's instructions into buffer A.

Apo-RsrR was prepared as with the holo form, followed by incubation with 1mM EDTA overnight.

2.3.4.8 *In vitro* DNA-binding assays

2.3.4.8.1 Electrophoretic mobility shift assay (EMSA).

EMSAs are carried out using non-denaturing PAGE gels. Four gels (7.65%) were poured from a 20 ml total volume containing 5.10 ml acrylamide (30%), 2 ml 10x TBE (89 mM Tris-HCl (pH8), 89 mM boric acid, 2 mM EDTA) 12.9 ml dH$_2$O, 0.2 ml 10% (w/v) APS and 30 µl TEMED. Gels were produced using a Mini Protean III system (BioRad). Before loading, the polyacrylamide gels were pre-run at 30 mA for 2 min.

A nested PCR system was used to amplify target probes. Probe primers were generally designed to have a UniFam or M13 sequence at the 5' end. Following the initial amplification of specific targets, a second amplification step was carried out using 6'Fam labelled primers (JM0017 or JM0062 and JM0063) to add a 6'FAM fluorescent label to the 5' ends, facilitating detection.

Unless otherwise stated, each binding reaction was prepared with ~20 ng 6'FAM labelled DNA probe, 2 µl buffer (100 mM Tris-HCl pH, 600 mM KCl, pH 7.52) X µl protein and Y µl of dH$_2$O (up to a final volume of 20 µl). Samples were incubated on ice for 10 min, 2 µl of loading dye (50% binding buffer, 50% glycerol, 0.03% Bromophenol blue) was added and then each sample was loaded onto gels and run at 200 V for 30 min. Gels were visualized on a molecular imager FX Pro (Bio-Rad) using an excitation wavelength of 488 nm and an emission wavelength of 530 nm.

2.3.4.8.2 DNaseI footprinting

Footprinting was carried out as previously described (Al-Bassam et al. 2014); with the following modifications. Primers were designed to produce PCR products of 120-200 bp in length encompassing predicted binding sites identified either bioinformatically with MEME or by EMSA. Per primer set, one primer end was labeled with γ-^{32}P, 3000Ci/mmol 10mCi/ml (Perkin Elmer) using T4 PNK (NEB) in a 20 µl labeling reaction (2.5 µl primer (10 pmol/µl), 11.5 µl water, 2 µl 10x T4 PNK buffer, 1 µl T4 PNK and 3 µl γ-^{32}P) incubated at 37°C for 2 h then 65°C for 20 min. To this labeling reaction 30 µl of PCR mix was added (2.5 µl second primer (10 pmol/µl), 1 µl template (100 ng/µl), 1 µl 10 mM dNTPs, 10 µl 5x Q5 buffer, 10 µl 5xGC enhancer, 5 µl dH$_2$O, 0.5 µl Q5) and thermal cycling conditions previously optimised using cold reactions were used. The subsequent PCR products were purified using QIAQuick columns (Qiagen) according to manufacturer's instructions. Binding reactions between DNA (30-150,000 cpm) and protein of interest (0-2 µM) were incubated for 30 min at room temperature in 40 µl of reaction buffer (10 mM Tris 54 mM KCl, 0.3%(v/v) glycerol, pH 7.5.), before treatment with 1-10 U DNaseI

(Promega) and 1 µl 100 mM $CaCl_2$ for 10-150 s. To terminate the reactions 140 µl of stop solution (192 mM sodium acetate, 32 mM EDTA, 0.14% (w/v) SDS, 70 µg/ml yeast tRNA) was added and mixed by vortexing. Samples were extracted with 190 µl phenol-chloroform and the DNA containing aqueous phase was ethanol precipitated by addition of 540 µl 96% ethanol. Pellets were dried in a miVac DNA concentrator (GENEVAC Ltd.) and resuspended in 4 µl of loading dye (80% (v/v) formamide, 10 mM NaOH, 1 mM EDTA, 0.1% (w/v) xylene cyanol, 0.1% (w/v) bromophenol blue). A 6% polyacrylamide sequencing gel (Severn Biotech) was loaded with each sample in 1x TBE running buffer. The gel was maintained at $50^{o}C$, 1200 V to ensure uniform DNA separation. Gels were transferred from glass plates to Whatman paper and dried for 30 min using a model 583-gel dryer/HydroTech vacuum pump (BioRad). Labeled DNA was visualised using a phosphoimager plate exposed for 16-24 h and scanned at 635 nM using the purple IP filter on a Typhoon FLA 9500 (GE Healthcare Life Sciences). G+A ladders were produced for each probe according to the Sure Track footprinting kit (Amersham Pharmacia Biotech).

2.3.4.9   Experiments for Next-generation sequencing

2.3.4.9.1   Chromatin-immunoprecipitation sequencing (ChIP-seq)

2.3.4.9.1.1   *S. coelicolor* NsrR

The NsrR ChIP-seq was carried out as described in the Ph.D. Thesis of Felicity J. Knowles (Knowles 2014). MS agar plates were prepared by adding sterile cellophane discs to the top of the agar. The experiment was carried out in parallel using M145 and *ΔnsrR*. A starting inoculum of 1 x 10^8 spores was added to each plate (20 plate/strain) and incubated at $30^{o}C$ for 48 h. Cellophanes were removed, inverted and the mycelium directly exposed to 10 ml 1% (v/v) formaldehyde solution for 20 min at RT. The cellophanes were then transferred to 10 ml of 0.5 M glycine for 5 min. Mycelia were harvested and pooled per experimental strain and washed twice with ice-cold PBS (1x PBS pH 7.4) centrifuging between each at 2,500 x g for 10 min. The subsequent pellets were then resuspended in 1 ml of lysis buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mg/ml lysozyme, 1x protease inhibitor (Roche, cOmplete™, Mini, EDTA-free)) and incubated at $25^{o}C$ for 25 min. To each sample 1 ml of IP buffer (100 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.5% (v/v) Triton X-100, 0.1% (w/v) SDS, 1x protease inhibitor (Roche)) was added before placing on ice to cool 2 min. Sonication was carried out at 50 Hz for 15 s and placed on ice for 45 s for 7 cycles followed by centrifugation at 15,120 x g for 10 min at $4^{o}C$. Supernatants were removed to fresh tubes and centrifugation repeated. Samples

were then split in half and 25 µl of sample was used for a total DNA sequencing control and 1 ml used for immunoprecipitation with the excess being stored at -20°C.

Preparation of 50% protein A sepharose beads was produced using 0.125 g of beads and washing them in ½ IP buffer 4 times for 15 min centrifuging between each at 2,500 x g for 3 min. Beads swell to 4x original weight.

The 1 ml of IP sample was pre-cleared using 100 µl of equilibrated 50% protein A sepharose beads and incubated at 4°C for 1 h on a rotating wheel before centrifuging for 15 min at 15,120 g for 15 min. Supernatants were then incubated with 100 µl (10 mg) of α-FLAG antibodies overnight at 4°C on a rotating wheel before adding 100 µl of equilibrated protein A sepharose beads and incubating for a further 4 h at 4°C. The sepharose α-FLAG NsrR 3x FLAG DNA complex was then cleared from the sample by centrifugation at 1,100 x g for 5 min before washing twice with 1 ml ½ IP buffer then twice more with 1 ml of IP buffer each taking 15 min and centrifuging for 1, 100 x g for 5 min between washes. To the bead pellet, 150 µl elution buffer (50 mM Tris-HCl pH7.6, 10mM EDTA, 1% SDS) was added and incubated at 65°C overnight, inverted 7 times then centrifuged 5 min at 15,120 x g. The supernatant was removed and stored and 50 µl of TE buffer (10 mM Tris-HCl pH 7.8, 1 mM EDTA) added to the beads that were incubated at 65°C for 5 min then centrifuged at 15,120 x g. The TE buffer supernatant and the previous 150 µl of supernatant were then pulled and treated with 2 µl of 10 mg/ml proteinase K at 55°C for 1.5 h. The samples were then phenol/chloroform extracted by adding 200 µl of phenol/chloroform, vortexing samples for 3 min, centrifuging for 3 min at 15,120 x g then removing the upper aqueous phase containing the purified DNA. The organic phase was re-extracted with 100 µl of TE buffer. The extractions were pulled and purified using a QIAquick PCR purification kit and eluted twice with the same 50 µl of H_2O (Sigma). DNA was quantified using a NanoDrop.

2.3.4.9.1.2 *S. venezualae* RsrR

A spore inoculum (~5-10 µl of $1x10^8$ spores) sufficient to reach an OD_{600} of 0.35 after 8 h of growth was added to 35 ml of MYM-tap media in 250 ml-flasks containing springs. Following growth to the chosen time point, the entire content of the flask was transferred to a 50 ml Falcon tube for crosslinking, which was carried out by incubation at 30°C for 30 mins with 1% (v/v) final concentration of formaldehyde. Crosslinking was quenched by incubation at room temperature with glycine (final concentration of 125 mM). Mycelium was harvested by centrifugation 3,345 x g at 4°C for 10 min and washed twice with ice-cold PBS before transfer to a 2 ml centrifuge tube. Pellets were resuspended in 0.75 ml lysis buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mg/ml lysozyme, 1x protease

inhibitor (Roche, cOmplete™, Mini, EDTA-free)) and incubate at 37°C for 10-25 mins. Then 0.75 ml 1x IP buffer (100 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.5% (v/v) Triton X-100, 0.1% (w/v) SDS, 1x protease inhibitor (Roche)) was added and samples mixed by pipetting up and down. Samples were sonicated 7x at 50Hz, 10 s/cycle with 1 min incubation on ice after each cycle. DNA fragmentation was checked by agarose gel electrophoresis following phenol extraction of 25 µl of the crude lysate mixed with 75 µl of TE buffer with 100-200 µl of phenol/chloroform. Contaminating RNA was removed with 2 µl RNase (1mg/ml) added to extracted DNA followed by an incubation for 30 min at 37°C. A smear of DNA from 200 to 1000 bp with the majority of DNA 200-400 bp was visible. Crude lysate was centrifuged at 21,130 x g for 15 min at 4°C to clear the sample of cell debris. M2 affinity beads (Sigma-Aldrich, ANTI-FLAG M2 Affinity Gel) were prepared by washing in ½IP buffer following manufacturer's instructions. The cleared lysate was incubated with 40 µl of washed M2 beads and incubated for 4 h at 4°C in a vertical rotor. The lysate was removed and the beads pooled into one 1.5 microfuge tube and washed in ½ IP buffer. The beads were transferred to a fresh microfuge tube and washed a further 3 times removing as much buffer as possible without disturbing the beads. The DNA-protein complex was eluted from the beads with 100 µl elution buffer (50 mM Tris-HCl pH7.6, 10mM EDTA, 1% SDS) by incubating at 65°C overnight. Removing the ~100 µl elution buffer, an extra 50 µl of elution buffer was added and further incubated at 65°C for 5 min. To extract the DNA, 2 µl proteinase K (10 mg/ml) was added to the 150 µl eluate and incubated 1.5 h at 55°C. To the reaction 150 µl phenol-chloroform was added. Samples were vortexed and centrifuged at full speed for 10 min. The aqueous layer was extracted and purified using the Qiaquick column from Qiagen with a final elution using 50 µl EB buffer (Qiagen).

The concentration of samples were determined using Quant-iT™ PicoGreen ® dsDNA Reagent (Invitrogen) or equivalent kit or by NanoDrop measurement. Sequencing was carried out by GATC using an Illumina's HiSeq 2500 platform with 50 bp reads. ChIP-seq data has been deposited with the Geo accession number GSE81073.

2.3.4.9.2 Data analysis

2.3.4.9.2.1 Sequence file data processing (fastq – wig)

The analysis pipeline commands used in this study were carried out through the command line terminal of either the Unix-based operating system Ubuntu (V12.04 LTS) or Mac OSX. Once all data and reference files are acquired, the pipeline can be summarised into 8 steps (stating the software, the version and the commands used): (1) Building/indexing the reference genome (Bowtie2, version 2.1.0, bowtie2-build (Langmead

& Salzberg 2012) (2) aligning raw sequencing reads to reference genome (Bowtie2, bowtie2), (3) this step was primarily necessary for d/RNA-seq samples. We split the sequence file into strand specific information (Perl, version 5.14.2, sam_split_pos_neg.pl). (4) Conversion of file type from a human readable .sam file to a binary .bam file, reducing the file size substantially (samtools, version 0.1.18, view (Li et al. 2009) (5) ordering the reads from the 1st to last base in the genome (samtools, sort), (6) producing a coverage file (.bedgraph) with the frequency in which each base position is sequenced (bedtools, version 2.17.0, genomeCoverageBed (Quinlan & Hall 2010), (7) removing all additional data and leaving a small file containing the frequency each position was sequenced (Perl, version 5.14.2, bedgraph2wig.pl (script below)) and finally (8) the . wig files each have a header added (example below), with the sample name changed each time it is used, to allow for visualisation in the Integrated Genome Bowser (IGB) software (IGB, version 6.2.2 - (Nicol et al. 2009)). All of the individual tools used are freely available and have full documentation.

2.3.4.9.2.2 Wig header:

----

track   type=wiggle_0   name=WT_18h   graphType=points   visibility=full color=168,130,88   fixedStep   chrom=Sven   start=1 step=1 span=1

2.3.4.9.2.3 Scripts

2.3.4.9.2.3.1 sam_split_pos_neg.pl

---Start   script---

```
use strict;
my $file=$ARGV[0];

open (INFIL, "<$file");
open (POS, ">$file.pos.sam");
open (NEG, ">$file.neg.sam");

while (<INFIL>){
chomp;
my @ar = split (/\t/, $_);
```

83 of 291

```perl
if ($ar[0] =~ m/^@/ or $ar[1] == 16){
print NEG "$_\n";
}
if ($ar[0] =~ m/^@/ or $ar[1] == 0){
print POS "$_\n";
}
}
close (INFIL); close (POS); close (NEG);
exit;
```

--End script--

2.3.4.9.2.3.1.1 bedgraph2wig.pl

---Start script---

=for comment

----

this is an example of a bedfile format output from the tools "genomeCoverageBed" in bedtools-

----

```
  -some weird reference        base     base
                   postion  position  frequency
                    from      to
i|408675720|ref|NC_018750.1|   0      1      7
gi|408675720|ref|NC_018750.1|  1      2      11
gi|408675720|ref|NC_018750.1|  2      3      12
gi|408675720|ref|NC_018750.1|  3      5      13
gi|408675720|ref|NC_018750.1|  5      7      14
gi|408675720|ref|NC_018750.1|  7      13     15
gi|408675720|ref|NC_018750.1|  13     14     16
gi|408675720|ref|NC_018750.1|  14     15     18
=cut
```

```perl
use strict;
use warnings;
my $file = <STDIN>;
my @array;
my $from;
my $to;
my $cc=1;
open (BED, "$file");

while (<BED>){
chomp $_;
@array = split (/\t/, $_);
```

```perl
$from = $array[1];
$to = $array[2];
my $freq= $array[3];
my $count = 0;
my $dif = $to-$from;
while ($dif-$count>0){
$count++;
my $tt = $from + $count;
print "$freq\n";


}
:
```
--End script--

2.3.4.9.2.3.2 Producing IGB reference genome

The below script was used to convert a .gbk genome file to a .bed file which can be used in IGB to view genome annotations:

--Start script--

```perl
#perl script to make a bed file from a gbk file. The bed output is
#used in IGV sortware to view ChIP-seq data.

use strict;
use Bio::SeqIO;
my $gbk = @ARGV[0];
my $in = Bio::SeqIO->new(-file=>"$gbk");
my $seq = $in->next_seq();

foreach my $feat($seq->get_SeqFeatures()){
```
--End script--

2.3.4.9.3 Peak calling

Peak calling was carried out using CLC genomics workbench 8 (Qiagen) using the default settings with cut-offs as described in the appropriate chapter. In addition to this analysis, manual visual inspection of identified peaks was carried out to define the target lists that were used for subsequent analysis.

2.3.4.9.4 dRNA-seq analysis

RNA-isolation was carried out following section (2.2.1.4), sequencing and post isolation sample processing was carried out by vertis Biotechnologie. In brief, total RNA was first ribodepelted, Tex (5'P-dependant terminator exonuclease) treated, split into two samples, TAP (Tobacco Acid Pyrophosphatase) treated and RNA adaptors ligated to the 5' end. cDNA was synthesised and sequencing carried out on an Illumina NextSeq 500 sequencing platform with 75 bp reads. Library construction and cDNA synthesis was carried out using commercially available kits. Data analysis was carried out as described in section (2.2.4.9.2) for visualisation. Expression profiling was carried out using the Tuxedo suite (Trapnell et al. 2012). Default settings were used for the TopHat analysis mapping the genome to NC_018750.1 genome. CuffDiff analysis was carried out using a default settings and a readcut off limit of ~112 reads per loci (achieved using the –c operation set to 500). TSSAR webservice was used for the dRNA transcription start site

(TSS) analysis (Amman et al. 2014). In addition to these a manual, by eye, processing approach was carried out.

2.3.4.9.5 Online data deposits

The RsrR data has been deposited online online with the GEO superSeries accession number GSE81105.

3   NsrR

3.1   Chapter summary

The work carried out in this chapter was to further our understanding of the *Streptomyces coelicolor* NsrR regulon. Some the research included below was ultimately published (Crack et al. 2016; Crack et al. 2015) along side biochemical studies of NsrR, its cofactor and its interaction with NO.

Prior to this work a bioinformatics approach was carried out to identify the NsrR regulon utilising binding site data available for the two known NsrR target promoters corresponding to *hmpA1* and *hmpA2* (Tucker et al. 2008; Tucker et al. 2010), from now on denoted as *hmpA1p* and *hmpA2p* respectively. Scanning the genome using this information resulted in a list of 322 putative NsrR targets. To experimentally validate these predictions ChIP-seq and *in vitro* DNA-binding studies were carried out. The ChIP-seq experiment (Knowles 2014) was carried out in an M145 *nsrRflp* background, the *nsrR* disruption being in frame with a small "scar" where the gene was removed, with a second copy of *nsrR* introduced at the ΦC31 integration site, containing a 3x FLAG-tag. The data acquired was analysed as part of this work where 3 targets were identified, both the known *hmpA1p* and *hmpA2p* targets with the only additional and novel target being the *nsrR* promoter (*nsrRp*). DNaseI footprinting and EMSA analysis was carried out on the 3 defined targets and binding sites for each identified corresponding to an 11-1-11 bp imperfect-inverted repeat sequence. Additional predicted targets were tested by EMSA, however, NsrR was either unable to bind or only bind weakly to these probes before non-specific binding was observed. Disruption mutants, single double and triple of each combination of *nsrR, hmpA1* and *hmpA2,* were produced and showed no observable phenotype under normal laboratory conditions. In this chapter we show that *S. coelicolor* NsrR has a small regulon of targets that are predicted to be specifically involved in NO-detoxification.

3.2   Results

3.2.1   The NsrR regulon: a disconnect of experiment and prediction

3.2.1.1   ChIP-seq analysis identifies a small, NO specialised, regulon

ChIP-seq (Chromatin-immunoprecipitation followed by sequencing) is a method to define the genome wide DNA target sequences of DNA-binding proteins to facilitate identification

of regulons. The method relies on having a DNA-binding protein which you can specifically detect using antibodies for either the protein or an added tag. Cells are grown to a chosen point where they are exposed to formaldehyde, covalently crosslinking cellular components, including DNA to DNA-binding proteins. Cells are lysed enzymatically, then sonicated to fragment the genome. Sonication is carried out until genome fragments are of the desired size (usually averaging 400-500 bp in length). The sonicated lysate is then exposed to protein specific antibodies and the DNA/Protein/antibody complex is extracted. The DNA is isolated from the complex by proteinase treatment and phenol/chloroform extraction. Next generation sequencing (NGS) of the DNA results in reads that are mapped to the genome, regions bound by the protein of interest are present in a higher abundance and as a result are sequenced more frequently and when mapped results in areas of enrichment called "peaks". When visualised these peaks correspond to sites where the protein of interest has bound specifically to DNA.

ChIP-seq, carried out by Dr Felicity Knowles, was used to identify the NsrR regulon. The experimental strain (*nsrRflp* pN5112) was produced by Dr Felicity Knowles containing an in-frame deletion of *nsrR* and pSET152 containing a C-terminally 3x FLAG tagged copy of *nsrR* under the control of its own promoter (Crack et al. 2015). ChIP-Seq required either the production of polyclonal antibodies specific to NsrR or to use a commercially available antibody specific for an epitope (e.g. an affinity tag) attached to the protein, in our case we chose a 3x FLAG tag (DYKDHDGDYKDHDIDYKDDDDK). The ChIP-seq experiment was carried out using WT M145 as a control and *nsrRflp* pN5112 as the experimental strain. Both strains were grown on cellophane discs on MS media at 30°C for 48 h with samples processed as described in section 2.3.4.9.1.1. Following ChIP-seq the sequencing files were processed as described in section 2.2.4.9.2 and visualised using the integrated Genome Browser IGB (Figure 3.1a-b). We were only able to identify three target genes above a 2 fold enrichment: *hmpA1* and *hmpA*2, both of which are known targets and *nsrR* itself*,* which was the only novel target identified from the data. Although not what we expected, based on the predictions, all three targets identified are specifically linked to NO detoxification. Both *hmpA* genes are involved in the detoxification of NO into nitrate and NsrR is the sensory protein repressing their production under non-stress growth conditions. From these results we were able to identify the NsrR binding sequence at the *nsrRp* (AAGGCGAACCTAGCATGCGCATT) using the previously identified *hmpA* sequences for *hmpA1p* (AACACGAATATCATCTACCAATT) and for *hmpA2p* AACAAGCATCTGAGATCCCAGTT) to produce a consensus binding sequence (Figure 3.1c).
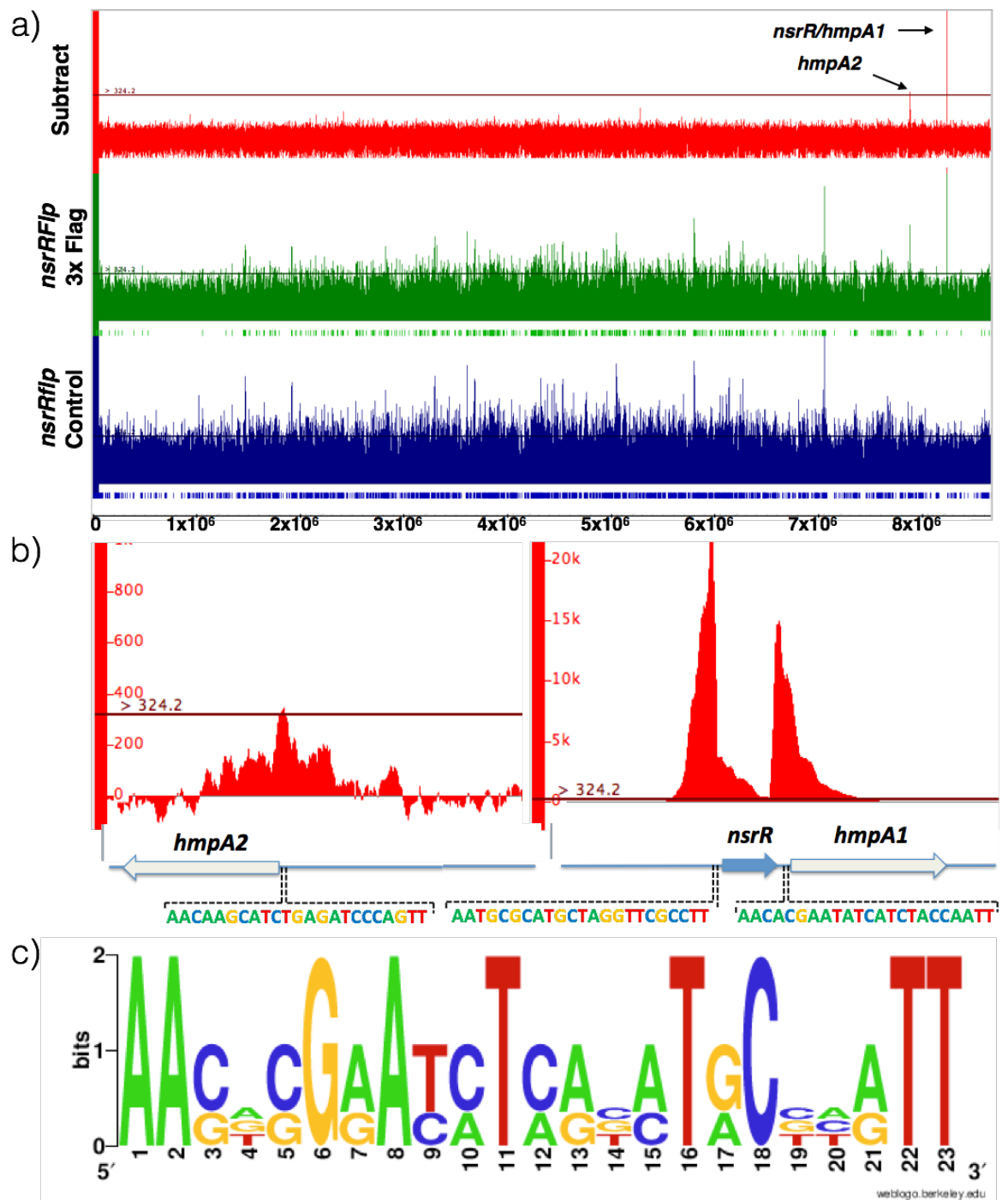
Figure 3-1 NsrR ChIP-seq results. (a) Shows the entire ChIP-seq data set. The *nsrRFlp* mutant (blue), the experimental strain *nsrRFlp* 3x FLAG (green) and the *nsrRFlp* subtracted from *nsrRFlp* 3x FLAG data (red) are each shown with number of reads on the y-axis and genomic position on the x-axis. A 2-fold cut off at 324.2 reads indicates only these 3 targets being above the threshold. (b) Shows the *hmpA2* and *nsrR/hmpA1* peaks with the location of each predicted binding site indicated below. (c) Shows the weblogo derived binding sequence from the 3 target binding sites orientated in the 5'-3' direction in regards to the target gene.

3.2.1.2 Refinement of the NsrR binding site, an 11 bp inverted repeat sequence

Results acquired from the ChIP-seq data allowed us to identify three NsrR target sites within the genome. The number of targets identified is more than 100-fold less than those predicted (322) bioinformatically using the *hmpA1*p and *hmpA2p* NsrR binding sites (Tucker et al. 2010). This was concerning and led us to further refine the NsrR binding site in *S. coelicolor* using DNaseI footprinting and then test some of the predicted binding sequences by EMSAs and test the *hmpA1p* binding site using short probes for the binding sites from *S. coelicolor*, *Bacillus subtilis* and *E. coli*.

We first produced ^{32}P labelled DNA probes, which allowed for visualisation of the DNA using a phosphoimaging plate, using primers JTM0005-10, which carried the promoter regions of *nsrR* (*nsrRp*), *hmpA1 (hmpA1p)* and *hmpA2 (hmpA2p)* including their NsrR binding sites. Optimisation of the experiment was carried out primarily using the *nsrRp* probe and then these conditions were used for all three probes. Optimisation of the DNaseI digestion time of the probe resulted in a very clear protected region, which spans the predicted NsrR binding site (Figure 3.2a). The NsrR binding site was further confirmed and identified for each of the promoter probes by incubating reactions with increasing concentrations of NsrR (Figure 3.2b). From these two experiments we see a protected region spanning the predicted NsrR binding sites (*nsrRp* CA<u>AATGCGCATGCTAGGTTCGCCTTT</u>ACCCGG, *hmpA1p* CTGTGGCCTAA<u>AACACGA ATATCATCTACCAATT</u>AAGAG, *hmpA2p* TCGGA<u>AAACAAGCATCTGAGATCCAGTT</u>CG GAG, the consensus binding site is underlined) at the promoters of these three genes.

To further analyse the *S. coelicolor* NsrR binding site we produced EMSA probes using a nested primer system incorporating short sequences to be tested for NsrR binding, summarised in Table 3.1. Firstly, probes were designed to determine the outer edge of the *hmpA1* promoter NsrR binding sites. JM0064 consists of the *hmpA1* sequence used previously to examine NsrR binding using analytical ultracentrifugation (Tucker et al. 2008). JM0065 is the full, predicted, *hmpA1*p site. JM0068 is the *hmpA1*p with the conserved double AA and double TT removed from the ends. We also wanted to determine if *S. coelicolor* NsrR can bind to the experimentally derived *E. coli* and *B. subtilis* NsrR binding sites at their respective *hmpA* promoters and if a binding site remained functional when made preferentially GC rich (JM0069-71). For probe JM0071, when a base was wholly conserved between the three sites, we retained this regardless of AT or GC nature. However if the site was a G or C in any of the three probes, this was selected preferentially over an A or T.
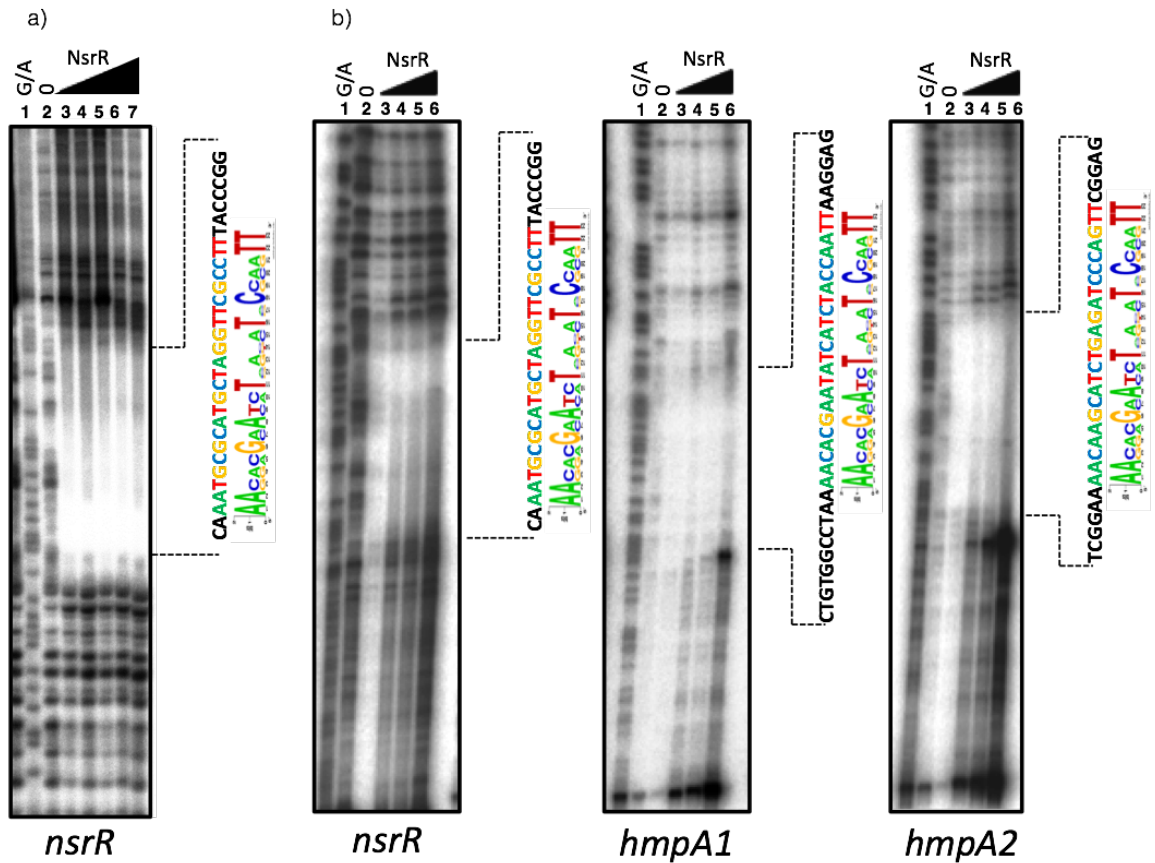
Figure 3-2 NsrR DNaseI footprinting. NsrR DNaseI footprinting results showing the binding sites of NsrR within the *nsrR, hmpA1* and *hmpA2* promoters modified from Crack et al. 2015. a) Footprinting experiment using ~30k cpm of probe and 2 µM NsrR with varying digestion times (Lane 2-7, 10 s, 1 min, 2.5 min, 5 min, 7.5 min and 10 min respectively). b) *nsrR, hmpA1* and *hmpA2* promoter (~30k cpm of each) with increasing concentrations of purified NsrR added and digestion times extended appropriately (lane 2-6, DNA only control, 0.1 µM, 0.25 µM, 1 µM and 2 µM digested for 10 s, 45 s, 90 s, 120 s, and 150 s respectively). In each, lane 1 contains G+A ladder for analysis.

The concentration of NsrR used in each EMSA throughout this chapter corresponds to the approximate molar ratio of [Fe:S]:DNA required to completely and specifically shift each of the target probes, 2.5:1.0, 5.0:1.0, and 8.0:1.0 for *hmpA1, nsrR and hmpA2* probes respectively, in addition to a DNA only control. These ratios were selected based on the EMSA results in our publication Crack et al., 2015 and shown in Figure 3.3. addtionally we were able to show that neither [2Fe-2S] or apo NsrR were capable of shifting target DNA, only Holo [4Fe-4S] NsrR is capable of this as indicated in Figure 3.4 and Figure 3.5 respectively. A ratio of 16.0:1.0 was also used as a higher limit for EMSA's in section 3.2.1.3.
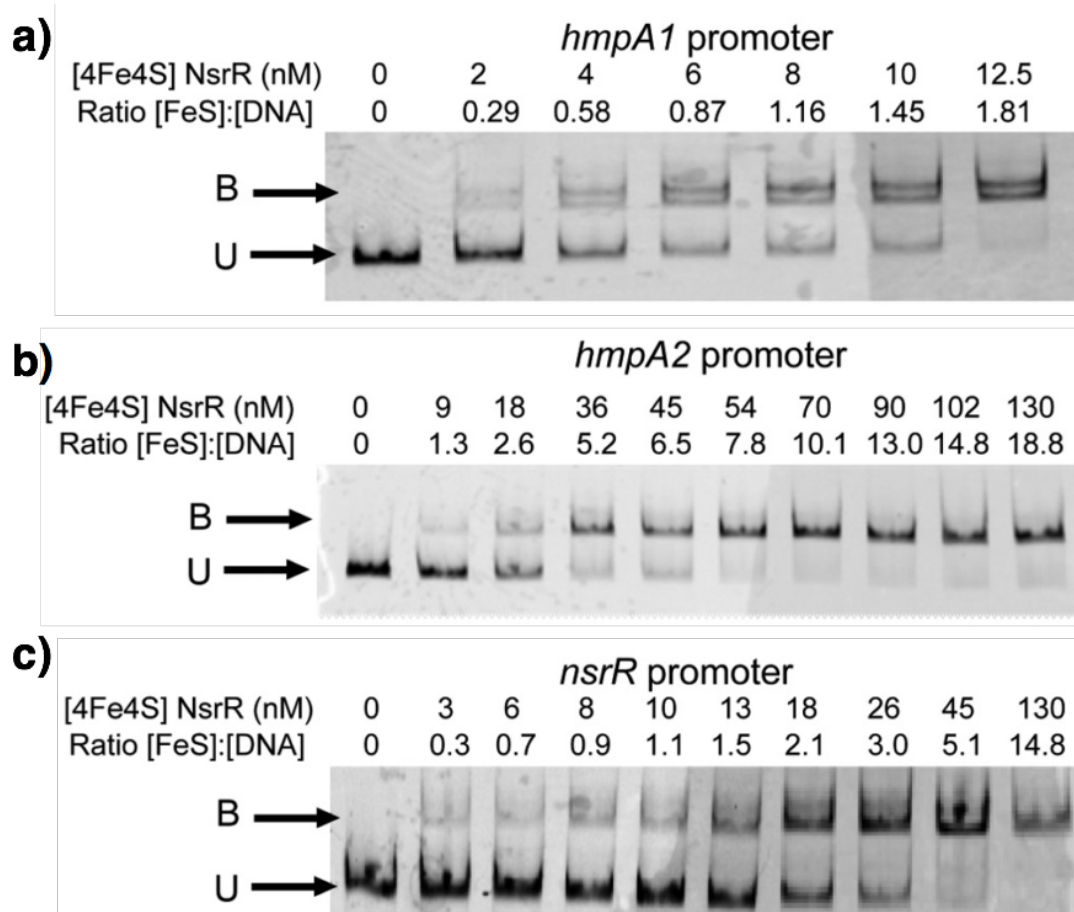
Figure 3-3 NsrR [4Fe-4S] EMSAs. Taken from Crack et al., 2015. EMSAs were carried out using holo NsrR ([4Fe-4S]) using (a) *hmpA1p* (6.9 nM), (b) *hmpA2p* (6.9 nM) and (c) *nsrRp* (8.8 nM) DNA probes. B = Bound, U = Unbound DNA probe. Each illustrates the approximate ratio required to shift each probe to near completion. These ratios were used for down stream experiments in this thesis.

a

hmpA1 promoter

[4Fe4S] NsrR (nM)	-	-	-	-	-	-	-	-	-	130
[2Fe-2S] NsrR (nM)	0	16	32	48	64	80	96	112	130	-
Ratio [FeS]:[DNA]	-	1.4	2.9	4.3	5.8	7.2	8.7	10.1	11.7	11.7

B →

U →

b

hmpA2 promoter

[4Fe4S] NsrR (nM)	-	-	-	-	-	-	-	-	-	130
[2Fe-2S] NsrR (nM)	0	16	32	48	64	80	96	112	130	-
Ratio [FeS]:[DNA]	-	2.4	4.7	7.1	9.4	11.8	14.1	16.5	19.1	19.1

B →

U →

c

nsrR promoter

[4Fe4S] NsrR (nM)	-	-	-	-	-	-	-	-	-	130
[2Fe-2S] NsrR (nM)	0	16	32	48	64	80	96	112	130	-
Ratio [FeS]:[DNA]	-	3.0	6.0	9.1	12.1	15.1	18.1	21.1	24.5	24.5

B →

U →
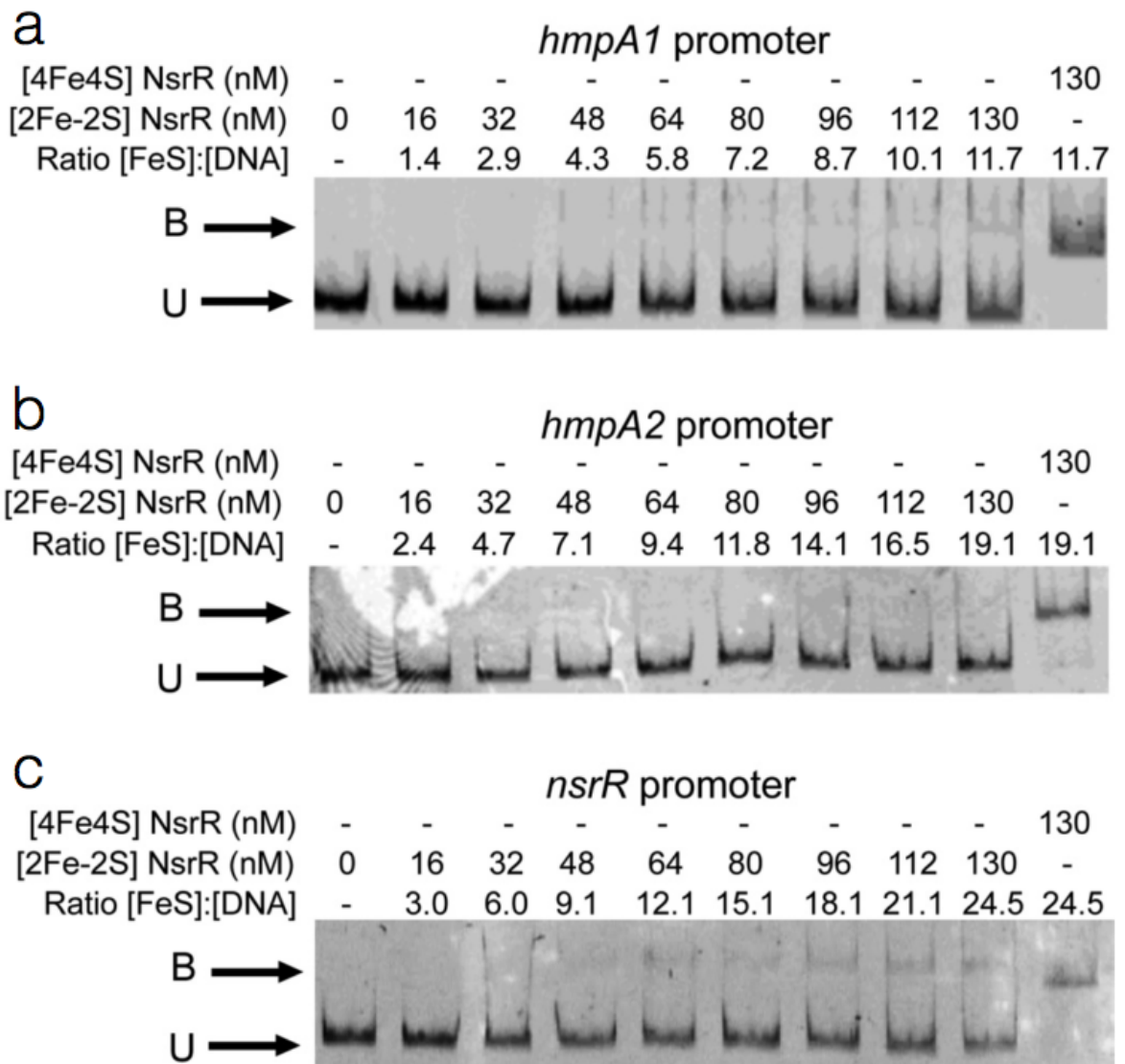
Figure 3-4 NsrR [2Fe-2S] EMSAs. Taken from Crack et al., 2015. EMSAs were carried out using NsrR [2Fe-2S] and NsrR ([4Fe-4S]) as a control using (a) *hmpA1p* (11.1 nM), (b) *hmpA2p* (6.8 nM) and (c) *nsrRp* (5.3 nM) DNA probes. Each illustrates the concentration of each form used (nM) and the ratio of [Fe-S]:[DNA].

a

Holo (nM)	-	-	-	-	-	-	-	-	-	130
Apo (nM)	0	16	32	48	64	80	96	112	130	0
Ratio [Apo]:[DNA]	0	1.8	3.6	5.5	7.3	9.1	10.9	12.7	14.8	

B →
U →

b

Holo (nM)	-	-	-	-	-	-	-	-	-	130
Apo (nM)	0	16	32	48	64	80	96	112	130	0
Ratio [Apo]:[DNA]	0	2.4	4.7	7.1	9.4	11.8	14.1	16.5	19.1	

B →
U →

c

Holo (nM)	-	-	-	-	-	-	-	-	-	130
Apo (nM)	0	16	32	48	64	80	96	112	130	0
Ratio [Apo]:[DNA]	0	1.8	3.6	5.5	7.3	9.1	10.9	12.7	14.8	

B →
U →
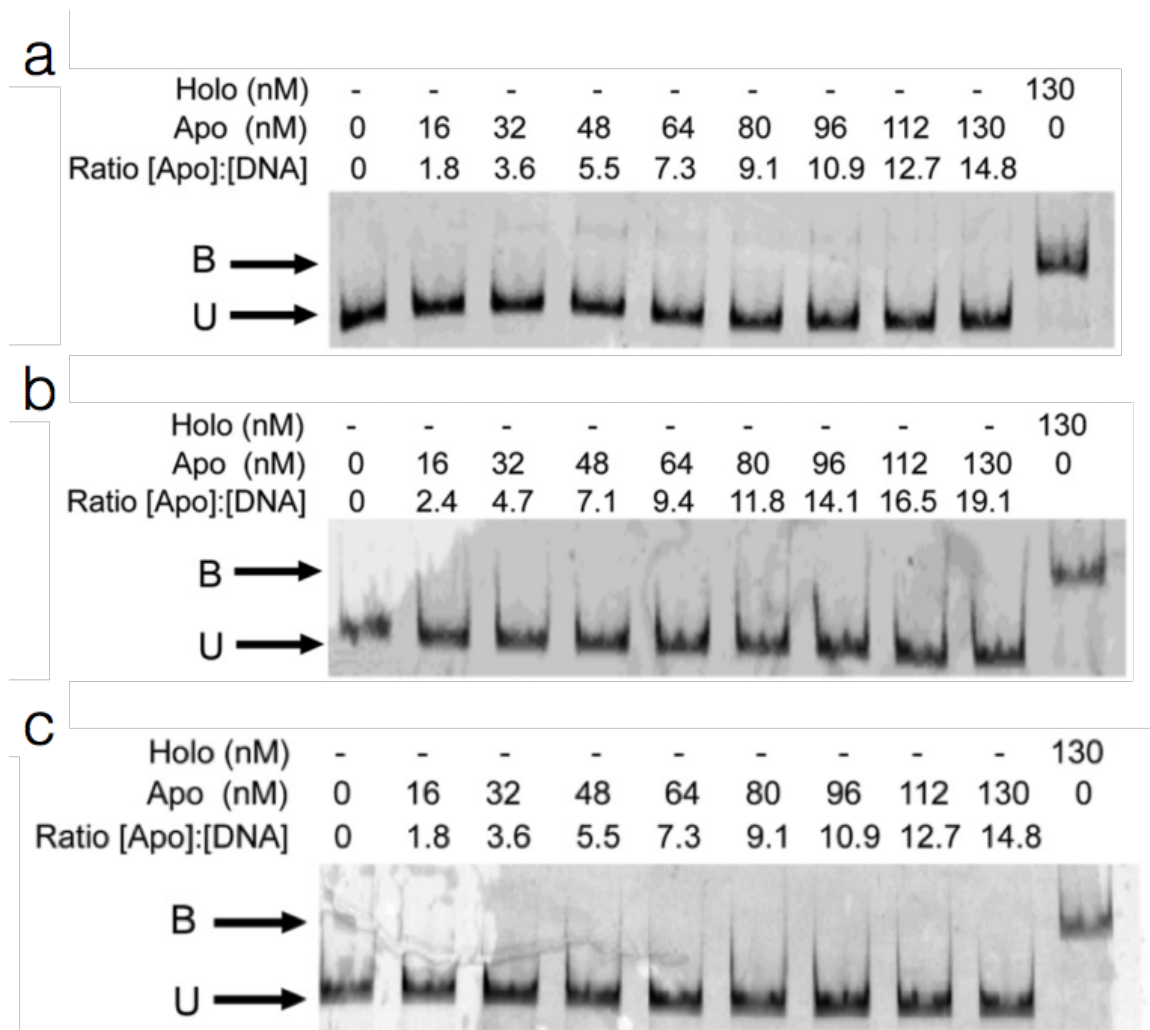
Figure 3-5 NsrR Holo vs. Apo EMSAs. Taken from Crack et al., 2015. EMSAs were carried out using holo NsrR ([4Fe-4S]) using (a) *hmpA1p* (6.9 nM), (b) *hmpA2p* (8.8 nM) and (c) *nsrRp* (8.8 nM) DNA probes. Each illustrates the concentration of each form used (nM) and the ratio of [Fe-S]:[DNA].

EMSA reactions were carried out using these probes and the results are summarised in Table 3.1 and shown in Figure 3.6. NsrR was able to bind to JM0064, our positive control and JM0065 the refined probe with the conserved AA and TT at the ends. However, NsrR did not bind to JM0068-70. This highlights the importance of the AA/TT at the ends of the binding sites, following the removal of these the protein is no longer able to bind. A clear difference in sequence binding between the *S. coelicolor, E. coli* and *B. subtilis* proteins and their target sequences. We observed that NsrR had a lower affinity for Short *hmpA1* (JM0068) which could be due to the loss of non-essential but important bases flanking the binding site. Additionally, due to ScoNsrR having a [4Fe-4S] cluster and only sharing between 30-40% amino acid sequence identity with *E. coli* and *B. subtilis* which [2Fe-2S] cluster, it is likely that the binding architecture is substantially different and as a result contribute to an inability to bind the ScoNsrR binding sequence.

Promoter	Sequence	NsrR binding	Primer
hmpA1 long	CTAAAACACGAATATCATCTACCAATTAAG	Y	JM0064
hmpA1	AACACGAATATCATCTACCAATT	Y	JM0065
Short *hmpA1*	CACGAATATCATCTACCAG	N	JM0068
Bsu *hmpA* 17	AAGATCATGTATTTTAAAGATATATTTTA	N	JM0069
Eco *hmpA*	ATAAGATGCATTTGAGATACATCAA	N	JM0070
GC rich	AACGCGCATCTGAGATGCGCGTT	N	JM0071
Consensus	AACACGAATCTNANATNCCAATT	-	-

Table 3-1 A summary of the short EMSA reactions. The EMSA column indicates if a shift was detected (Y = yes, N = no).
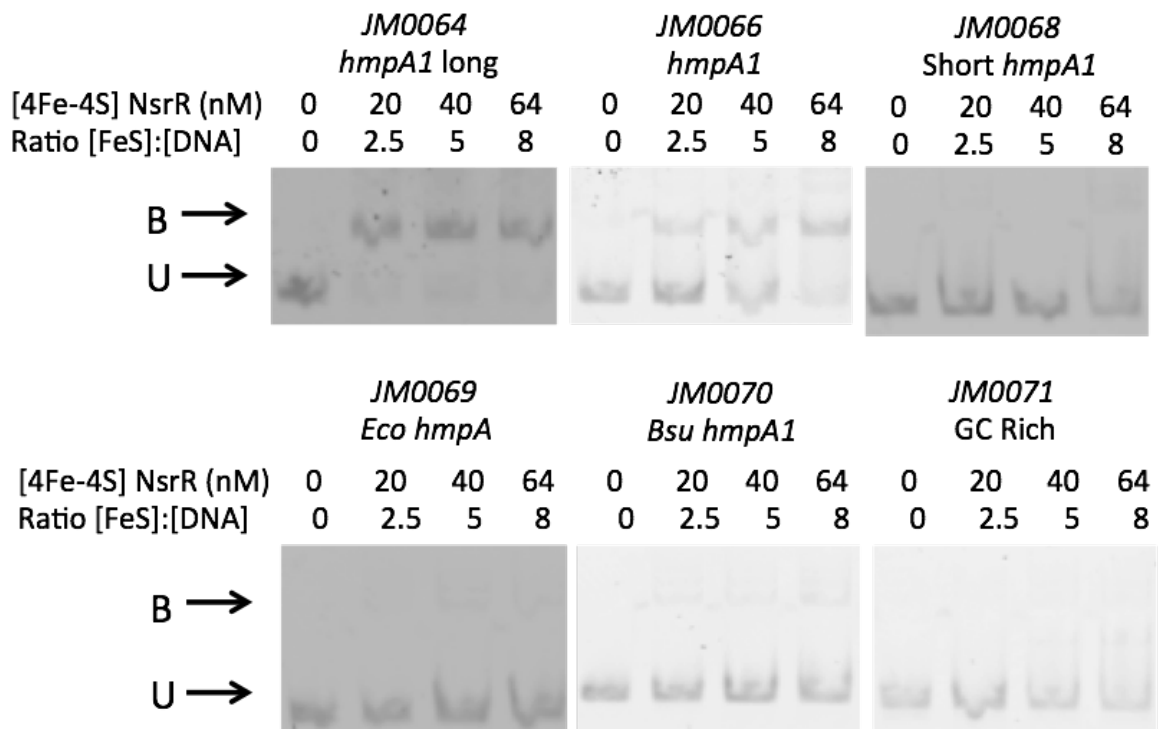
Figure 3-6 Short probe EMSA reactions. Using short probes JM0064-JM0071 showing that ScoNsrR can bind the Sco *nsrR hmpA1* and *hmpA2* sequence very specifically (taken from Crack et al. 2015). Ratio of cluster to DNA was determined using 8 nM DNA probe in each case. Black arrows show unbound (U) and Bound (B) DNA to NsrR.

3.2.1.3 Predictions dramatically overestimate the NsrR regulon

Predictions of the NsrR regulon were carried out (Tucker et al. 2010) using the computationally predicted NsrR binding site for *Streptomyces* / *Bacillus* (Rodionov et al. 2005) to search the 70 bp of DNA upstream of every coding sequence in the *S. coelicolor* genome. The prediction correctly identified the *hmpA1* and *hmpA2* promoters as targets but missed *nsrR* and dramatically overestimated (in regards to the ChIP-seq results) the number of regulon targets, estimating 322 targets (Supplementary data S3.1). The ChIP-seq results show clearly that both *nsrR* and *hmpA1* are highly enriched but *hmpA2* has only an ~2-fold enrichment. Recent results suggest that *hmpA2* binding by NsrR is first to be relieved at only 2 NO molecules per [4Fe-4S] cluster suggesting it has lower affinity for *hmpA2* and this is first to be induced upon exposure to NO (Crack et al. 2016). On inspection of the three DNA-binding sequences we see that the *hmpA1/2p* NsrR sites are far more similar to each other, than to the *nsrRp,* which is likely why the target was missed in the bioinformatics approach.

All three ChIP-seq identified targets are bound by holo NsrR *in vitro* and the results of the EMSAs are summarised in Table 3.2. Ultimately we chose three major paths to progress.

1) we checked key targets based on annotation, including another putative *hmpA* homologue (*hmpA3*) (Table 3.2a and Figure 3.7). (2) We narrowed down the list of 322 targets to those with multiple predicted sites (Table 3.2b and Figure 3.8). (3) Following the work by Partridge and colleagues (Partridge et al. 2009), where they identified NsrR half site targets in *E. coli*, we produced our own half site sequence and scanned the genome using virtual footprinting tool (Münch et al. 2005). The resulting full list of targets identified can be found in Supplementary data S3.2 with those checked in Table 3.2c and Figure 3.9.

During the experimentation we observed that many of the target probes could be bound non-specifically. We define non-specific binding based on the appearance of the EMSA shift. Typically this appears as smearing with much of the sample residing in the well (likely as a huge aggregated protein/DNA complex). We define specific binding as a protein/DNA band, which completely shifts before the onset of non-specific binding. This characterisation is similar to work that has been carried out before (Yamazaki et al. 2000; Kaiser & Stoddard 2011).

Several of the targets (*sco3773*, *sco0447 sco0622, sco1434 sco2610 sco7459)* appear to form partial shifts before the onset of what we have characterised as non-specific binding. Non-specific binding occurs inconsistently between the samples and is likely due to innate differences in the DNA and its overall fold/interaction as part of the protein/DNA aggregate. The targets that shift, beyond *nsrRp, hmpA1p* and *hmpA2p* do so in a non-specific manor.

Promoter	Description	EMSA shift	Primers JM
a			
sco7428	*hmpA1*, positive control	Yes	JM0015-16
sco0103	Putative *hmpA3*	No	JM0073-74
b			
sco3773	Putative *lysR* family transcriptional regulatory protein	No	JM0024-25
sco1447	Putative ROK-family transcriptional regulatory protein	No	JM0026-27
sco2014	Pyruvate kinase	No	
sco6108	*fusH,* esterase	No	JM0022-23
c			
sco0166	Possible regulator, similar to pyoverdine regulator, PvdS	No	JM0028-29
sco0447	Putative MarR-family regulatory protein	No	JM0030-31
sco0622	Putative TetR-family transcriptional regulator	No	JM0032-33
sco1343	Uracil-DNA glycosylase	No	JM0034-35
sco1434	Putative CbxX/CfqX family protein	No	JM0036-37
sco1570	Argininosuccinate lyase	No	JM0038-39
sco1663	Putative cysteinyl-tRNA synthetase	No	JM0040-41
sco2494	Putative pyruvate phosphate di-kinase	No	JM0042-43
sco2610	Rod shape-determining protein	No	JM0044-45
sco3485	Putative LacI-family transcriptional regulator	No	JM0046-47
sco4908	Putative RNA polymerase sigma factor	No	JM0048-49
sco5085	Actinorhodin cluster activator protein	No	JM0050-51
sco6535	Conserved hypothetical protein	No	JM0052-53
sco7168	Putative GntR-family transcriptional regulator	No	JM0054-55
sco7459	Putative ABC transporter, ATP-binding component	No	JM0056-57
sco7705	Putative oxidoreductase	No	JM0058-59

Table 3-2 All bioinformatics targets predicted and tested for NsrR binding by EMSA. a) *hmpA1 and hmpA3* genes chosen as a positive control and test respectively. (b) genes with multiple 11-1-11 bp binding sites within their promoter sequence from the Tucker et al., 2010 predictions. (c) Selected targets from the Virtual Footprinting (Münch et al. 2005) predictions using a conserved half site sequence. EMSA shift results Yes= shift No =no shift.
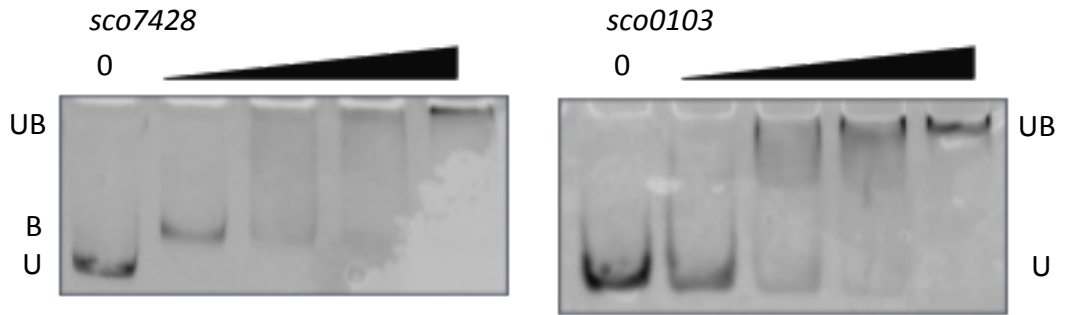
Figure 3-7 EMSA reactions showing attempts to shift *hmpA1* (*sco7428)* and a putative *hmpA3* (*sco0103*) with NsrR. Reactions were carried out utilising ratios of Cluster:DNA of 0, 2, 5, 8 and 16 (0, 40 nM, 100 nM, 160 nM and 320 nM respectively) left to right using 20 nM DNA.. Unbound (U) bound (B) and non-specifically bound (NB) DNA is indicated where present.
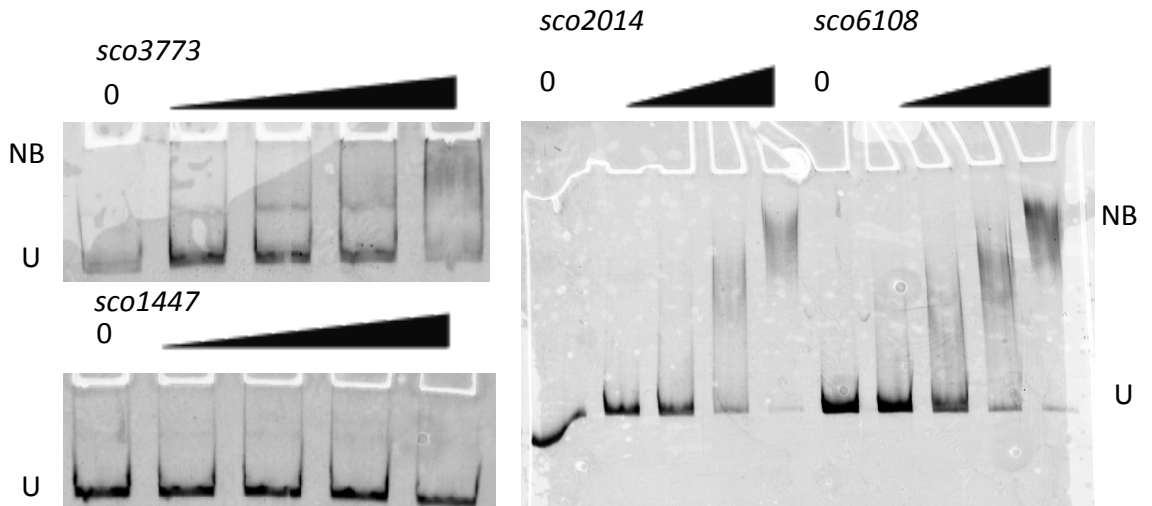
Figure 3-8 EMSA reactions showing attempts to shift targets (*sco3773, sco1447, sco2014 and sco6108*) with multiple predicted binding sites identified from the Tucker 2010 predictions with NsrR. Reactions were carried out utilising ratios of Cluster:DNA of 0, 2, 5, 8 and 16 (0, 40 nM, 100 nM, 160 nM and 320 nM respectively) left to right using 20 nM DNA. Unbound (U) and non-specifically bound (NB) DNA is indicated where present.
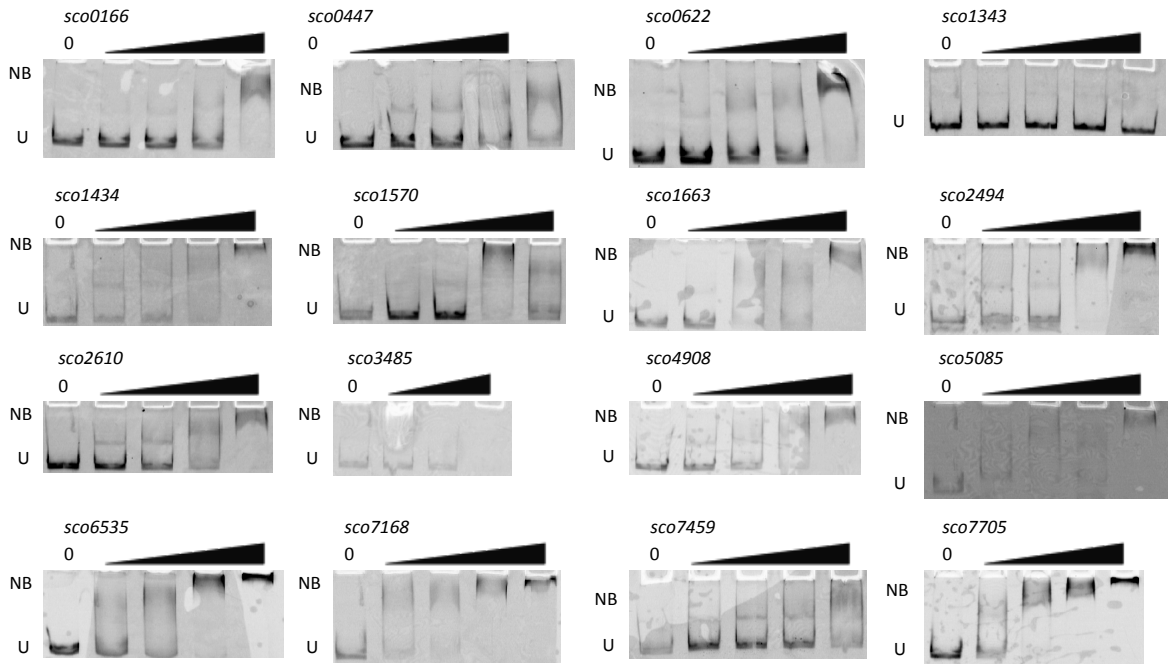
Figure 3-9 EMSA reactions showing attempts to shift key virtual footprinting targets with NsrR. Reactions were carried out utilising ratios of Cluster:DNA of 0, 2, 5, 8 and 16 (0, 40 nM, 100 nM, 160 nM and 320 nM respectively) left to right using 20 nM DNA. Unbound (U) bound (B) and non-specifically bound (NB) DNA is indicated where present.

We think it unlikely that the partial shifts observed in many of these targets corresponds to legitimate NsrR regulated targets. We instead hypothesize that NsrR can bind weakly to closely conserved binding sequences with ultimately no biological impact. Although strong enough to observe binding *in vitro*, the interaction is likely weak enough to be considered transient *in vivo* as we suspect is the case in regards to the ChIP-seq results. Following on from this we wished to discover if there

3.2.2 Disruption of the NsrR regulon: Single, double and triple mutations

3.2.2.1 Disruption mutants do not suffer from developmental phenotypes

Previous work on the *S. coelicolor nsrR* gene indicated that a disruption mutant had a white developmental phenotype under wild type growth conditions (Knowles 2014). To confirm that this was the case and to investigate the other members of the regulon, disruption mutants of each along with double and triple mutants were produced. All constructs utilised for disrupting the NsrR regulon genes were produced as part of this work. Using *S. coelicolor* cosmid 6D11 (Redenbach et al. 1996), we produced disruption cosmids (as described in section 2.3.3) for the single and double disruption of *nsrR::apr, hmpA1:apr* and *nsrR/hmpA1::apr* using pIJ773 as a source of the disruption cassette. The *hmpA2::hyg* disruption cosmid was produced from cosmid 3A4.2.A04 (Fernández-Martínez et al. 2011), with pIJ10700 used as a template for the disruption cassette. All mutated cosmids were confirmed by PCR using combinations of disruption cassette specific and gene specific test primers (Figure 3.10). These cosmids were introduced into wild-type *S. coelicolor* M145 to produce the *nsrR::apr (JTM021), hmpA1::apr (JTM022), hmpA2::hyg (JTM023)* and *nsrR/hmpA1::apr (JTM024)* mutants (as describe in section 2.3.3). To produce the *hmpA1::apr/hmpA2:hyg (JTM025)* double and *hmpA2::hyg nsrR/hmpA1::apr* triple mutant (*JTM026*), the *JTM023* strain was used as the background. Each strain was produced in triplicate (non clonal colonies) and confirmed by PCR (Figure 3.11). None of these strains have any developmental issues when grown under standard conditions, on MS agar at 30°C (Figure 3.12). We additionally produced a *hmpA1* over expression construct, under the control of the *ermE** promoter, pJM022, introduced this into M145 and confirmed it by PCR (Figure 3.13). This was to confirm that constitutive expression of HmpA has no developmental effects under wild-type growth conditions as would occur in the Δ*nsrR* strain.
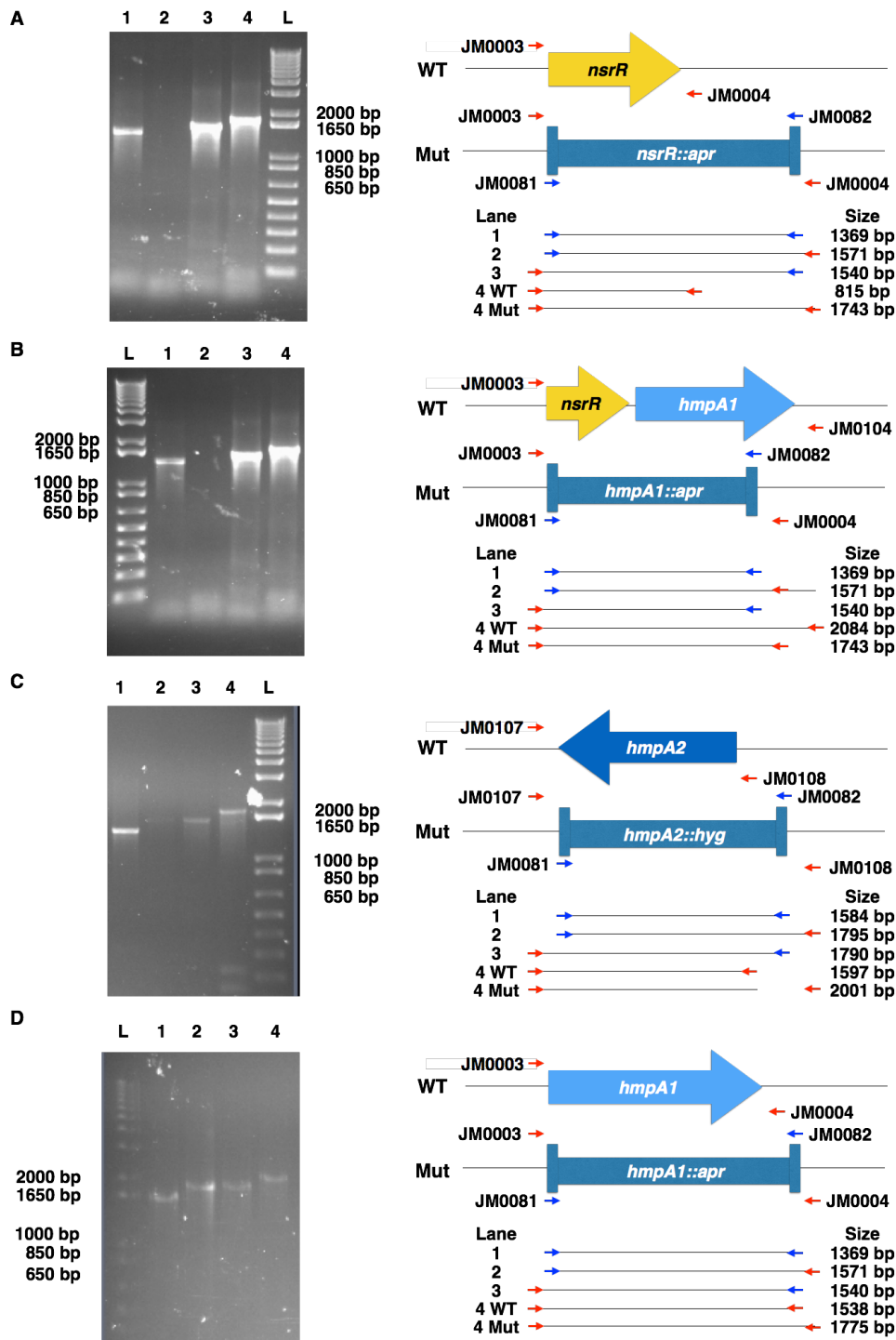
Figure 3-10 Confirmation of *nsrR, hmpA1* and *hmpA2* disruption cosmids. The left panel in each shows PCR results separated on 1% agarose gels and the right panel indicates the PCR products produced, the primers used, the lane for each product and the size in bp. Disruption cosmid production confirmation by PCR for (a) *nsrR::apr* using cassette specific primers JM0081+JM0082 along side target specific primers JM0003+JM0004, (b) *nsrR/hmpA1::apr* using primers JM0003+JM0104, (c) *hmpA2::hyg* using primers JM0107+JM0108 and (d) *hmpA1::apr* using primers JM0103+JM0104. Predicted product sizes (a) *nsrR::apr,* lane 1: JM0081+JM0082, 1369 bp, lane 2: JM0081+JM0004, 1571 bp,

lane 3: JM0082+JM0003, 1540 bp and lane 4: JM0003+JM0004, 1742 bp (WT gene size 815 bp). Predicted product size (b) *nsrR/hmpA1::apr,* lane 1: JM0081+ JM0082, 1369 bp, lane 2: JM0081+JM0104, 1540 bp, lane 3: JM0082+JM0003, 1572 bp and lane 4: JM0003+31, 1743 bp (WT gene size 2084 bp). Predicted product size (c) *hmpA2::hyg,* lane 1: JM0081+JM0082, 1584 bp, lane 2: JM0081+JM0108, 1795 bp, lane 3: JM0082+JM0107, 1790 bp and lane 4: JM0107+JM0108, 2001 bp (WT gene size 1597 bp). Predicted product size (d) *hmpA1::apr,* lane 1: JM0081+JM0082, 1369 bp, lane 2: JM0081+JM0104, 1571 bp, lane 3: JM0082+JM0103, 1540 bp and lane 4: JM0103+JM0104, 1775 bp (WT gene size 1538 bp). Invitrogen 1 kb plus DNA ladder (L) was used.
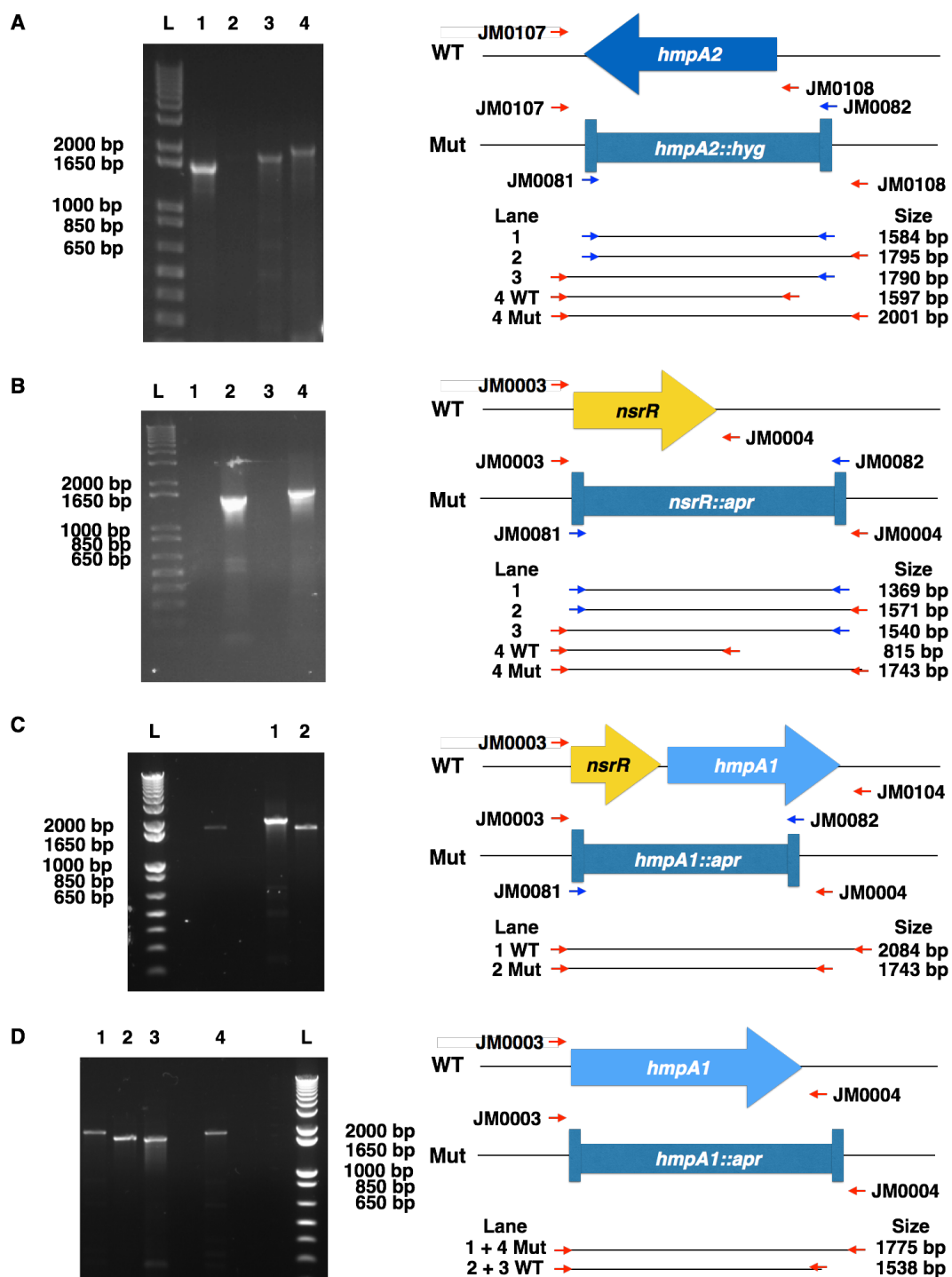
Figure 3-11 Confirmation by PCR of *nsrR, hmpA1* and *hmpA2*, single, double and triple gene disruptions in M145. The left panel in each shows PCR results separated on 1% agarose gels and the right panel indicates the PCR products produced, the primers used, the lane for each product and the size in bp. Confirmation of (a) *hmpA2::hyg* which acts as the starting strain for the triple mutant, (b) *nsrR::apr*, (c) *hmpA2::hyg nsrR/hmpA1::apr triple mutant* and (d) *hmpA2::hyg hmpA1::apr* double mutants. Gene disruptions were confirmed using combinations of primers JM0082 + JM0083 and gene specific primers. Confirmation of (a) *hmpA2::hyg, hmpA2::hyg,* lane 1: JM0081+JM0082, 1584 bp, lane 2:

JM0081+JM0108, 1795 bp, lane 3: JM0082+JM0107, 1790 bp and lane 4: JM0107+JM0108, 2001 bp (WT gene size 1597 bp). Confirmation of (b) *hmpA2::hyg nsrR::apr, nsrR::apr,* lane 1: JM0081+JM0082, 1369 bp, lane 2: JM0081+JM0004, 1571 bp, lane 3: JM0082+JM0003, 1540 bp and lane 4: JM0003+JM0004, 1742 bp (WT gene size 815 bp).. Confirmation of (c) *nsrR/hmpA1::apr* using primers JM0003+JM0104: lane 1, WT genomic DNA (2084 bp), Lane 2, mutant DNA (1743 bp). Confirmation of (d) *hmpA1::apr,* using primers JM0103+JM0104, lane 1 + 4, successful mutants, lane 2, failed mutant, lane 3 genomic DNA (mutant size: 1775 bp, WT gene size 1538 bp). Invitrogen 1 kb plus DNA ladder (L) was used.
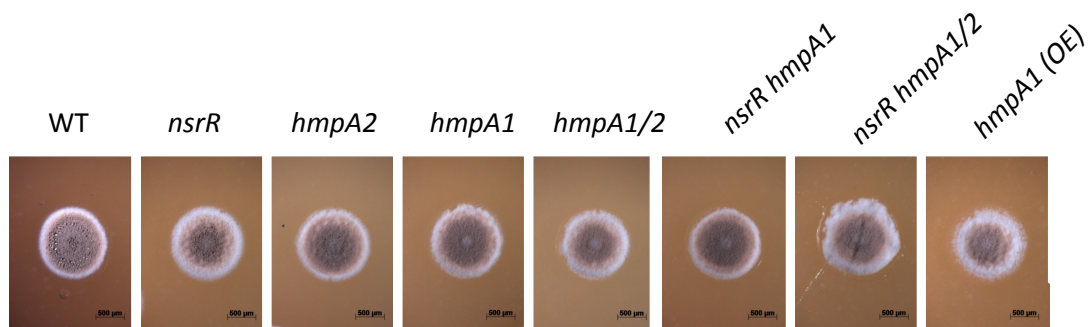
Figure 3-12 Light microscopy of the *nsrR, hmpA1* and *hmpA2,* single, double and triple mutants and the *hmpA1* overexpression (OE) strain. Samples where grown on MS agar for 5 days at (30°C).
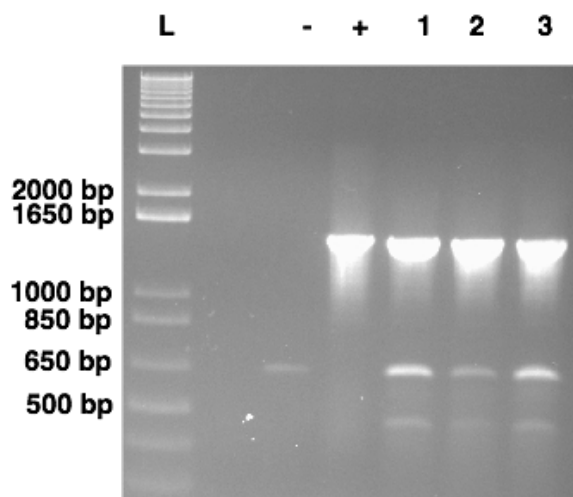
Figure 3-13 Confirmation of insertion of pIJ10257 *hmpA1* into M145 using primers JM0113+JM0114. M145 genomic DNA (-) and vector DNA (+) were used as negative and positive controls respectively. Lane 1-3 corresponds to M145 strains containing the pJM022 in triplicate.

107

3.2.3 *S. coelicolor* and NO: Is sporulation inhibition the tip of an ice-berg?

3.2.3.1 Inoculation of *S. coelicolor* with NO inhibits sporulation

Following the confirmation of the NsrR regulon, reserved specifically for NO detoxification, we wished to investigate the effect NO donors have on aerobically grown *S. coelicolor*. Incubation of wild-type *S. coelicolor* with the donor, DETA NONOate, resulted in a white-colony morphology (Figure 3.14). We wished to investigate whether the white phenotype was due to inhibition of sporulation or just a loss of the brown WhiE spore pigment using cryo-SEM (Scanning Electron Microscopy) (Figure 3.15). It is apparent from the SEM imaging results that sporulation has indeed been inhibited with a dramatic reduction in the number of spores observed in wild-type M145 treated with NO compared with untreated sample but no evident differences in colony size or biomass. A few rare spore chains are visible in the NONOate treated samples however.

To test if NsrR or its regulon are involved in this phenotype we repeated the NONOate experiments with the WT strain and the *nsrR*, *hmpA1* and *hmpA2* mutants. We see a similar delay in the onset of sporulation with no observable difference in phenotype across the strains (Figure 3.16).
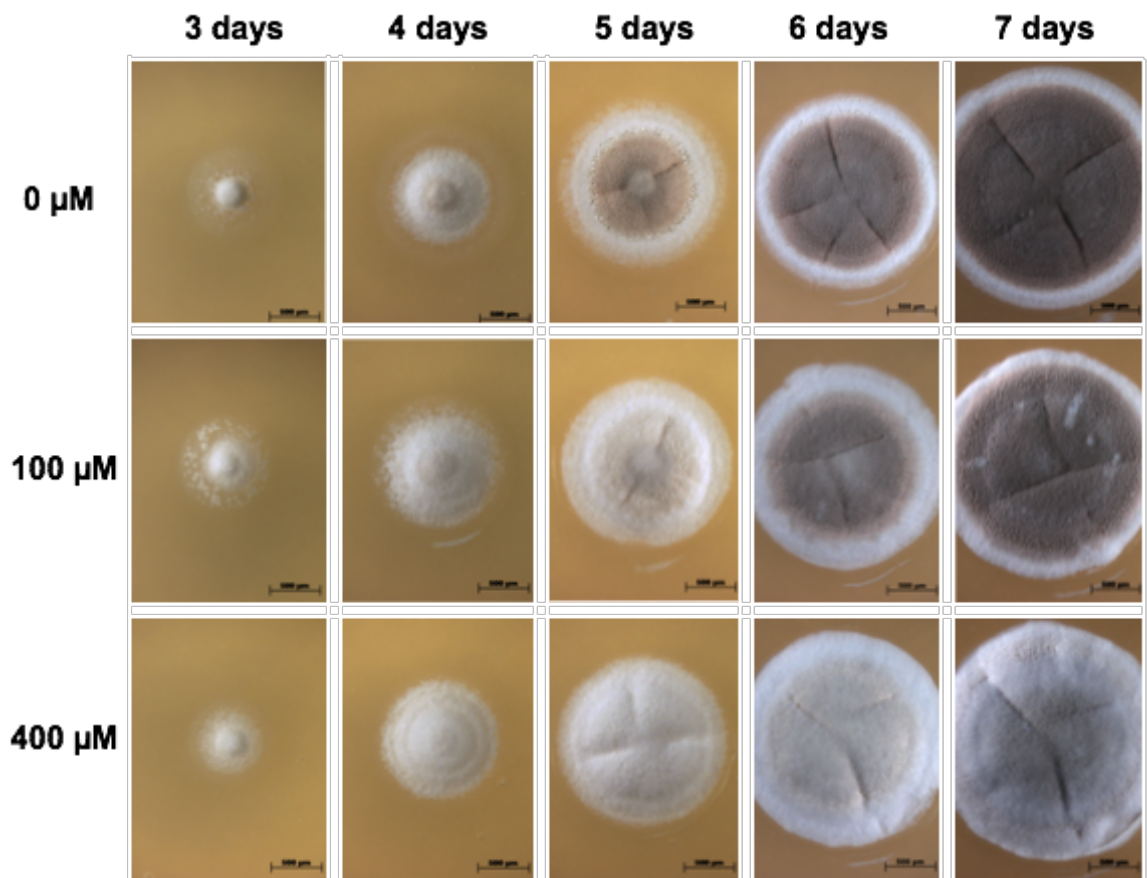
Figure 3-14 Light microscopy showing colony morphology of wild-type *S. coelicolor* M145 grown in the absence and presence of DETA NONOate; an NO donor. Images were taken daily between days 3 and 7. A final concentration of 0 μM, 100 μM and 400 μM DETA NONOate was incorporated into MS media before inoculation and growth at 30°C.
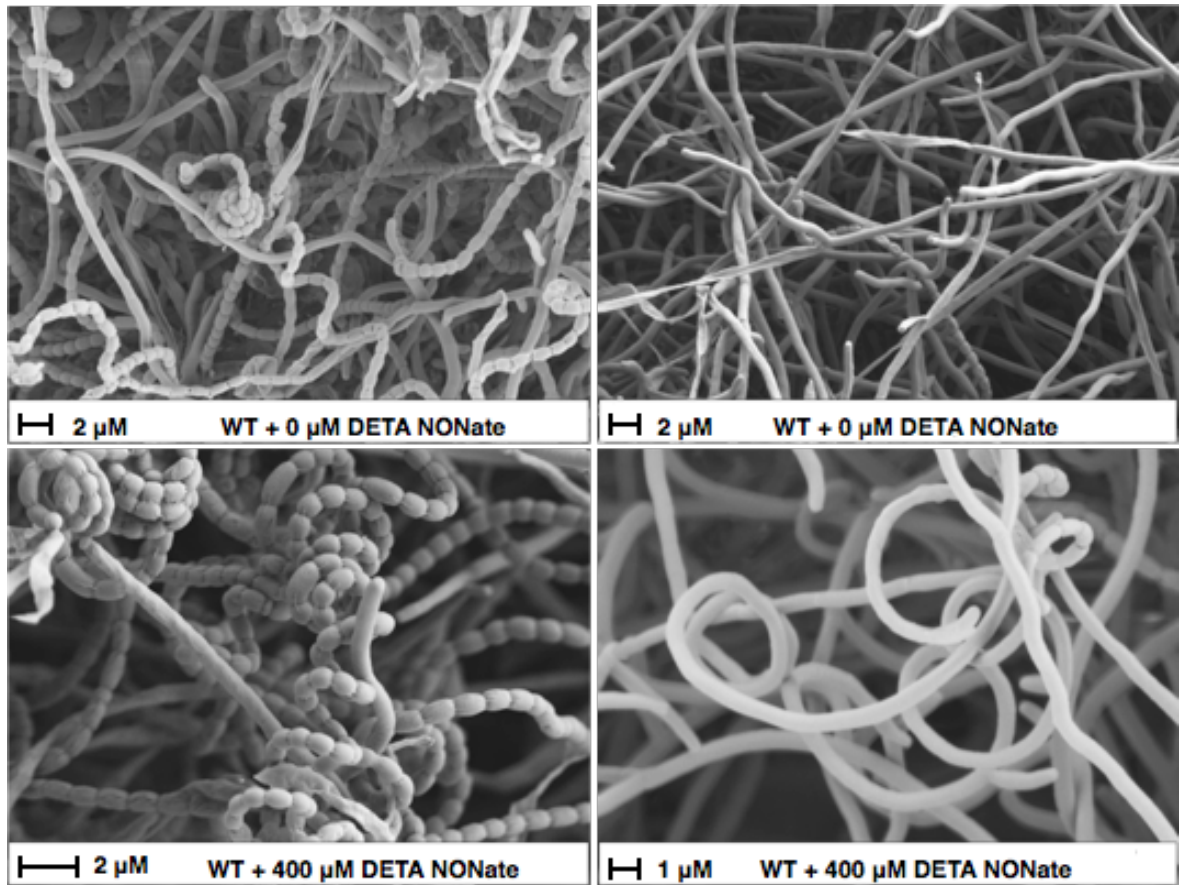
Figure 3-15 Cryo-SEM images of *S. coelicolor* M145 in the presence and absence of DETA NONOate. Samples were grown on MS at 30°C for 4 days.
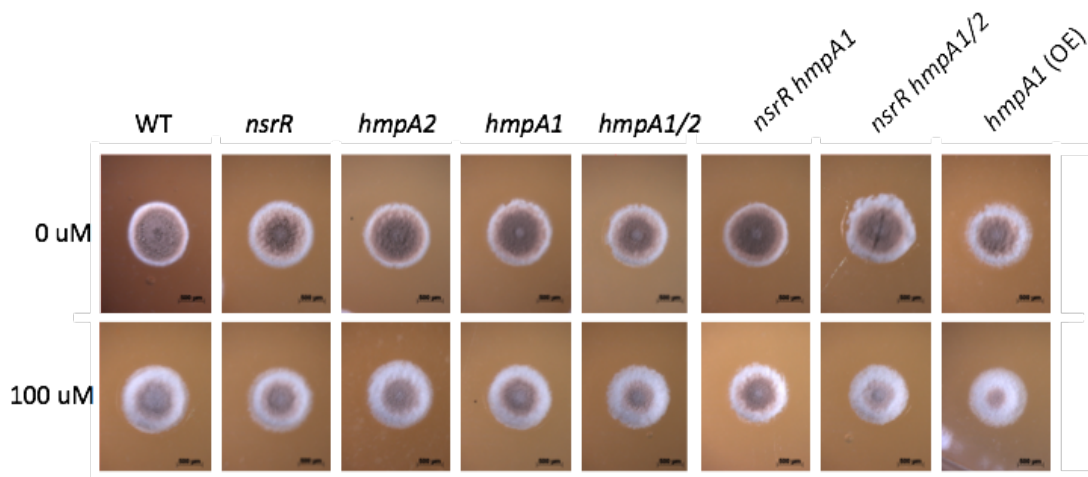
Figure 3-16 Light microscopy of the *nsrR, hmpA1* and *hmpA2,* single, double and triple mutants and the *hmpA1* overexpression (OE) strain. Samples where grown on MS agar for 5 days at (30°C) in the presence or absence of 100 µM DETA NONOate.

3.3   Summary

We conclude, based on the ChIP-seq EMSA and disruption data that *S. coelicolor* NsrR has a three-gene regulon consisting of *nsrR, hmpA1 and hmpA2.* The regulon is specifically suited for the sensing and detoxification of NO *in vivo* where NsrR is an NO sensor, and two predicted HmpA genes, for the conversion of NO to nitrate. Single, double and triple disruption mutants of the regulon targets have no growth or developmental phenotypes and indeed none we have yet been able to identify under the conditions tested when compared to WT. We have identified that NONOate treatment of wildtype cells results in the specific inhibition of sporulation as observed by light microscopy and SEM.

3.4   Discussion

NsrR has been studied throughout the bacterial kingdom, in the high GC Gram-positive actinomycetes *Streptomyces* (Crack et al. 2016; Crack et al. 2015; Tucker et al. 2008), the low GC Firmicute *B. subtilis* (Kommineni et al. 2012; Henares et al. 2014), as well as in the Gram-negative γ-proteobacteria *E. coli* and *Salmonella* (Partridge et al. 2009; Karlinsey et al. 2012; Branchu et al. 2014) and beta-proteobacteria *Neisseria* (Vincent M. Isabella et al. 2009)*.* The *S. coelicolor* NsrR regulon contains 3 target genes (Crack et al. 2015). *B. subtilis* and *E. coli* K12 on the other hand, control >35 genes (Kommineni et al. 2013; Härtig & Jahn 2012; Henares et al. 2014; Kommineni et al. 2012; Kommineni et al. 2010) and >60 genes (Tucker et al. 2008; Bodenmiller & Spiro 2006a; Partridge et al. 2009) respectively, some of which have no direct link to NO protection and often through interaction with other regulators. Some of the *E. coli* NsrR target genes are regulated by binding to half sites (Partridge et al. 2009) however we have been unable to show this for *S. coelicolor*. The *Neisseria* NsrR regulon, more similar in size to *S. coelicolor*, contains only 5 genes, *aniA* (nitrite reductase)*, norB* (NO reductase)*, mob* (involved in molybdenum metabolism)*, dnrN* (involved in [Fe-S] repair of nitrosylated clusters) and *nsrR* itself (Vincent M Isabella et al. 2009; Edwards et al. 2012). Lacking *hmp, Neisseria* rely on the conversion of nitrite to nitrous oxide, avoiding NO production, using *aniA* and *norB* (Barth et al. 2009). Although NsrR proteins appeared to have evolved generally to perform the role of a global NO stress response system, we suggest that in *S. coelicolor* it plays a far more refined and specialised role, similar to that for the NorR regulon, directly controlling a small set of NO detoxification genes (D'Autréaux et al. 2005). Unlike many other regulators, including FNR and SoxR, NorR is a direct and specific sensor for NO, acting through a non-haem iron (D'Autréaux et al. 2005). NorR exists as a dimer, bound to DNA, each monomer consisting of tripartite domain architecture with a C-terminal HTH domain, an AAA+ domain and a N-terminal GAF domain containing the non-haem iron (D'Autréaux

et al. 2005). Under non-stress conditions, The GAF domain binds the AAA+ domain inhibiting ATPase activity (D'Autréaux et al. 2005). Upon nitrosylation under NO stress conditions, the GAF domain releases the AAA+ domain, allowing ATPase activity to proceed and the activation of transcription of NorVW, a flavorubredoxin, to detoxify NO (D'Autréaux et al. 2005). This NorR contains a small regulon, regulating only *norVW* expression (D'Autréaux et al. 2005). No NorR homolog exists in *Streptomyces* however it is possible that NsrR, although a very different protein, carries out a similar functionality.

3.4.1 ChIP-seq, EMSA and binding site

Bioinformatics approaches to define the NsrR regulon, both before and during this work, have implicated a wide range of targets based on a conserved binding sequence from the *hmpA1* and *hmpA2* sequences. We have carried out a range of EMSA reactions on these predicted targets and have found little to no evidence for specific DNA-binding to any targets other than the three identified in the ChIP-seq experiment. Specifically of interest was a third *hmp* gene identified in *S. coelicolor* genome however NsrR was unable to shift the promoter of this gene. We suggest that this *hmp* gene was horizontally acquired, based on its position within the arms of the chromosome (a common characteristic for acquired genes) however we have provided no functional data for any of the three proteins so, even though *sco0103* is not under the direct control of NsrR, it does not exclude it from being involved in the NO stress response.

Several of the EMSA targets appeared to have specific, partial shifts when approaching high holo-protein to DNA ratios. Part of the rational for each EMSA reaction was to include a range that covered complete DNA binding (Crack et al. 2015) for each confirmed target: *hmpA1*, *nsrR hmpA2* (~2.5:1, 5:1 and 8:1 respectively) cluster to DNA along with a 0 control sample of DNA only and a reaction sufficient to achieve non-specific binding at ~16:1 cluster to DNA. In each case, we were unable to achieve a complete band shift, before non-specific binding occurred. We suggest the targets, as is the case in *B. subtilis*, could be class II (cluster independent) targets bound by apo-protein. We think this is unlikely but not out of the question from these results. Each NsrR sample had a range of ~33-75% incorporation of [4Fe-4S], depending on the purification batch, during experimentation. As a result, in some cases, we have more apo protein than holo and, based on experimentation with the apo protein, we see little to no shift of the major targets within our experimental concentration range but EMSA with 100% apo-NsrR was not carried out for the majority of samples tested. It is possible weak binding occurred with either apo or holo-protein in the samples and due to the accumulative amount of protein we begin seeing non-specific shifts before complete specific shifts occurred. We suggest from this that NsrR specificity is stringent in regards to strong, biologically relevant

reactions, however weak interactions occur that have an unknown or no biological relevance e.g. do not inhibit RNA polymerase or other DNA-binding interactions even when bound (e.g. interaction is so weak they are displaced).

We have provided two major suggestive pieces of evidence for the stringency of NsrR target binding sites on top of testing a range of predicted targets, which failed to shift. Firstly, replacement of the end AA-TT bases with GT-CA (ends of the nest primers) we see that DNA-binding to the *hmpA1* target is abolished. Based on how the experiment was designed this result is not due to the protein having insufficient binding space as amplification primers of the probe continue for at least 10 bases either side. Secondly, our GC rich probe (Table 3.1) was designed to contain the universally conserved base from all there target probes and at each position where a base different between binding sites, the G or C base option was preferentially selected. For the purpose of clarity this is illustrated in Table 3.3. Only 5 bases differ between the GC rich probe and *hmpA2* probe however looking at these positions (3, 4, 16, 18 and 19) in the *hmpA1* probe we see that the bases match (positions 3, 4, 16 18 and 19 match in *hmpA1* and the GC rich probe) while the rest match in the *hmpA2*/GC rich probes. On further inspection of the binding sites/probes, *hmpA1, nsrR* and *hmpA2* are all imperfect inverted repeats, while the GC rich probe is a perfected inverted repeat. It is possible that the imperfection of the repeat is important to the binding architecture. It is difficult to ascertain from the data provided which of these bases or base groupings are essential however it would be interesting to investigate this further following a point mutation based approach. Although the data is not complete, we are confident that based on these results there are no other direct targets.

Promoter	Sequence	NsrR binding
hmpA1	AACACGAATATCATCTACCAATT	Y
GC rich (1)	AACGCGCATCTGAGATGCGCGTT	N
nsrR	AAGGCGAACCTAGCATGCGCATT	Y
GC rich (N)	AACGCGCATCTGAGATGCGCGTT	N
hmpA2	AACAAGCATCTGAGATCCCAGTT	Y
GC rich (2)	AACGCGCATCTGAGATGCGCGTT	N
GC rich (all)	AACGCGCATCTGAGATGCGCGTT	N

Table 3-3 A comparison of the EMSA probe base conservation with *hmpA1 nsrR* and *hmpA2* are compared to the GC rich probe. Each NsrR target is compared to the sequence of the GC rich probe highlighting the concerned bases (yellow for *hmpA1*, green for *nsrR*, cyan for *hmpA2* and red if conserved in all targets).

Previous work in the *B. subtilis* field has implicated ResD and Fur as co regulators of NsrR targets (Henares et al. 2014; Kommineni et al. 2012). ResD has been shown to be necessary for the activation of these genes following the release of NsrR repression. It is possible that in *S. coelicolor* there are also additional levels of regulation on these targets we are yet to define. This would be interesting to investigate further. It will be important to bare in mind, for any expression work carried out in the future, that In the case of *E. coli nsrR*, it is a highly expressed gene but its translation is inefficient resulting in a relatively low abundance of NsrR being present (Chhabra & Spiro 2015) this may be the case with ScNsrR.

3.4.2   NO, phenotyping and possible links to the delay in sporulation

Following the disruption of all three NsrR regulated genes, we have observed no developmental or non-wildtype phenotype under our test conditions.

As indicated in our results, we have so far been unable to determine a condition where we can induce the NsrR regulon, except in an *nsrR* disruption mutant. Currently no experimentally derived NO specialised regulons have been identified in *S. coelicolor*, beyond NsrR which is a predicted NO sensor based on *in vitro* (DNA-binding/spectroscopic studies in the presence of NO) experimentation (Crack et al. 2016; Crack et al. 2015), *in vivo* identification (ChIP-seq) and *in silico* predictions of its regulon (i.e. *hmpA* genes) and comparison to other organisms (*E. coli, B. subtilis etc.*) which have been shown to respond to NO (Vine et al. 2011; Henares et al. 2014). We have recently shown that NsrR is an NO binding regulator with NO disrupting DNA-binding to the three reported targets (Crack et al. 2016) and based on the function of the targets we suspect that NO detoxification and detection is the sole/primary function of the NsrR regulon.

In addition to NsrR there are other candidates as NO sensors, including the Wbl proteins generally, with specific focus on WblE and WhiD as previously discussed (section 1.3.4.3.1). We suggest that, NsrR is not linked directly to the cause of the NONOate induced white phenotype but that nitrosylation of the Wbl protein [4Fe-4S] results in the specific delay of sporulation (Figure 3.17). Although we have not been able to confirm this link yet, work has already been undertaken to investigate the possible WhiD and WblE regulons and any involvement they might have in the NONOate phenotype. Previous studies have linked the *wbl* genes to NO through co-conservation of genes (Chandra & Chater 2014) or through *in vitro* studies showing them to be NO binding proteins *via* their [4Fe-4S] cluster (Singh et al. 2007; Smith et al. 2010; Crack et al. 2011; Crack et al. 2013; Stapleton et al. 2012). To further investigate the phenotype and indeed NO/NsrR within *S. coelicolor*, we have submitted several NONOate treated samples for RNA-seq analysis,

to investigate any change in expression resulting from the NONOate treatment. These results will guide our hand in future experimentation and hopefully highlight important genes.
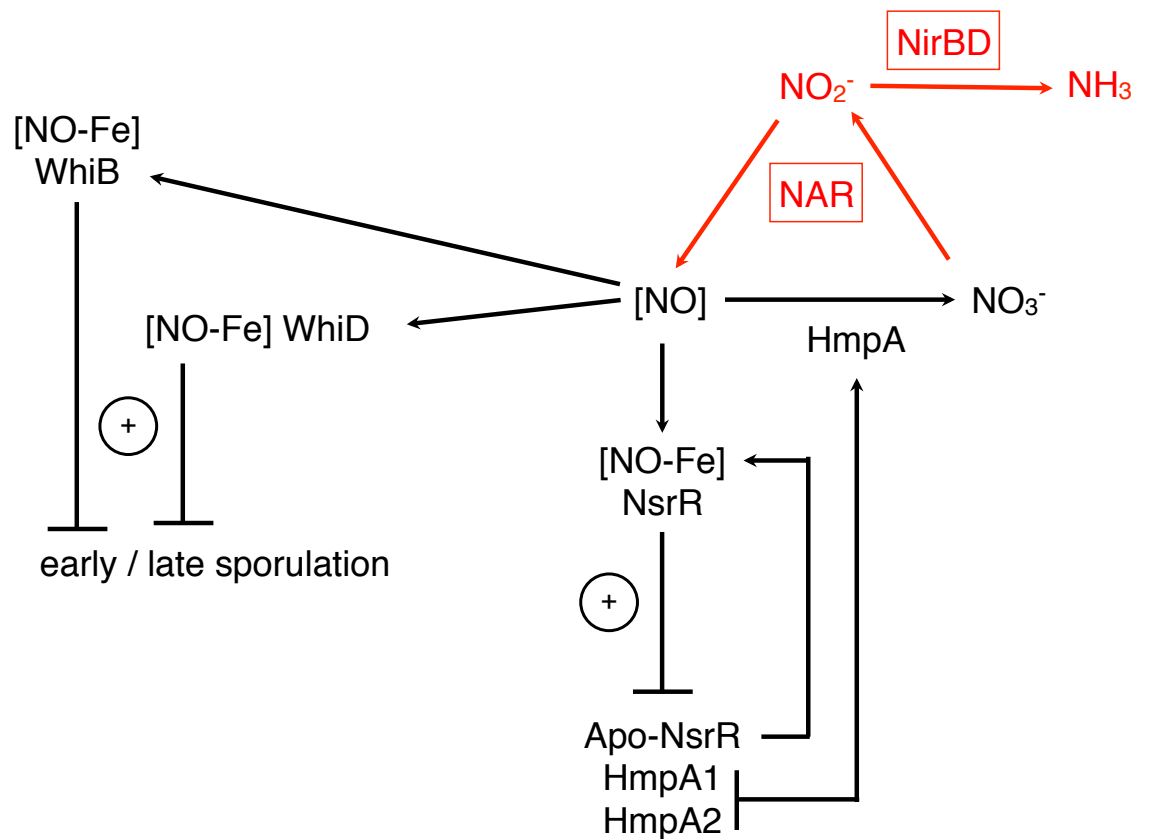
Figure 3-17 A diagrammatic summary of the ScoNsrR pathway showing: NO production and detoxification and the putative targets of NO involved in the white developmental phenotype following NONOate treatment.

4   RsrR: a novel member of the Rrf2 superfamily in *S. venezulae*

In parallel to the ScoNsrR research, the closest known ortholog in *S. venezuelae, rsrR* (putatively named <u>r</u>edox <u>s</u>ensitive <u>r</u>esponse <u>r</u>egulator), was investigated. Interest in *rsrR* began following the determination of the small ScoNsrR regulon that, as previously discussed, is dissimilar to other reported NsrR. *rsrR* was identified as an *nsrR* homologue through reciprocal blast analysis with *S. coelicolor nsrR* using StrepDB and ultimately became a natural candidate to continue research into *rrf2* genes in *Streptomyces*.   A comparison of the two proteins however highlights that they have significantly different primary sequences (Figure 4.1) sharing only 26% amino acid identity. A table of all *rrf2* genes and their reciprocal BLAST results found on StrepDB was produced and a multiple alignment containing all results was carried out (Supplementary data S4.1 and S4,2). Of the three predicted *rrf2* genes, *sven6563* was selected for further study due to the presence of five cysteine residues, three of which are commonly conserved and have been shown to be to be important in NsrR for ligation of a [4Fe-4S] cluster (the 4th ligand, in the case of ScoNsrR is likely E85) and ultimately its capacity to bind DNA (Crack et al. 2015).

To investigate *rsrR,* a similar repertoire of experiments to those carried out to study *SconsrR* was employed. It was initially hypothesised that *rsrR,* although fairly dissimilar to *nsrR*, would play a similar role*.*

```
sco7427     VRLTKFTDLALRSLMRLAVVRDGDEPLATREVAEVVGVPYTHAAKAITRLQHLGVVEARR
sven6563    MKLSGGVEWAL---HCCVVLTAASRPVPAARLAELHDVSPSYLAKQMQALSRAGLVRSVQ
            ::*:  .: **       .*:  ...*: : .:**:  *  :: ** :  *.: *:*.: :


sco7427     GRGGGLTLTDLGRRVSVGWLVRELEGEAEVVDCEG-------DNPCPLRGACRLRRALR
sven6563    GKTGGYVLTRPAVEITLLDVVQAVDGPDPAFVCTEIRQRGPLATPPEKCTKACPIARAMG
            *: ** .**   . .:::  :*: ::*    .. *            *     ** : **:


sco7427     DAQEAFYAALDPLTVTDLVAAPT---GPVLL-GLTD---RPSG*
sven6563_   AAEAAWRASLAATTIADLVATVDDESGPDALPGVGAWLIEGLG*
             *: *: *:*    *::****:      **  * *:       .  **
```

Figure 4-1 An alignment of NsrR (*sco7427*) and RsrR (*sven6563*) amino acid sequences. Cysteine residues are highlighted in red and the glutamic acid residue, the likely 4th ligand in NsrR, is highlighted in blue. Alignments were carried out using Clustal Ω (Sievers et al. 2011).

4.1.1 Summary of work

To investigate RsrR both ChIP-seq and RNA-seq were carried out with the primary aim of determining the core regulon of this transcription factor. Disruption of RsrR was carried out producing an apramycin marked deletion mutant (JTM034). This mutant then had a second, *in trans* copy of *rsrR* introduced containing a C-terminal 3x FLAG-tag to facilitate ChIP-seq using α-FLAG antibodies. More than 600 target binding sites were identified from ChIP-seq each containing either a full (11-3-11 bp inverted repeat) or half (~11 bp) site similar to previously reported Rrf2 binding sequences. The regulon includes many genes associated with cellular redox homeostasis including many NAD+/NADH or NADP+/NADPH utilising proteins. dRNA-seq was carried out to determine the location of transcriptional start sites (TSS) within the *S. venezualae* genome and to investigate the differential gene expression of WT vs. *rsrR* mutant. Following an independent analysis of the expression results we observed but up and down regulated genes. Between 8-25% of these correlate with ChIP-seq targets indicating the two data sets do not correlate well. A major target of interest in both is *sven6562*. *sven6562* the divergently expressed gene of *sven6563, rsrR*, that was highly expressed (a $\log_2$ fold change of 5.4) following the disruption or *rsrR*. On investigation, *sven6562* was determined to be LysR-type transcriptional regulator now denoted as NmrA. NmrA contains a C-terminal DNA-binding domain and an N-terminal NmrA domain hypothesised to sense cellular redox poise by binding NAD(P)+ but not NAD(P)H. *In vitro* studies using purified RsrR were carried out to validate the ChIP-seq results and to determine key biochemical properties of the protein. It was determined that RsrR contains a [2Fe-2S] cluster, functions as a dimer and can be reversibly redox cycled between oxidised and reduced states. Much of the biochemistry was carried out by collaborators and as such will not be discussed in great detail. Based on the available data it is speculated that RsrR and NmrA co-regulate/have overlapping regulons controlling a redox responsive regulon acting as an initial sensor of redox stress and a late stage sensor of damage/the depletion of the NAD(P)H pool, respectively, within the cell as a result of severe or prolonged redox stress.

4.2 Results

4.2.1 *rsrR* genetics

4.2.1.1 Disruption of *rsrR* (*sven6563*) in *S. venezuelae*

In an attempt to elucidate the function of *rsrR*, a disruption mutant was produced to determine if the gene was essential or identify any resulting phenotypes and begin the production of a ChIP-seq suitable strain. Disruption of *rsrR* was carried out, as described,

using the cosmid SV-5-F05 (a gift from the John Innes Centre and hereafter denoted as 5F05) and an apramycin resistance cassette amplified from pIJ773 as described. The *sven6563* disruption cassette was produced using JM0109/JM0110 and introduced into *E. coli* strain BW25113 pIJ790 containing 5F05. Confirmation of successful gene disruption in the *E. coli* background was carried out by PCR (Figure 4.2a) using a combination of cassette specific primers and those flanking the gene (JM0081/JM0082, JM0081/JM0112, JM0082/ JM0111, JM0111/JM0112) showing that both wild-type 5F05 (WT *rsrR*= 800 bp) and mutated (pJM026) 5F05 *rsrR::apr* (*rsrR::apr* = 1768 bp) copies, were present. Both mutant and wild-type 5F05 cosmids were isolated from BW25113 and used to transform Top10 *E. coli* cells and ET12357 pUZ8002 (from now denoted as ETpUZ) and apramycin resistant colonies were selected. The ETpUZ pJM026 strain, was conjugated with wild-type *S. venezuelae* and exconjugants were screened for apramycin resistance and kanamycin sensitivity. These successful exconjugants were checked by PCR as previously described and the resulting gel are illustrated in Figure 4-2b. The isolation of *rsrR* mutants (n=3), JTM034, had no observable developmental of growth phenotype when grown under standard laboratory conditions.
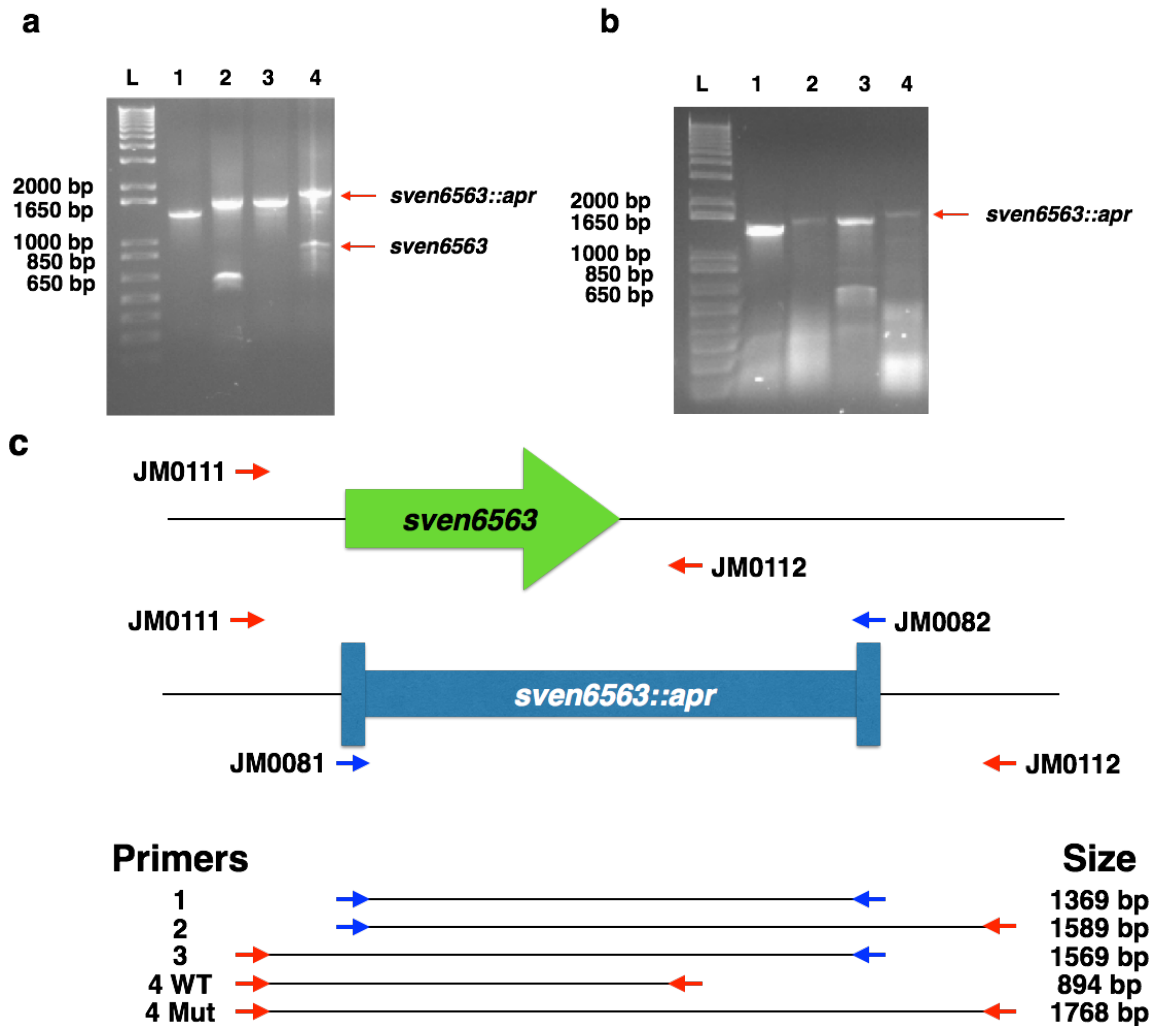
Figure 4-2 PCR confirmation of cosmid 5F05 *rsrR::apr* (pJM026) isolated from *E. coli* strain BW25113 (a) and on genomic DNA isolated from *S. venezuelae* exconjugants (b). (c) Illustrates the genotype of WT and *sven6563::apr* mutants and the expected PCR products from each primer set along with the predicted sizes indicating the well each was analysed in. Confirmation of was carried out using primer sets 1-4: (1) JM0081/JM0082, with a predicted size of 1369 bp, (2) JM0081/JM0112, with a predicted size of 1586 bp, (3) JM0082/JM0111 with a predicted size of 1569 bp and (4) JM0111/JM0112 with a predicted size of 894 bp (*rsrR* - WT) and 1768 bp (*rsrR::apr* - mutant). The DNA ladder (L) used was the Invitrogen 1kb plus DNA Ladder. Bands corresponding to *rsrR::apr* and *rsrR* are shown with arrows*.

4.2.1.2 Construction of RsrR 3x FLAG vector/strain for ChIP-seq analysis

As with ScoNsrR we carried out the ChIP-seq experiment using a 3x FLAG tag epitope and α-FLAG antibodies. For RsrR we used a C-terminal 3x FLAG epitope as this worked effectively for ScoNsrR (Crack et al 2015). The construct (Figure 4.3) was synthesised by GenScript and delivered in pUC57. It consists of (5' to 3') a HindIII site, 200 bp upstream

of the *rsrR* start codon to include the WT promoter (*rsrRp*), the *rsrR* gene minus the stop codon, a 45 bp linker (Bush personal communication), a 66 bp 3x FLAG tag sequence, a stop codon and a KpnI site to facilitate directional cloning.   The *rsrR* construct was subcloned from pUC57 into pMS82 (Gregory et al. 2003) using HindIII/KpnI digestion and ligation and positive clones were identified by colony PCR (Figure 4.4a) and then confirmed by test digestion using *Hind*III/*Kpn*I (Figure 4.4b).   The final construct, pJM027, was then sequenced using primers JTM105 and JTM106 to confirm it was correct (Supplementary data S 4.3).
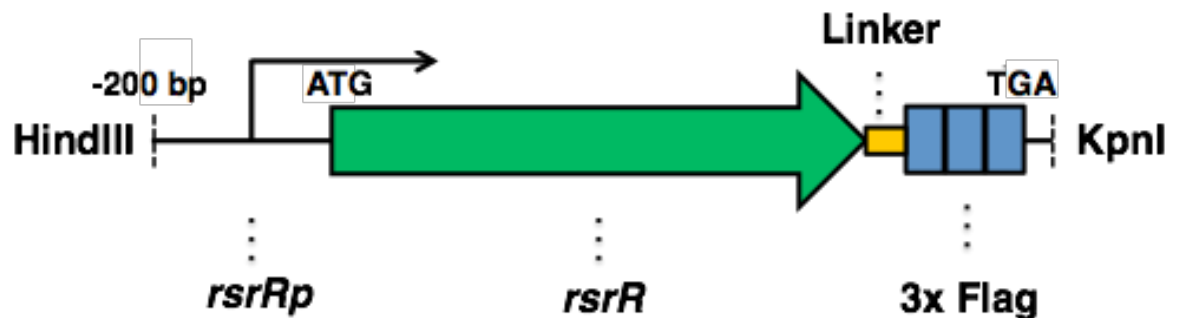
Figure 4-3 Graphical representation of the *rsrR* 3x FLAG tagged construct ordered from GenScript.
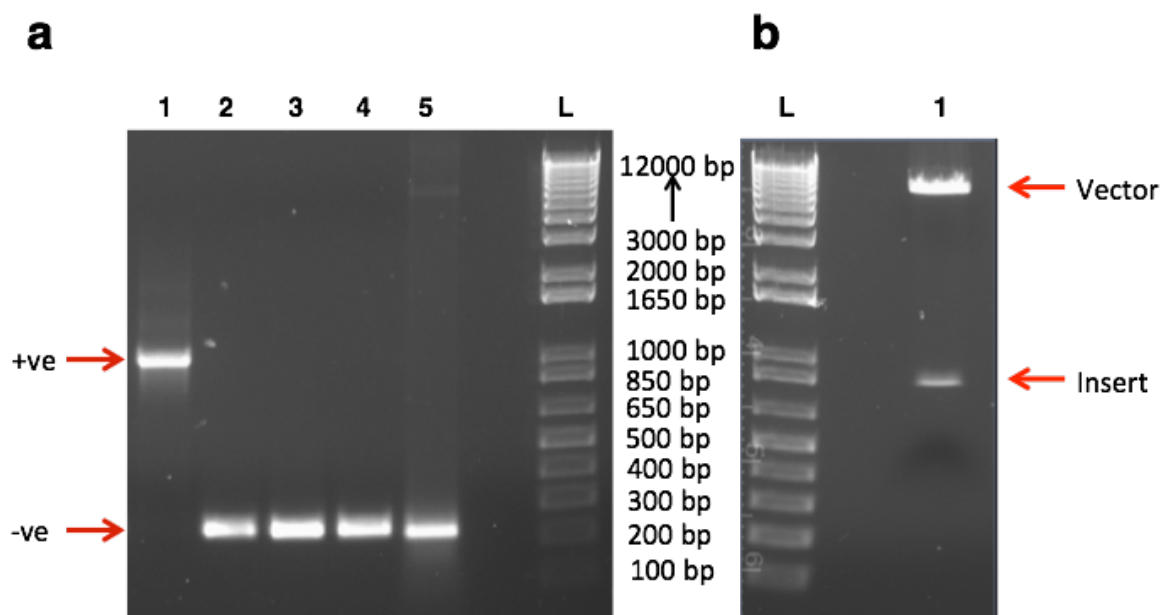
Figure 4-4 Identification and confirmation of positive pMS82 *rsrR* 3x FLAG (pJM027) clones.   (a) Colony PCRs using JM0113/JM0114 with a successful clone indicated by a band of 1012 bp and a negative, empty vector result indicated by a band of 230 bp.   (b) Restriction digestion (HindIII/KpnI) of the positive vector (from a) showing the vector (6332 bp) and insert (800 bp) bands.   The Invitrogen 1kb plus DNA Ladder (L) was used.

4.2.1.3 RsrR 3x FLAG is constitutively expressed under wild-type conditions

The sequenced *rsrR* 3x FLAG construct (pJM027) was introduced into ETpUZ and then the *S. venezeulae rsrR::apr* strain by conjugation. The resulting exconjugants were confirmed by PCR (Figure 4.5a) and immunoblotting was carried out using anti-FLAG antibodies (purchased from Sigma-Aldrich) to detect the FLAG-tagged protein (Figure 4.5b). Whole cell extracts were prepared from mycelium harvested at 12, 24, 36 hours. These samples were resolved on SDS-PAGE gels and a single band was detected that corresponded to the RsrR-3x FLAG molecular weight of 20.2 kDa.   These results suggest RsrR is constitutively expressed in *S. venezuelae*.
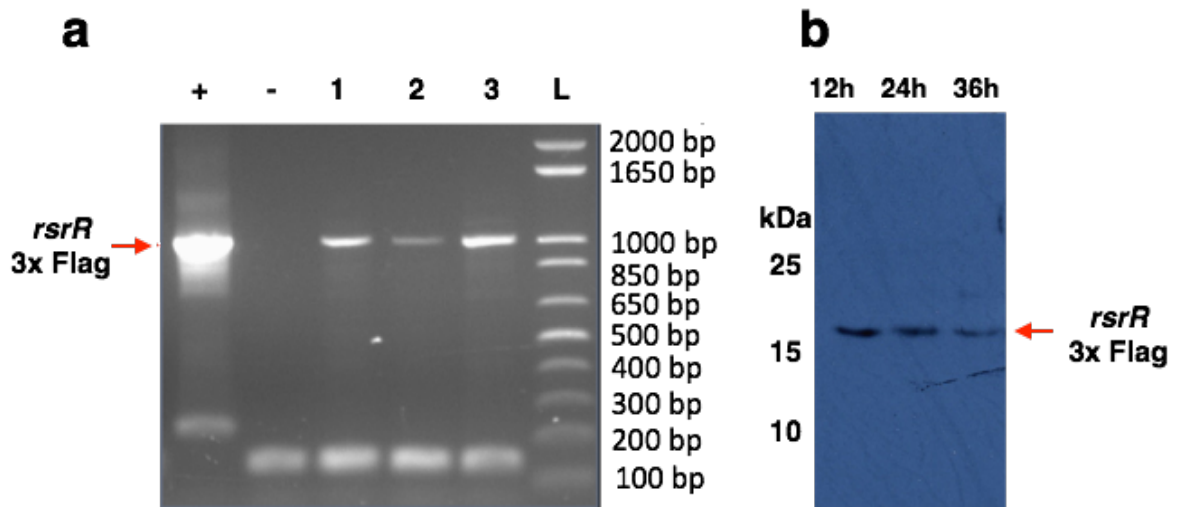
Figure 4-5 Detection of *rsrR* 3x FLAG construct pJM026 and RsrR 3x FLAG protein in the *S. venezuelae rsrR::apr* background. (a) PCR confirmation of strains using primers JM0113/JM0114. Construct DNA (+), JTM034 genomic DNA (-) and 3 replicates of the putative JTM035 strain (1-3) were tested. The PCR product using the JM0113/JM0114 primer set (if the construct is present) should be 1012 bp. The Invitrogen 1kb plus DNA Ladder (L) was used as a marker.   (b) Detection of RsrR 3x FLAG (~20kDa) by immunoblotting of whole cell extracts, over a time course spanning 12-36 hours. RsrR 3x FLAG has a predicted size of 20.2 kDa.

4.2.2 RsrR NGS approaches and bioinformatics regulon analysis

4.2.2.1 ChIP-seq identification of the RsrR regulon

The *S. venezuelae rsrR::apr rsrR* 3x FLAG strain (JTM035) was used to carry out ChIP-seq using an α-FLAG antibody along with a WT control (processed identically) as a background/genomic control. ChIP-seq was performed using cells grown for 18 h and processed as described in section (section 2.3.4.9.1.2). Wild-type *S. venezuelae* was used as a negative control. Library production (using confidential, commercially available kits) and sequencing (Illumina HiSeq 2500 platform, 50 bp read length, single ended) was performed by GATC biotech. Additionally we simultaneously ChIP-sequenced a *S. venezuelae rsrR::apr rsrR* 3xFLAG sample exposed to DETA-NONOate to investigate any change in target binding in the presence of NO. At the onset of this experiment we expected *rsrR* to be an *nsrR* homolog and wished to determine if the regulon was NO sensitive. The strain was grown under identical conditions with 50 µM DETA-NONOate added (the lowest concentration where no observable growth defects occurred). The sequencing files were processed as described (section 2.3.4.9.2) and visualised using IGB (Figure 4.6). CLC workbench 8 was used to carry out peak calling (Figure 4.6c).

The sequencing results for both samples (NONOate +/-) were analysed and resulted in near identical sequencing with identical data sets produced. This indicated, that at the concentration tested, RsrR does not respond to NO and as such was considered a replicate for peak calling. The processed ChIP-seq can be found in Supplementary data S4.4. Utilising CLC workbench 8 we identified 2696 peaks using the default settings with the WT control being subtracted from the dataset. For completeness all 2696 targets are included in Supplementary data S4.4 but a subset of the data is shown in Table 4.1. This includes targets chosen based on their annotation, those with a range of enrichments and targets that have been investigated further experimentally including the s*ven1847/8*, *sven3827/8* and *sven6562/3* promoters, which are the targets used for *in vitro* experiments later in this chapter.
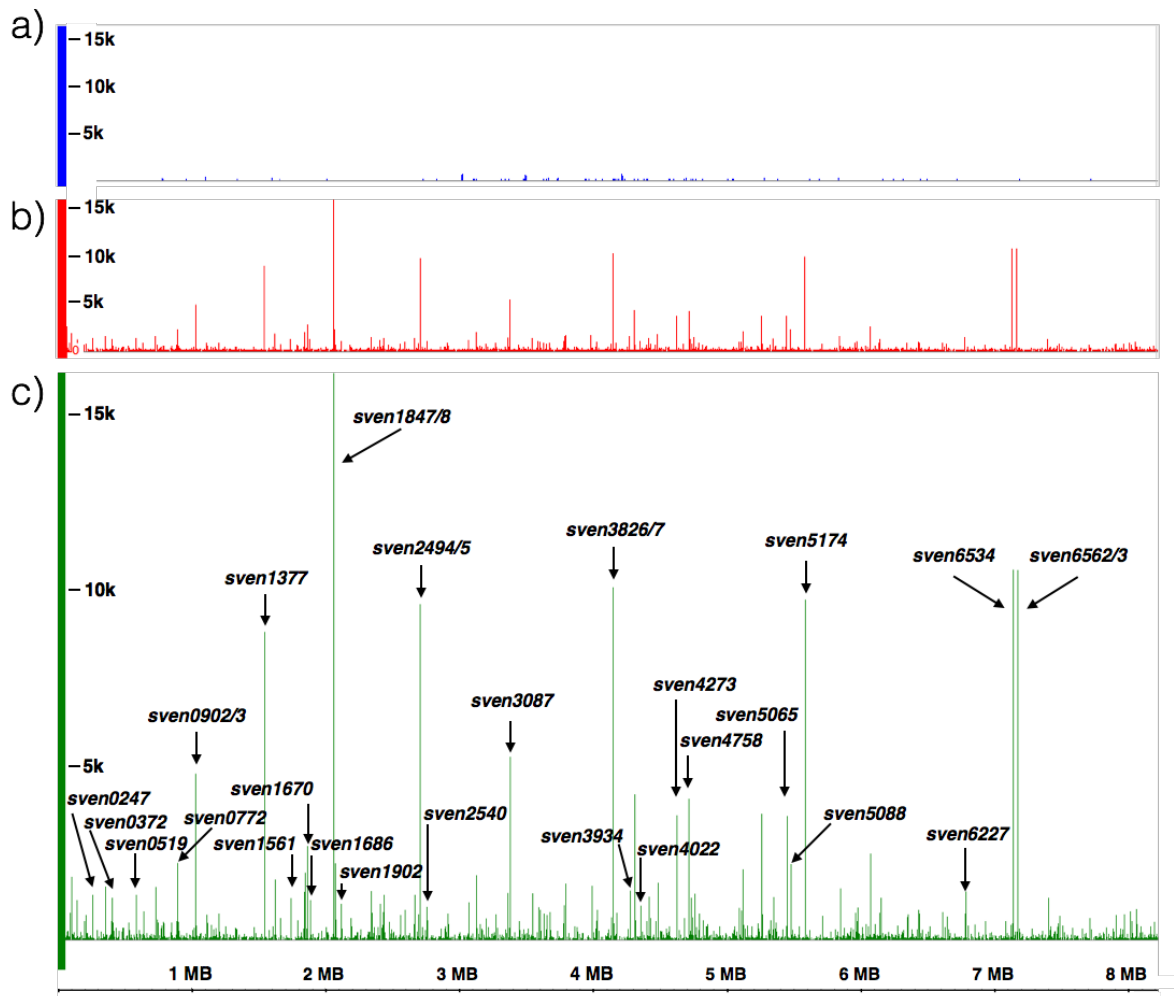
Figure 4-6 IGB visualisation of ChIP-seq results. (a) Wild-type *S. venezuelae* genomic coverage data (blue), (b) *S. venezuelae ΔrsrR* RsrR 3x FLAG (red) and (c) the subtracted data set to remove background results (green) are shown. Indicated along the subtracted data set are key targets from CLC workbench 8 peak calling results indicating the major represented peaks and many of those of interest through out this chapter, are indicated as visualised in IGB.

Flanking gene[a]		Distance[b]	Annotation
Left (-1)	Right (+1)		
	sven0247[e]	11	Universal stress protein
	sven0372[e]	7	Putative two-component system sensory histidine kinase
sven0519[e]		-993	Sulfate permease
sven0772		-408	Xaa-Pro aminopeptidase
sven0902	sven0903	413; -340	Hypothetical protein; Uracil-DNA glycosylase, family 1
	sven1377	29	Transcriptional regulator, DeoR family
	sven1561[e]	103	nodulin or glutamate-ammonia ligase protein
	sven1670	17	Pyridoxamine 5'-phosphate oxidase
	sven1686	-41	Citrate lyase beta chain
sven1847[e]	sven1848[e]	6; -64	3-oxoacyl-[acyl-carrier protein] reductase; hypothetical protein
	sven1902	-1643	Clutamine synthase adenyltransferase
	sven2494	90	Hypothetical protein
	sven2540	221	Glucose fructose oxidoreductase
	sven3087	51	Hypothetical protein
	sven3827[e]	26	SAICAR synthase
	sven4273	5	NADH-ubiquinone oxidoreductase chain I
	sven4759	429	Dipeptide-binding ABC transporter
	sven5065	-221	Putative transcriptional regulator, ArsR family metallo-re
	sven5088	-77	Epimerase/dehydratase
	sven5174[e]	-119	Quinone oxidoreductase
	sven6227	73	NADH-FMN oxidoreductase
	sven6534	-100	Hypothetical protein
sven6562[e]	sven6563[e]	72; -35	nmrA; rsrR

Table 4-1 Selected ChIP-identified RsrR targets (those highlighted in Figure 4-6).

[a] – Genes flanking the ChIP peak are listed

[b] – Distance to the translational start codon

The published *S. venezualae* genome (Pullan et al. 2011) contains 7455 genes with the CLC derived peaks comprising ~36% of these. Although possible we are concerned that this is unlikely in light of no obvious phenotype in the disruption mutant. We hypothesis that many of the weaker binding sites (those with a lower enrichment/abundance of reads) do not constitute legitimately regulated targets and although bound, do not bind strongly enough to modulate the regulation of the targets while still being visible through the ChIP-

seq experiment. As a result, we investigated the target binding sequences of all targets and in the process narrowing down the list to the most likely/strongly bound sites. We identified a DNA-binding sequence for the targets (next section) and use this, alongside read depth cut offs to begin analysis of the RsrR regulon gene repertoire.

4.2.2.2 Identification of a conserved binding sequence.

Following the identification of the RsrR DNA binding targets we wished to use the ChIP-seq data to define the RsrR binding site. To do this we utilised the MEME suite tools (Bailey et al. 2009). The main MEME utility requires a set of sequences (ideally in Fasta format) to search for conserved DNA motifs. The analysis report primarily includes information on the identified consensus motif, the sequences containing the motif and a score representing the relative match of each sequence specific motif to the consensus motif. Based on our previous work with ScoNsrR and from studying other Rrf2 protein DNA binding sequences on the RegPrecise database (Novichkov et al. 2009) we hypothesize that the RsrR binding sequence will form an imperfect inverted repeat. Certain Rrf2 proteins, however, have been reported to have half site binding capacity (potential monomer/apo-protein binding) and have been shown to regulate different classes of targets (Branchu et al. 2014; Partridge et al. 2009).

For this MEME analysis we produced several target sequence lists based on read depth cut off and target position from the ChIP-seq data. Each list used was analysed independently through MEME and results, description and cut off used along with the number of sites and type identified are summarised in Table 4-2. A MEME analysis of the 16 most highly enriched ChIP-seq targets that fall within 300 bp upstream of the gene start site (list 1), resulted in the identification of a 25 bp motif in 14 of the targets. From the data available we hypothesize that the sequence appears to be either a 9-7-9 bp or an 11-3-11 bp inverted repeat (Figure 4-7a), with the perfect inverted repeat consisting of AACTCGGACGGCGGGTGTCCGAGTT. On inspection of the binding site location 12 of these sequences fall almost exactly in the centre of the peak (the most highly enriched sequence), as would be expected. Of these targets, *sven1847/8 sven3826/7* and *sven6562/3* where used to illustrate this in Figure 4-8. Following this, each subsequent target list (Table 4-2, 2-4) analysed with MEME, however, reports a motif of ~11 bp which corresponds closely to one half of the inverted repeat sequence (Figure 4-8b-d). A visual inspection of the motif regions indicate that in some cases, but not all, there appears to be a full site sequence, however, one half appears to have lower conservation and as a result is not identified by the MEME algorithm.

List	Description	No. of sites[a]	Site type[b]	Method[c]
1	Read cut off – 1000, within 300 bp upstream	16/14/12	full	Website
2	Read cut off - 1000	58/44/42	half	Website
3	Read cut off - 200	696/295/257	half	Terminal
4	No cut off	2696/2696/1010[d]	half	Terminal

Table 4-2 Summary of MEME results using ChIP-seq targets

[a] – Number of targets tested in MEME/No. of binding sites identified by MEME/no. of sites located directly in the centre of a ChIP-seq peak

[b] – The type of site each list returned e.g. half or full binding site

[c] – The method used for data submission to MEME, web service or terminal based user service

[d] - Sites include those that do not fall in the centre but within the peak area, list not included in supplementary documents.
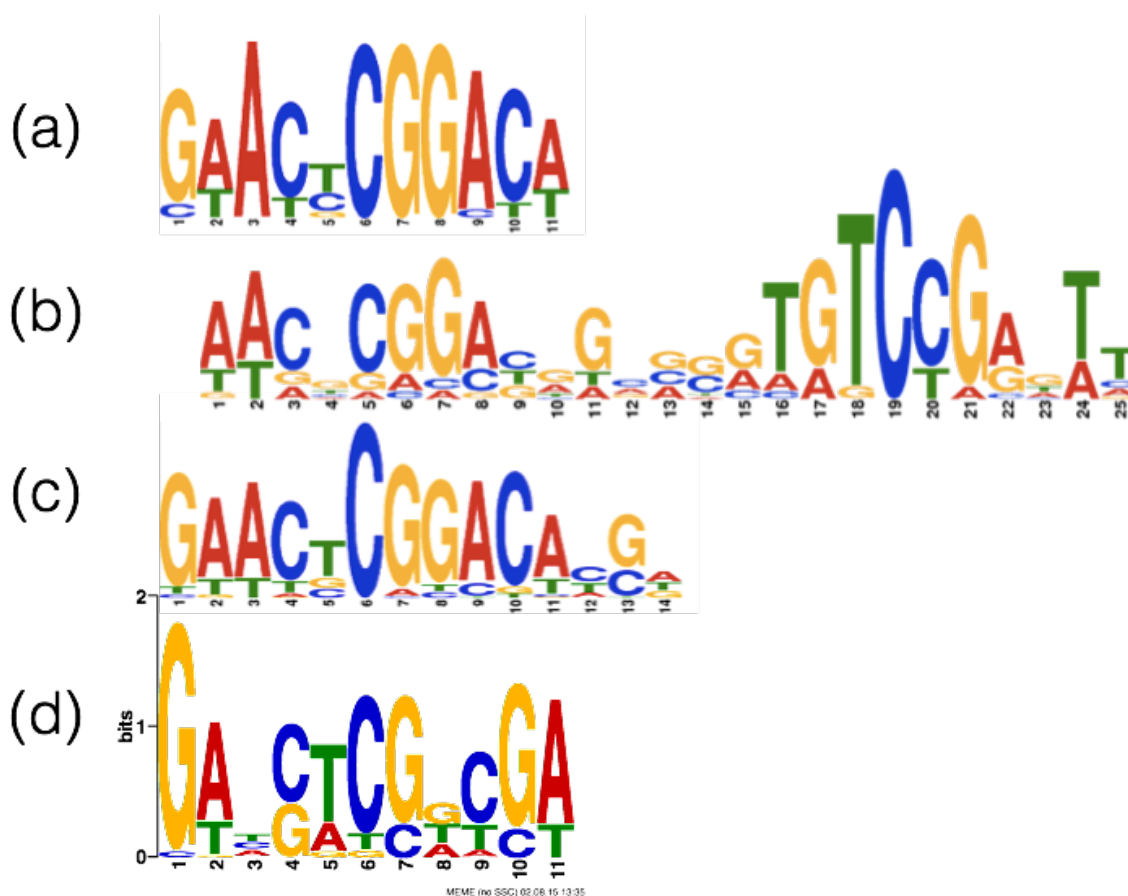
Figure 4-7 MEME identified DNA binding motifs from the RsrR ChIP-seq targets. These MEME consensus binding sequences were produced using target lists from the RsrR 3x FLAG ChIP-seq experiment, additional information available in Table 4-3. The motifs

correspond to list 1, the 16 targets above the read cut off of 1000 reads and within 300 bp of a gene start site (a), List 2, the 58 targets above 1000 reads (b), list 3, the 696 targets above 200 reads (c) or list 4, the 2696 targets using no read cut off (d). Each motif was produced with 14, 44, 295 and 2696 identified sequences respectively.
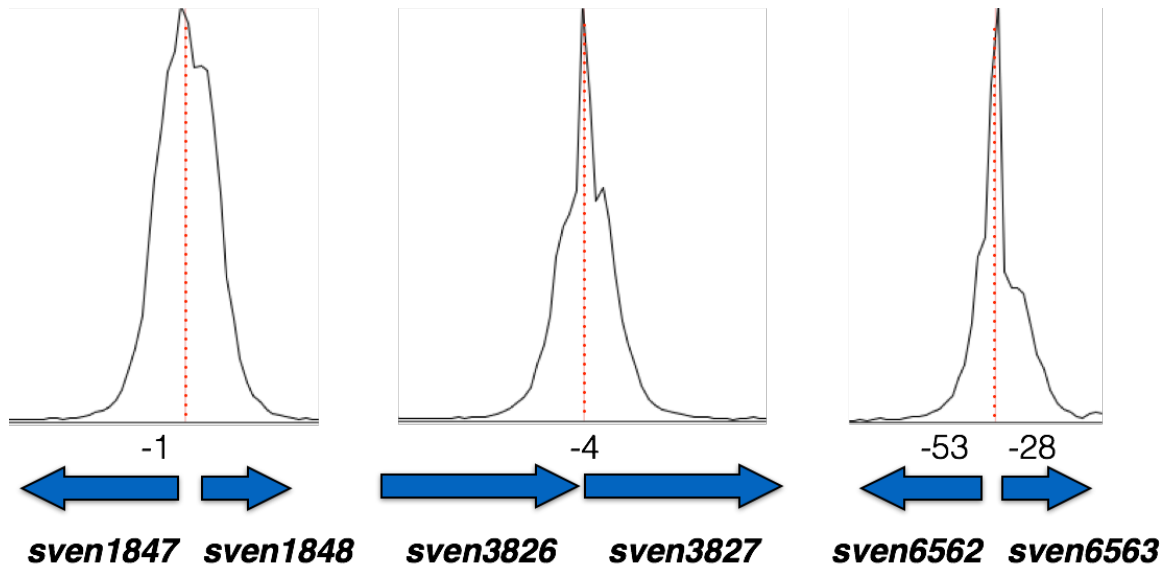
Figure 4-8 A graphical representation of the MEME identified binding motif location in regards to the ChIP-seq peak. ChIP-seq peaks (black line) showing where optimal MEME predicted motifs (red dashed-line) lies in regards to the gene (blue arrow) translation start site (numbers above arrow) for the three largest target peaks *sven1847/8, sven3826/7 and sven6562/3.*

We can see, in comparison particularly to our ScoNsrR binding site but also to several reported Rrf2 protein sites including *B. subtilis* CymR and *C. acetobutylicum* IscR (Figure 4.9) that our predicted binding sequence matches up well with an expected Rrf2 binding site. We hypothesize that either RsrR has two classes of binding site (class 1 – full site, Class 2 – half site) or that due to the reduced conservation half to half or within the centre of the sequence, MEME cannot accurately resolve full site motifs and indeed all sites are ultimately full-sites. In attempt to answer this, the binding site will be further investigated experimentally later in this chaptery EMSA study focusing on full-site vs. half site binding.
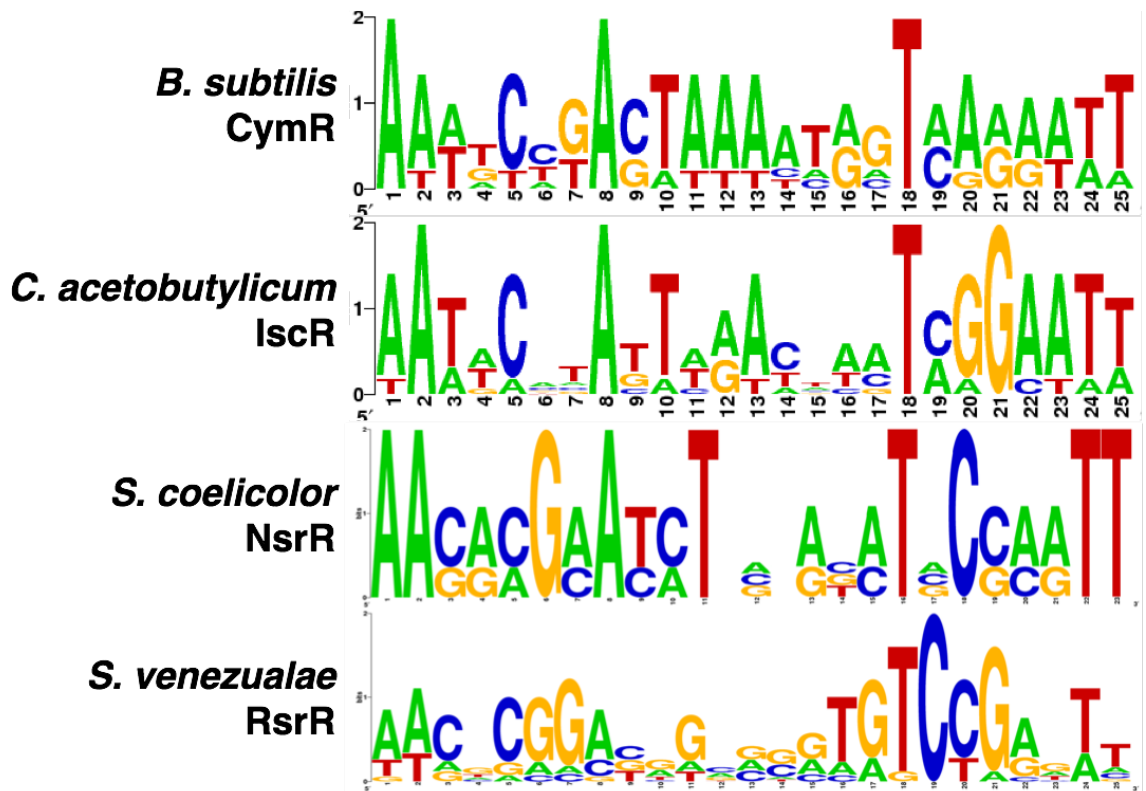
Figure 4-9 Weblogo representations of the NsrR and RsrR target sequences. The RsrR sequence corresponds to a 16 gene subset of the data referred to as list 2. Weblogos produced for *B. subtilis* CymR and *C. acetobutylicum* IscR from data acquired from the RegPrecise website.

4.2.2.3 RsrR ChIP-seq targets

To analyse the regulon function we investigated the 116 peaks above a 500 read threshold, primarily chosen as a manageable sample size. Fourteen of the ChIP-seq peaks have a set of divergent genes that have been included during the analysis. A manual annotation of the of these targets, primarily using the Pfam service (Finn et al. 2016), allowed us to elucidate at least in part, many of the gene functions and group them into broad overlapping categories summarised in Table 4.3. Each group will be discussed in detail below before an overall attempt is made to rationalise a function for this regulator.

Annotation	No.[a]	Description
No annotation	46	No significant annotation following Pfam analysis
Metabolism	22	Involvement in primary or secondary metabolism
Regulators	21	Annotated or putative transcriptional regulators
NAD/FAD	16	Enzymes associated with NAD/FAD binding, regulation, or redox
Amino acid	16	Protease enzymes or linked to amino acid metabolism
RNA/DNA	15	Enzymes involved in the modification of RNA and DNA
Transporters	13	Transporters and permeases, primarily small molecule
TCS	7	Parts of two-component systems, either the histidine kinase and or the sensory domain.

Table 4-3 Annotation of the core RsrR regulon genes into broad functional groups.

[a] – Number of targets in the sub group, annotations in some cases fall into multiple categories.

The largest category, and because of that, the most frustrating is the "No annotation" group, making up 46 of the targets. Based on the Pfam results, we were unable to assign a function for most of these targets either by overall protein structure or constituent domains. A minority however, did have functions, these simply were difficult to rationalise in *S. venezulae* e.g. *sven0630* has an annotation for a discoidin domain, which is major domain in many blood coagulation factors. As a result we have investigated them no further for the purposes of this section.

The "Metabolism" category (n=22) is the least specialised group and by nature the most thoroughly overlapping group with the others. Most of the genes within this group are categorised based on their specific enzymatic function being core to primary and secondary metabolism (e.g. *sven2540,* glucose fructose oxidoreductase) while having specific characteristics (e.g. *sven2540,* has a NAD-binding Rossmann fold) or inferred roles based on pathways/upstream/downstream products. Although not the most descriptive of grouping, it is useful until more is known about the global effect RsrR has on the cell.

The most exciting aspect of the results was the number of transcription factors (the "Regulator" category) that were identified (n=22). This, if true, indicates RsrR plays a larger role within the cell than previously hypothesised by being linked both directly and

indirectly to specific gene responses and expression. Many of these regulators are putative or hypothetical, however the list includes: SsgA, involved in regulating *ftsZ*, RsrR itself and the divergent gene *sven6562,* and regulators of the DeoR family, the ArsC family, PadR family, the AsnC family along with several sigma factor subunits. Of these, the most thoroughly studied is SsgA.

The next category we will discuss is the "NAD/FAD" category (n=16). The majority of these target genes contain $NAD^+/NADH$, $NADP^+/NADPH$ or $FAD/FADH_2$ binding domains or utilise these di-nucleotides during their enzymatic function. Some of these genes will be discussed in other section or only contain hypothetical links (e.g. contain the NAD-binding Rossman fold) but the category includes: *sven0675, sven1561, sven1847, sven2540, sven4022, sven4272, sven4455, sven5088, sven5173, sven5832, sven6227, sven6562, sven6563, sven6836, sven7195 and sven7248.* The first, *sven0675,* a D-beta-hydroxybutyrate dehydrogenase, carries out the reaction: (*R*)-3-hydroxybutanoate and NAD^+ to acetoacetate and NADH and H^+(Cortese et al. 1982). *sven1847,* a 3-oxoacyl-acyl-carrier protein reductase is involved in fatty acid biosynthesis utilising NADH/NADPH to reduce CH-OH bonds (Wickramasinghe et al. 2006). *sven4272* is a NADH-ubiquinone oxidoreductase, part of the respiratory chain forming the proton gradient across the membrane, utilises NADH to reduce the target acceptor ubiquinone (Friedrich et al. 2016). *sven4455*, carries out the reaction: inosine 5'-phosphate and NAD+ and H2O to xanthosine 5'-phosphate and NADH and H+ (Hedstrom 2009). *sven5173* is a quinone oxidoreductase. Recent work has associated a quinone oxidoreductase to the SoxR regulon (Naseer et al. 2014). Quinones are often parts of bioactive molecules being formed from the oxidation of aromatic compounds and are essential intermediates for efficient aerobic respiration. *sven6227*, NADH-FMN oxidoreductase, carries out the reversible reaction: $FMNH_2$ and NAD^+ to FMN and NADH and H^+. We have named *sven6562,* the divergent gene to *rsrR, nmrA* because it contains an NmrA-type NAD(P)+ binding domain which we suspect may be involved in sensing redox poise, perhaps sensing the NAD(P)+ pool within the cell similar to the Rex protein (Pagels et al. 2010). *sven6836*, the Succinate dehydrogenase flavoprotein subunit A, containing a FAD binding site and carries out the deprotonation of succinate to fumerate.

The "Amino acid" category (n=16) includes targets mostly made up of amino acid modifying enzymes (*sven0641, sven1561, sven1670, sven1902, sven3711, sven4418, sven4888, sven7195*) or proteases (*sven0772, sven0774, sven0979, sven3468 and sven6534*). The former group, which will be the major focus here, includes a range of enzymes involved in glutamate and glutamine metabolism and those involved with pyridoxal phosphate (PLP - the active form of vitamin B6). *sven0641,* an aromatic-L-

amino-acid decarboxylase, catalyses the decarboxylation of tryptophan to tryptamine utilising PLP (Kalb et al. 2016; Yuwen et al. 2013). *sven1561*, a nodulin or glutamate-ammonium ligase (or Glutamine synthase - GS) that carries out the ATP dependent conversion of glutamate and ammonium to glutamine (Gill & Eisenberg 2001). *sven1670*, pyridoxine 5'-phophate oxidase, involved in the vitamin B6 metabolic pathway (Mashalidis et al. 2011). *sven1902*, a glutamate-ammonia ligase adenylyltransferase that carries out the adenylation and deadenylation of GS, reducing or increasing GS activity respectively (P. Jiang et al. 2007). *sven3711*, which results in the liberation of glutamate, carries out the general reaction of: an aromatic amino acid and 2-oxoglutarate to an aromatic oxo acid and L-glutamate, with PLP as a cofactor (Huang et al. 2009). *sven4418*, a glutamine fructose-6-phosphate transaminase that carries out the reaction: L-glutamine and D-fructose 6-phosphate to L-glutamate and D-glucosamine 6-phosphate (Yamazaki 2014). *sven4888*, glutamate-1-semialdehyde aminotransferase, which carries out the PLP dependant, reversible reaction of L-glutamate to 1-semialdehyde 5-aminolevulinate (Grimm 1990). Finally, *sven7195*, a glutamine-dependent asparagine synthase which carries out the ATP dependent transfer of NH_3 from glutamine to aspartate, forming asparagine (Tesson et al. 2003).

The category, "DNA/RNA" (n=15) includes: *sven0142, sven0189, sven1604, sven2256, sven2397, sven3826, sven3848, sven3970, sven4076, sven4324, sven4393, sven4474, sven4749, sven5493, sven6761*. All targets within this section are linked to DNA or RNA metabolism, biosynthesis and or replication. *sven0142*, a Ribonuclease BN, is both an endoribonuclease and exoribonuclease (Dutta et al. 2012). *sven2256*, a DNA primase, is essential in facilitating the start of DNA replication as no known polymerase can start replication without a primer (Lewis et al. 2016). *sven3827*, a SAICAR synthase, is involved in purine metabolism (Manjunath et al. 2015). *sven3848, sven4474* and *sven5493* are ATP-dependent DNA helicases that facilitate the separation and unwinding of DNA (Lewis et al. 2016). *sven3970*, a rRNA methylase (SpoU family), are involved in aminoglycoside resistance (Anantharaman et al. 2002; Doi & Arakawa 2007). *sven4076*, an aminoacyl tRNA synthase (similar to the archael seryl-tRNA synthase), involved in the transfer of amino acids on to its cognate tRNA (Berg et al. 2002). *sven4324*, a Methylated-DNA-protein-cysteine methyltransferase, is involved in the cellular defence of O6-methylguanine (methylated guanine) which can cause base pair transitions, often as a result of N-nitroso compounds (Miggiano et al. 2016).

When we consider each of these categories independently we see NAD utilising enzymes, enzymes linked to glutamate and glutamine metabolism, PLP utilising enzymes, various DNA processing enzymes as well as proteases and various small molecule transporters.

This broad functional diversity of targets, the strong links to important cellular resources, the large number of regulators, along with our previous knowledge of Rrf2 proteins and there links to stress responses, we suggests implicates RsrR as a potential stress response regulator. At this stage it is still unclear what this "stress response" might be and will be discussed further in the discussion section of this chapter. Following the ChIP-seq analysis the next aspect of this project was to determine the transcriptional response of *S. venezualae* to the loss of *rsrR*. To do this we carried out dRNA-seq on the wild-type and *rsrR* mutant strain.

4.2.2.4 dRNA/RNA-seq analysis in an RsrR background.

Following the expansive ChIP-seq results, we aim to further characterize the RsrR regulon using transcriptome data acquired through RNA-seq. RNA-seq is a process where the total RNA pool is ultimately sequenced to elucidate the genome wide expression profile of a sample. dRNA-seq (differential RNA sequencing) is a form of RNA-seq that facilitates the identification of TSS and non-coding RNA's, while still allowing for expression quantitation, by differentiating between primary and processed transcripts (Sharma et al. 2010). Based on the recommendations of the sequencing provider (vertis-Biotechnologie AG) the process was modified from the Sharma et al 2010 method and is illustrated in Figure 4.10. The cellular RNA pool consists of primary and processed transcripts, which contain 5' triphosphates (5'PPP) and 5'monophosphates (5'P) respectively. Total RNA is first isolated, followed by ribodepletion to remove the predominant and ultimately interfering rRNA. The sample is split in two with both samples treated with 5'P-dependant terminator exonuclease (Tex), degrading the processed transcripts. Following this a TAP (Tobacco Acid Pyrophosphatase) treatment is carried out on one half, producing a TAP(+) and, the second half, a TAP(-) sample. The TAP treatment converts 5'PPP ends into 5'P ends. RNA adaptors, which facilitate specific cDNA synthesis, are added which will only bind to 5'P transcripts. cDNA is synthesised to facilitate NGS of the samples. We have additionally used the TAP(-) sample data for expression profiling. Library construction was carried out by vertis Biotechnologie using commercially available kits that were kept confidential.

Total RNA

|

Ribodepletion

Addition of poly-A-tails

Tex treatment

TAP(+)   TAP(-)

Adaptor ligation to only 5'P

cDNA synthesis

Next Gen. Sequencing

DATA: Data:
5'PPP + 5'P   5'P
transcripts   transcripts
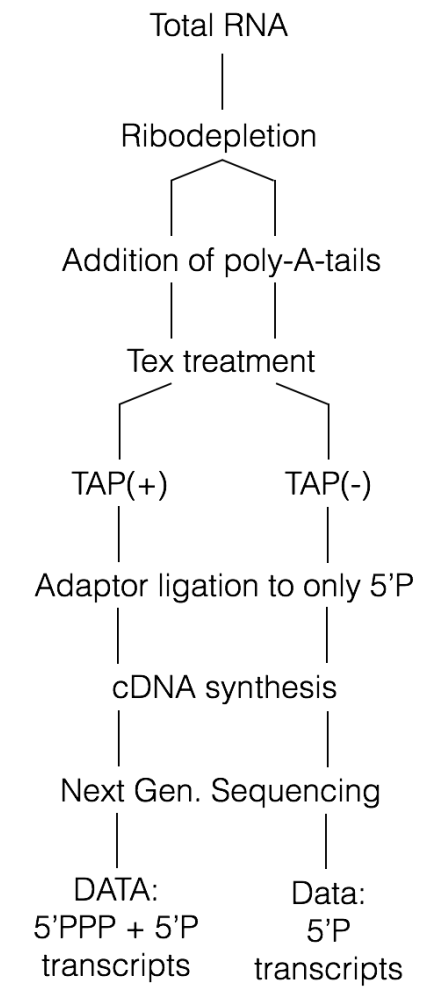
Figure 4-10 A summary of the dRNA-seq sample processing pipeline until sequencing data is produced. Each step is carried out with branch points indicating where the sample is split during processing. The TAP(+) sample (left route) ultimately results in data containing both 5'PPP and 5'P data where 5'PPP transcripts are enriched, where as the TAP(-) samples (right) contains no 5'PPP and only 5'P data.

Expression analysis of the data was carried out using the Tuxedo suite of software (as described). TSS analysis was carried out using the TSSAR web service (Amman et al. 2014). All data sets were processed for visualisation in IGB, as previously described (section 2.3.4.9.2). The entire data set can be found in the Supplementary data 4.4. To begin the analysis of the RNA-seq data we produced M/A plots to assess any over all trends in the data, determine if there is any specific visible biasing of the data and to map and correlate selected ChIP-seq targets (Figure 4.11) using the associated gene identifier number (those listed in Table 4-1) before processing the RNA-seq data independently. M/A plots, those showing distribution of intensity or fold change (M) by the average intensity (A). The M/A plots were produced with M and A values corresponding to:

$$M = \log_2(\text{mutant/WT}) \qquad A = 0.5\log_2(\text{Mutant} \times \text{WT})$$

Where Mutant and WT corresponds to the number of reads mapped to a specified genetic loci from each of the test strains *S. venezulae rsrR::apr* and *S. venezualae* respectively. We would expect, and observe, that much of the data would centre on a value of M=0. This is due to expectation that much of the transcriptome is unaffected.
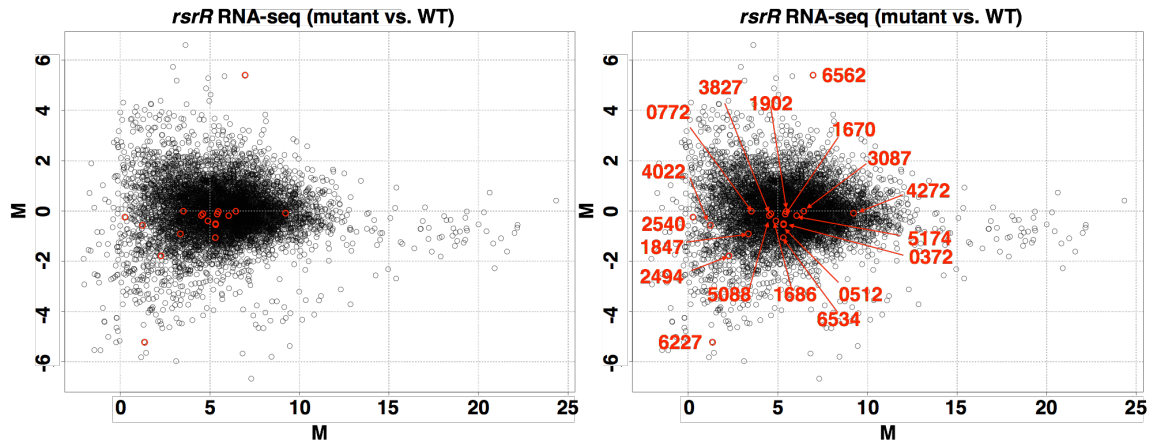
Figure 4-11 MA plot of dRNA-seq expression data and selected ChIP-seq targets. M = $\log_2(\text{mutant/WT})$ and A = $0.5(\log_2(\text{Mutant} \times \text{WT}))$. Each data point corresponeds to a single gene. The left and right panel contain the same data with red circles indicating points of interest highlight with the *svenxxxx* identifier.

Much of the data as expected centres along M=0. Of the targets indicated in Figure 4-11, we see that almost all fall within the $\log_2$ 2-fold range. To assess genes with significant fold changes, we chose several conditions, each are indicted in Table 4.4. These include a fold change cut off of 1 or 2 (up and down regulation) and a read count cut off (112) based on the observation that many of the largest fold changes occur in targets with low read counts. The read cut off of 112 was chosen, this being closest value to 100 reads achievable using 0-decemial place integers in the Tuxedo suite processing.   §

Two target lists were used in the comparison of RNA-seq and ChIP-seq, the short list discussed in the previous section corresponding to the unique, highest enriched ChIP targets above a 500 read cut off (129) and all uniquely identified targets from the data set (2273). When comparing the refined target list (129 targets) to the RNA-seq data, we see limited correlation, the least stringent cut off containing only 28.7% of the targets. At the most stringent cut off, this falls to 3.1%. A similar trend is observed with the full target list. Based on this observation, it is possible that the ChIP-seq results include targets that are not legitimately regulated by RsrR as has already been discussed. However, considering that the strongest bound (most highly enriched) targets do not appear to be within those regulated, we hypothesize that it is more likely there are multiple levels of regulation several putative regulators. This has led to two routes of investigation: a standalone

analysis of the RNA-seq data to try and determine the direct effect of losing RsrR and a targeted approach to attempt to understand why only some of the ChIP-seq "targets" are modulated, the latter being a discussion of the functions of the regulated genes. Before proceeding with this analysis, we will briefly discuss the TSS data produced.

Fold change	Read cut off	RNA-seq targets	Correlated RNA-seq targets to ChIP-seq targets with percentages to (RNA-seq/ChIP-129/ChIP-2273)	
			Targets = 129[a]	Targets = 2273[b]
<-1 and >1	-	2107	37 (1.8% / 28.7% / 1.6%)	625 (29.7% / 27.5%)
<-1 and >1	112	676	10 (1.5% / 7.8% / 0.4%)	196 (29.0% / 8.6%)
<-2 and >2	-	552	12 (2.2% / 9.3% / 0.5%)	166 (30.1% / 7.3%)
<-2 and >2	112	164	4 (2.4% / 3.1% / 0.2%)	52 (31.7% / 2.3%)

[a] – Target list contains the unique refined ChIP-seq encriched above a 500 read cut off.

[b] – All unique targets identified through ChIP-seq.

Table 4-4 A summary of the combined RNA-seq/ChIP-seq data. Fold change and read depth cut offs were used as indicated in columns 1 and 2 with the resulting number of targets indicated in number 3. The correlated number of RNA-seq targets to the refined ChIP-seq target list is shown in column 4 with this as a percentage of RNA-seq/ChIP-seq (129)/ ChIP-Seq (2273) targets. The correlated number of RNA-seq targets to the total ChIP-seq target list is shown in column 5 with this as a percentage of RNA-seq / ChIP-Seq (2273) targets.

4.2.2.4.1 TSS identification in WT and *rsrR* mutant backgrounds

TSS analysis was carried out using the TSSAR web service (Amman et al. 2014) implementing the default settings. TSS are categorised into 3 major groups. **P**rimary, those within 250 bp of the translational start codon of a gene, **I**nternal, those existing within the coding region of a gene and **A**ntisense, those either inside or within 30 bp of the stop codon of a gene, on the opposite strand. Those that do not fall within these 3 categories are denoted as **O**rphan sites. Based on the data, we were able to identify a range of TSS, summarised in Figure 4.12. This data is available in supplementary data 4.4
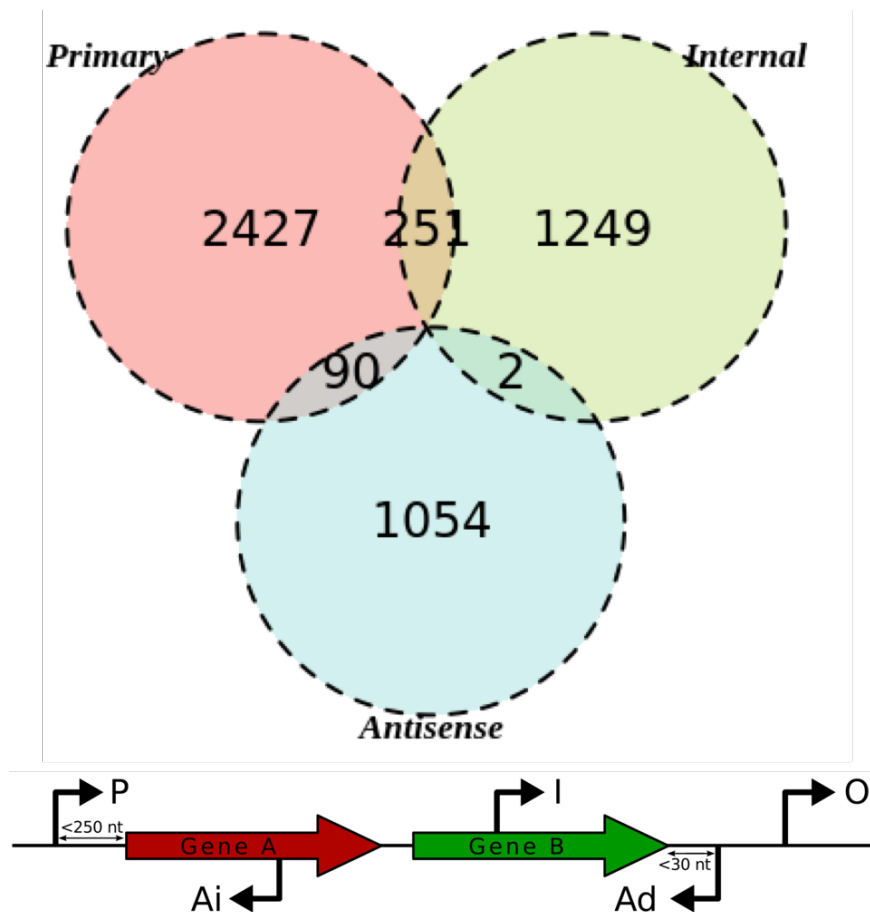
Figure 4-12 Visualisation of the TSSAR defined TSS. A Venn diagram was used to show the resulting distribution of TSS within *S. venezualae* with a graphical indication of how the different positions were defined below. P = primary, I = internal, O = orphan, Ai = antisense and Ad = antisense downstream. This diagram was produced as part of the TSSAR analysis of the data (Amman et al. 2014) using the direct image to represent the TSS positions.

When investigating the location of the TSS, in regards to the translational start codon of the down stream gene, we see a high proportion of short to leaderless transcriptions (Figure 4.13). We observe two predominant distribution of TSS at 0-4 bp from the translational start and between 30-38 bp, similar to the results published for *S. coelicolor* (Jeong et al. 2016). This will allow for the identification of operonic units through out the genome and has been made available online (GEO accession number GSE81104). In addition to this, the data allows for the identification of putative small RNAs (sRNA). A superficial investigation with putative sRNAs defined primarily by the presence of sequence within a coding region, on the opposite strand and have a reported p-value of 0 resulting in identifying ~776 putative sRNA. Often putative sRNAs are difficult to experimentally define. Commonly qRT-PCR is employed to do this (Gaimster et al. 2016) however using Northern blotting has been suggested to be the most appropriate route.
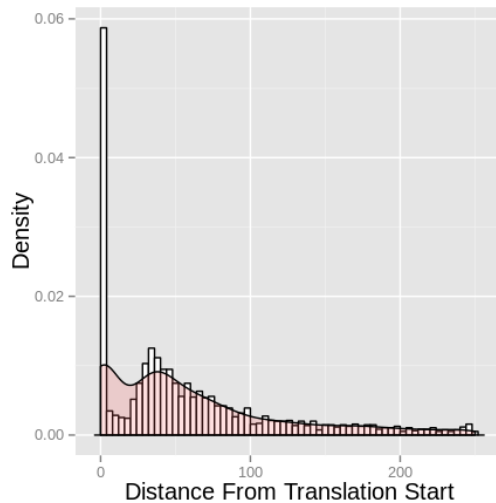
Figure 4-13 A plot of TSS locations relative to the translational start site. The plot represents the relative proportion of TSS (y-axis) against the distance in bp from the translational start codon (x-axis). This plot was produced as part of the TSSAR analysis of the data (Amman et al. 2014).

When investigating the TSS for *sven6562*, we see an interesting feature. Upstream of the TSS there is a leading tail of transcription associated, however, this does not appear to be enriched in the TAP treated samples. We indeed see this quite commonly throughout the sequencing. We suspect that a small proportion of those reads associated with *sven6562*, may in part, be caused by transcription from the upstream apramycin resistance cassette inserted within the *sven6563 (rsrR)* loci. The presence at other sites, however, is more difficult to explain. Relative to the TSS, these points have a very a low read count. It is unclear what these sequences represent. It is possible they correspond to TSS that are not highly transcribed under these conditions but we think this unlikely due to its global nature. Its possible that they are caused by very small amounts of DNA contamination, however, we think this also unlikely, as the sequence is associated with the TSS specifically. We do see sites throughout the genome, especially in the TAP positive sample of small, low count points of sequence. It is difficult to rationalise what these points may be, especially, following the use of a relatively unique treatment process. As a result, we have focused on TSS identified by TSSAR. We may be able to shed light on the biological significance or, indeed a methodological rationale if they are artefacts of the process.

4.2.2.5 The RsrR regulon: RNA-seq and ChIP-seq data

Upon investigating the ChIP-seq targets, specifically the binding region of RsrR, especially focusing on the RNA-seq targets that have modulated expression, we do not observe

fluctuations in transcription specifically around the RsrR binding sites. We would expect, if RsrR were acting as a repressor, to observe a sharp rise in transcription following the RsrR binding site. The two methods considered for RsrR repression involve either a "roadblock" method, where RsrR physically blocks the advancement of RNA-polymerase, or RsrR occludes the binding of other transcription factors responsible for target gene expression. If RsrR was solely acting as an activator, we would expect to see a universal drop in transcription at target sites. In both cases we would expect transcription to be controlled tightly around the binding site. We do not clearly observe this. An assumption made at this point is that RsrR solely regulates target genes. If RsrR co-regulates the regulon along side other regulators, this may explain the lack of observed transcriptional fluctuation.

With that mind, a striking feature of the RNA-seq data comes from *sven6562,* the divergently transcribed gene from *rsrR,* encoding a LysR family regulator. *sven6562* is upregulated 5.4 fold (log$_2$) the Δ*rsrR* background. Following the MEME analysis we were able to identify two full binding sites within the *sven6562/3* intergenic region with a 1 bp overlap. We suspect that each full site regulates one of the divergent genes with RsrR regulating its own expression (as is the case for ScoNsrR). We have named *sven6562, nmrA* because it contains an NmrA-type NAD(P)+ binding domain which we suspect may be involved in sensing redox poise, perhaps sensing the NAD(P)+ pool within the cell similar to the Rex protein (Pagels et al. 2010). However due to time restrictions, *nmrA* has not be investigated experimentally.

Looking at the data with a 1 or 2 fold cut off with the 112 read cut off for the data (Table 4.4) without correlating the data to the ChIP-seq targets we have 676 and 164 targets respectively that are either up or down regulated (available in Supplementary data S4.4). Immediately upon looking at the annotations we see the expected high percentage of hypothetical proteins (41.1%/45.1%) with another substantial portion of the annotated genes being of putative function (12.3%/9.8%). Within the remaining list we see several transcription factors/putative regulators (*sven0624* and sven5957- Xylose repressor XylR, *sven1046* - ROK-family, *sven1332* - GntR-family, sven2720, sven3606 - Leucine-responsive regulatory (*lrp*), sven3945 - AraC family, sven5852 - LuxR family). In addition we see genes involved in xanthine metabolism (*sven5952-6*), the phenylacetate-CoA oxygenase (*sven3613-7*), protein degradation (*sven1650* - Trypsin protease precursor, *sven3837* - secreted peptidase, *sven4769* - Oligopeptide transport ATP-binding protein OppF, *sven5514* - putative peptidase) to name a few. The lower fold cut off results in a list that includes several glutamate and glutamine associated enzymes (*sven0203* - glutaminase, *sven0679* - glutamate racemase, *sven1676, sven1677* - glutamate synthase,

sven1863 - glutamine synthetase, *sven5431* - glutamate binding periplasmic protein, *sven6870* - Glutamate--cysteine ligase) and NAD and FAD utilising or binding proteins (*sven0215* - NADH dehydrogenase, *sven1576* - NAD-dependent glyceraldehyde-3-phosphate dehydrogenase, *sven4308* - NADH ubiquinone oxidoreductase chain A, *sven1412* - oxidoreductase, FAD-dependent protein, *sven5952* - Xanthine dehydrogenase, FAD binding subunit).

We hypothesize that *rsrR* and *nmrA* are functionally linked and, in some way, regulate a shared set of genes, primarily due to the high level of expression of *nmrA* in the disruption strain and its putitive links to utilising NAD as a sensing molecule. As a result we wished to investigate if the two genes are genetically linked in other bacteria.

4.2.2.5.1 Conservation of the putative RsrR regulon and NmrA

To investigate briefly whether the link between NmrA and RsrR correlates with biological function, we investigated how frequently the genes are conserved in other strains. A range of BLAST analyses identified a high similarity to many IscR annotated homologues from phylum including Actinobacteria, Frimicutes and Proteobacteria. We also carried out delta BLAST searches, looking at domain conservation instead of sequence conservation. We investigated 14 putative RsrR containing strains that constitute the best Acintobacteria that did not return direct IscR hits. From this we carried out a BLAST analysis comparing each of the putative ChIP-seq targets with greater than a 500 read cut off to determine if the regulon targets were present in these other strains. The resulting heat map table can be found in Supplementary data S4.5 with an excerpt of the data in Figure 4.14 *rsrR/nmrA* are co-conserved in each strain tested with *S. flavochromogenes* containing the highest number of conserved targets of the regulon.

	Amycolatopsis decaplanina	Amycolatopsis orientalis	Catelliglobosispora koreensis	Kitasatospora NRRL B-11411	Kitasatospora phosalacinea	Micromonospora parva	Streptacidiphilus rugosus	S. 303MFCol5.2	S. ATexAB-D23	S. avermitilis	S. flavochromogenes	S. HmicA12	S. NRRL F-5123	S. TAA040	S. venezuelae ATCC 10712	
SVEN_2494	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	1.00
SVEN_4758	0.66	0.66	0.75	0.91	0.91	0.76	0.87	0.92	0.85	0.94	0.98	0.95	0.87	0.83	1	0.86
SVEN_3826	NaN	NaN	0.3	0.7	0.7	NaN	0.43	0.84	0.91	0.82	0.97	0.85	0.78	0.81	1	0.77
SVEN_4272	0.55	0.55	0.53	0.78	0.77	0.52	0.76	0.85	0.89	0.87	0.96	0.85	0.78	0.72	1	0.77
SVEN_3970	0.62	0.62	0.47	0.68	0.7	0.49	0.7	0.81	0.82	0.81	0.97	0.82	0.8	0.8	1	0.75
SVEN_6563	0.72	0.72	0.73	0.84	0.83	0.72	0.83	0.72	0.81	0.74	0.93	0.84	0.74	0.72	1	0.74
SVEN_1561	0.41	0.41	NaN	NaN	NaN	NaN	NaN	0.76	NaN	0.76	0.92	0.74	NaN	NaN	1	0.71
SVEN_6562	0.66	0.67	0.45	0.79	0.79	0.67	0.76	0.68	0.76	0.64	0.86	0.81	0.73	0.63	1	0.68
SVEN_0902	0.5	0.53	NaN	NaN	NaN	0.5	NaN	0.63	0.75	0.71	0.98	0.61	NaN	0.45	1	0.67
SVEN_6534	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.01	NaN	NaN	0.89	NaN	NaN	NaN	1	0.63
SVEN_1847	0.49	0.63	0.01	0.62	0.69	0.64	0.58	0.67	0.01	0.01	0.89	0.65	0.59	0.67	1	0.55
SVEN_0519	0.31	0.01	0.28	0.28	0.79	0.3	0.27	0.85	0.84	0.01	0.88	0.84	NaN	0.73	1	0.49
SVEN_0372	0.01	0.68	0.01	0.7	0.74	0.01	0.73	0.73	0.01	0.77	0.95	0.01	0.01	0.01	1	0.40
SVEN_0372	0.01	0.68	0.01	0.7	0.74	0.01	0.73	0.73	0.01	0.77	0.95	0.01	0.01	0.01	1	0.40
SVEN_1377	0.56	0.56	0.58	0.01	0.01	0.01	0.01	0.01	0.56	0.01	0.94	0.01	0.61	0.65	1	0.39
SVEN_5065	0.42	0.01	0.01	0.01	0.37	0.38	0.4	0.64	0.69	0.01	0.88	0.68	0.58	0.01	1	0.38
SVEN_3087	0.01	0.01	0.01	0.54	0.53	0.59	0.55	0.57	0.01	0.01	0.63	0.54	0.53	0.01	1	0.35
SVEN_5173	0.44	0.45	0.01	0.01	0.01	0.01	0.01	0.49	0.6	0.01	0.01	0.66	0.63	0.49	1	0.33
SVEN_0247	0.01	0.01	0.01	NaN	NaN	0.01	0.01	0.77	NaN	NaN	0.94	NaN	0.01	NaN	1	0.31

Figure 4-14 A heat map display of 14 delta BLAST identified actinomycete strains containing a putative *rsrR* gene comparing part of the core regulon of *S. venezuelae* RsrR to 14 RsrR containing strains. On a scale of 0-1 (red to green) representing percentage conservation 1 representing 100% and NaN representing strains lacking the specific gene.

4.2.3 RsrR biochemistry and *in vitro* studies

To further our understanding of what RsrR is and the processes it regulates, biochemical experiments on the purified protein were undertaken. We aimed to define the multimeric state of the protein determine if it binds a co-factor and to carry out *in vitro* DNA-binding assays to confirm target genes from the ChIP-seq experiment, determine binding sites and the functional state of the protein. Finally we wished to use the purified protein to raise polyclonal antibodies against RsrR, upon completion however these antibodies ultimately provided less specificity than the FLAG antibodies, so have not been used for subsequent rounds of experiment.

Several constructs were produced to address different biochemical questions and to facilitate purification: 1) a wild type RsrR expression construct with no tags (pJM028), (2) an RsrR expression construct with a C-terminal hexa-His tag (pJM029), (3) an RsrR expression construct sharing the linker from the FLAG construct, a 2xFLAG tag and a hexa-His tag (pJM030) and finally (4) an RsrR expression construct with all three C-terminal cysteines: C91A, C108A, C112A (putative [Fe-S] ligands) changed to alanine (pJM031). Constructs 1 and 2 were produced and purified by our collaborators, Nick Le Brun and Maria Pellicer Martinez in Chemistry at UEA, whereas 3 and 4 were produced as part of this work. Constructs were synthesised by Genscript and subcloned into pGS-21a. Purification was carried out as described in section 2.2.4.7.2.

4.2.3.1 Purification of RsrR-His and RsrR-3Cys-Ala-His

Purification of wild-type RsrR utilising pJM029 under aerobic conditions resulted in the isolation of a reddish brown solution whereas purification of RsrR C91A, C108A, C112A (pJM031) resulted in a clear protein solution. The reddish brown colour is indicative of an [Fe-S] containing protein, similar it ScoNsrR (Crack et al. 2015). EPR and UV visible spectroscopy carried out by our collaborators showed that *rsrR* contains a [2Fe-2S] cluster that can switch between oxidised and reduced states in the presence and absence of oxygen, respectively (Munnoch, Teresa, et al. 2016) Anaerobic purification of RsrR results in its reduced form being isolated, exposure to oxygen results in its oxidation. Oxidised RsrR (RsrRox) can be reduced (RsrRred) by the addition of reducing agents (such as dithionite) with no obvious/dramatic loss of cluster or precipitation of the protein based on UV/visible spectrum. The reduced and oxidised forms of RsrR have clearly distinct UV-Visible spectrums (Figure 4.15). Purity was confirmed by resolving on SDS PAGE followed

by coomassie staining (Figure 4.16). Identity was confirmed by mass-spectrometry (Jason Crack unpublished) and trypsin digestion MALDI-TOF. Protein and iron concentrations were determined and the ratio of [2Fe-2S] cluster to RsrR monomer was determined. Wild-type RsrR-His was purified with a 41% load of [2Fe-2S] cluster per monomer but no iron was detected in the RsrR C91A, C108A, C112A His sample.
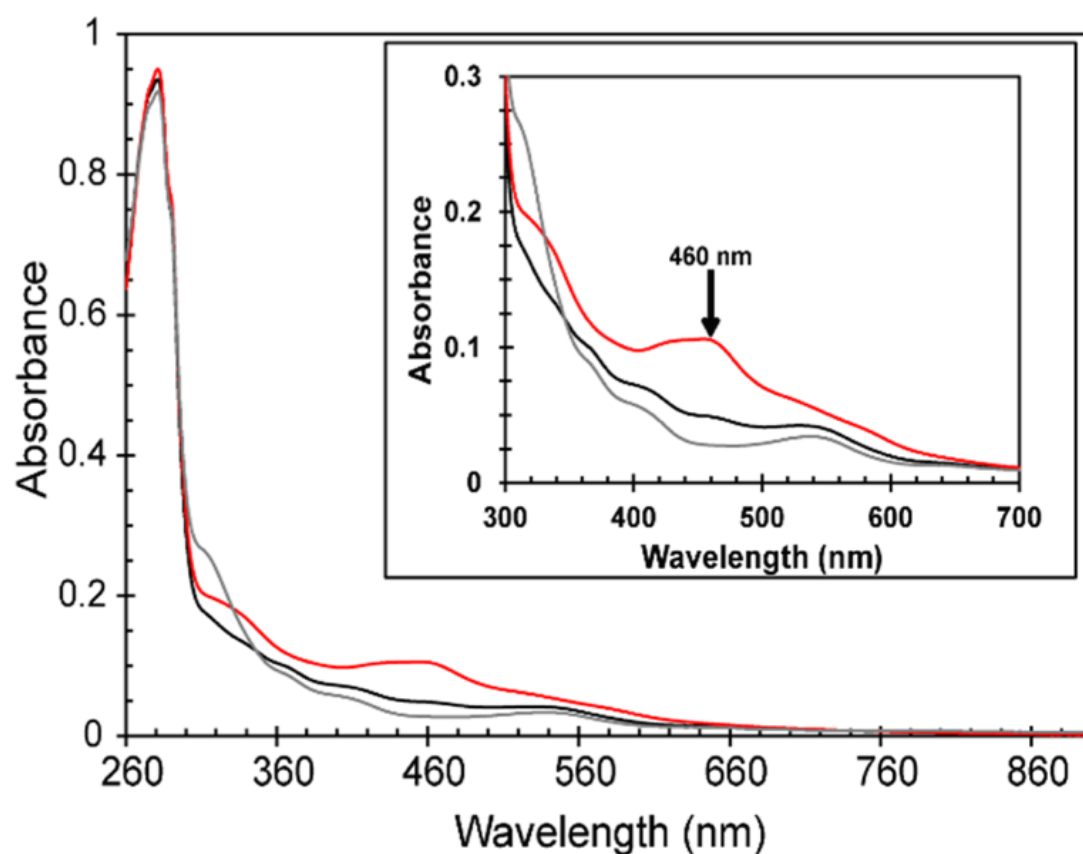
Figure 4-15 UV-Visible spectrums RsrR. Spectrum show RsrR under oxidising (black) and reducing (red) conditions. Data produced by Maria Pellicer-Martinez.
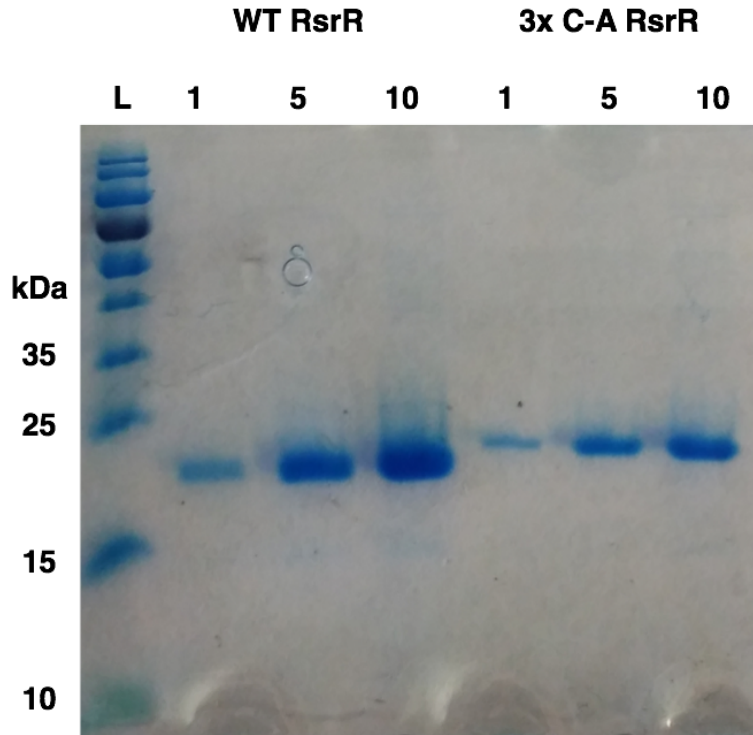
Figure 4-16 SDS-PAGE gel analysis of purified RsrR proteins by coomassie staining. The left sample contained in the gel represent wild-type RsrR-6xHis (WT) and on the right RsrR-3CA-6xHis (3xC-A). Volumes, 1, 5 and 10 µl were added, Prestained PAGE ruler was used as a size marker (Biorad).

To summarise the data produced by our collaborators during the study, RsrR exists as a dimer in solution (analytical gel filtration) and contains a 2Fe-2S cluster (identified based on UV-vis absorbance (ref) and confirmed by EPR) that can redox cycles between oxidised, with a brown colouration and reduced with a pink colouration (Pellicer-Martinez and Le Brun personal communication).

4.2.3.2   EMSA analysis of RsrR target sites

To determine if the purified protein was functional and to validate gene targets identified by the ChIP-seq, we carried out electrophoretic mobility shift assays (EMSAs). We focused on three targets identified by ChIP-seq: *sven1847/8p, sven3827p* and *sven6562/3p,* the RsrR encoding gene. These targets were selected because they contained the largest peaks identified in the ChIP-seq experiment. Following MEME analysis of binding sites we predicted that each of these promoters contains a "full-site" inverted repeat sequence while *sven6562/3p* contains two "full-sites" as previously discussed. A comparison of the predicted binding sequences can be found in Table 4-4. EMSAs were carried out as

described in section 2.2.4.8.1).   We were able to detect RsrR binding to each target (Figure 4.17) and RsrR bound them all under similar conditions and concentrations. The *sven6562/3p* probe shows two shifted bands consistent with the presence of two predicted RsrR binding sites.   We show that apo-RsrR is unable to shift the *sven1847p* that suggests that the cluster bound form is essential for DNA-binding.

Gene/Name	Binding site sequece
sven1847	AAACCAGACAGAAGATGTCTGATTT
sven3827	AAACCAGACAGAAGATGTCTGATTT
sven6562/3 (1)	ATCTCGGACATCGGGTATCCGAGTT
sven6562/3 (2)	TACTCGGATAGTCTGTGTCCGAGTC
Conserved	.---*-**.*.--.-*.**-**-*.
Perfect	AACTCGGACGGCGGGTGTCCGAGTT

Table 4-5 The binding sites of each of the major targets tested comparing conservation of bases to each other and a perfect inverted repeat proposed in section 4.2.2.2.   The numbering corresponds to the frequency of conservation between the 4 sites (1-4). Stars (*) represent bases conserved in all 4 sites, periods (.) in 3 and dashes (-) for 2. (1) and (2) represent each full site, each corresponding to either *rsrR* or *nmrA*.
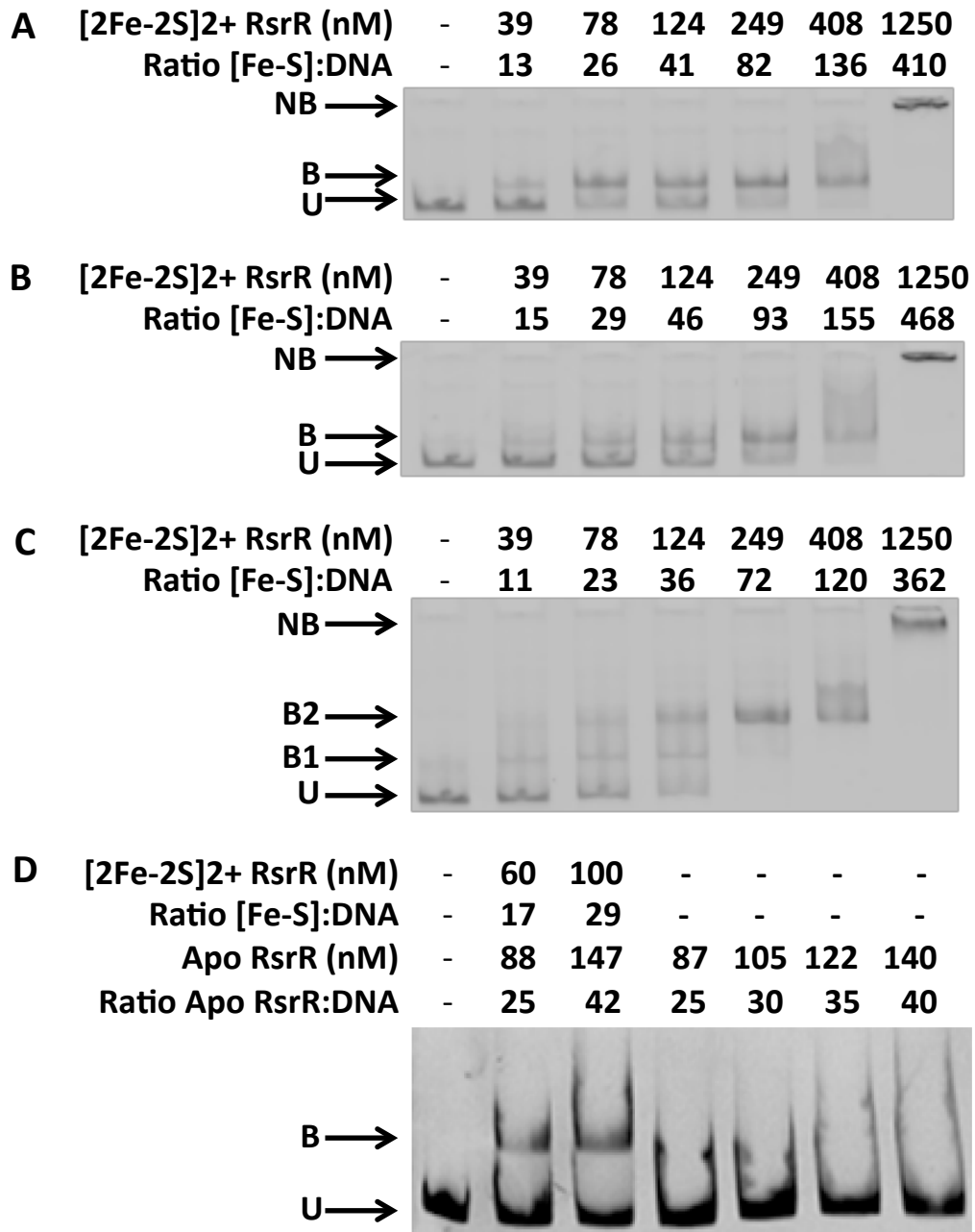
Figure 4-17 Cluster dependent binding of RsrR to RsrR-regulated promoter DNA. EMSA reactions using ChIP-seq identified target DNA probes *sven1847/8* (A and D), *sven3826/7* (B) and *sven6562/3* (C) showing unbound (U) bound (B) and non-sepcific binding (NB). Panels A-C are carried out using protein partially incorporated with $[2Fe-2S]^{2+}$ cluster (~41%) while section D shows both $[2Fe-2S]^{2+}$ (~41%) and apo results at comparible concentrations as a negative control. Probe concentrations were (A) 3 nM, (B) 2.6 nM (C) 3.5 nM and (D) 3.5 nM respectively

for *sven1847/8p*, *sven3826/7p*, *sven6562/3p* and *sven1847/8p.* Data in section (D) was produced by Maria Pellicer-Martinez.

MEME analysis identified putative half (class 1) and full (class 2) binding sites at target promoters. It is unclear that the half sites are real, rather than partially called full sites. Subsequently, we investigated if RsrR can bind to half sites. To do this we carried out EMSA reactions using probes for 2, class 2 targets (*sven0247* and *sven0519*) and probes consisting of combinations of the 4 detectable half site sequences within the *sven6562/3p.* These are numbered 1-4, where 1-2 are thought to form the full site that regulates *sven6562* expression and 3-4 the full site that regulates *rsrR* expression. Each reaction was carried out at a ratio 80:1 [2Fe-2S]:DNA with 4 nM DNA used in each case. It can be seen that each natural full and half site completely shifts (in the case of 1-4 both shifts have occurred by this concentration as demonstrated in Figure 4.18) but we only see partial shifts at best for each artificial half site. This suggests that RsrR can indeed bind half sites, though, with much weaker affinity but indicates that the detected, natural, half sites for *sven0247p* and *sven0519p* are in fact full binding sites. It is possible that the artificial half sites have poor binding affinity to RsrR and that *sven0247p* and *sven0519p* have a greater affinity however we suspect that in each case, successful binding is occurring at full site sequences which it has been unsuccessful to identify with MEME.
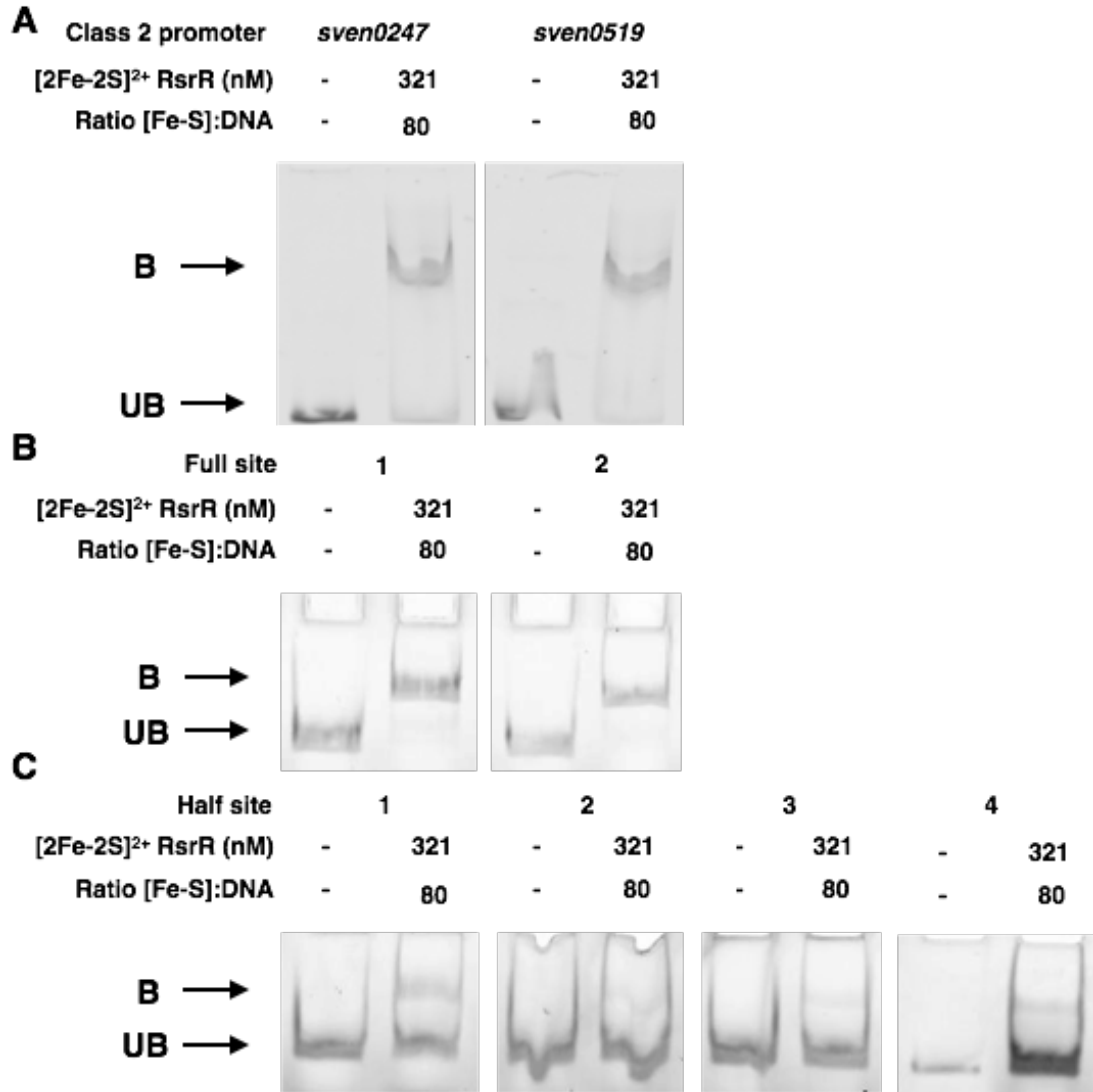
Figure 4-18 Oxidised RsrR binding to full site (class 1) and half site (class 2) RsrR targets. EMSAs showing DNA probes unbound (U) and bound (B) by $[2Fe-2S]^{2+}$. Ratios of [2Fe-2S] RsrR to DNA are indicated. A DNA concentration of 4 nM was used for each probe. EMSA's using class 2 promoters *sven0247* and *sven0519* (A), class 1 probes from the RsrR *rsrR* binding region (B) and the four possible half sites from the *rsrR* class 1 sites (C) were used.

4.2.3.3 Footprinting analysis of target sites

To analyse the RsrR binding sites in the *nmrA-rsrRp* intergenic region, to define the limits of the binding sites and check that the MEME predicted consensus was correct DNaseI footprinting was carried out (as described in section 2.2.4.8.2). DNaseI footprinting works on the principle that DNA bound by protein is protected from nuclease activity as described in section 3.2.1.2. Using a range of protein concentrations we are able to see a distinct zone of protection that corresponds almost exactly to our predicted binding sites for RsrR to the *nmrA-rsrRp* sequence (Figure 4.19).
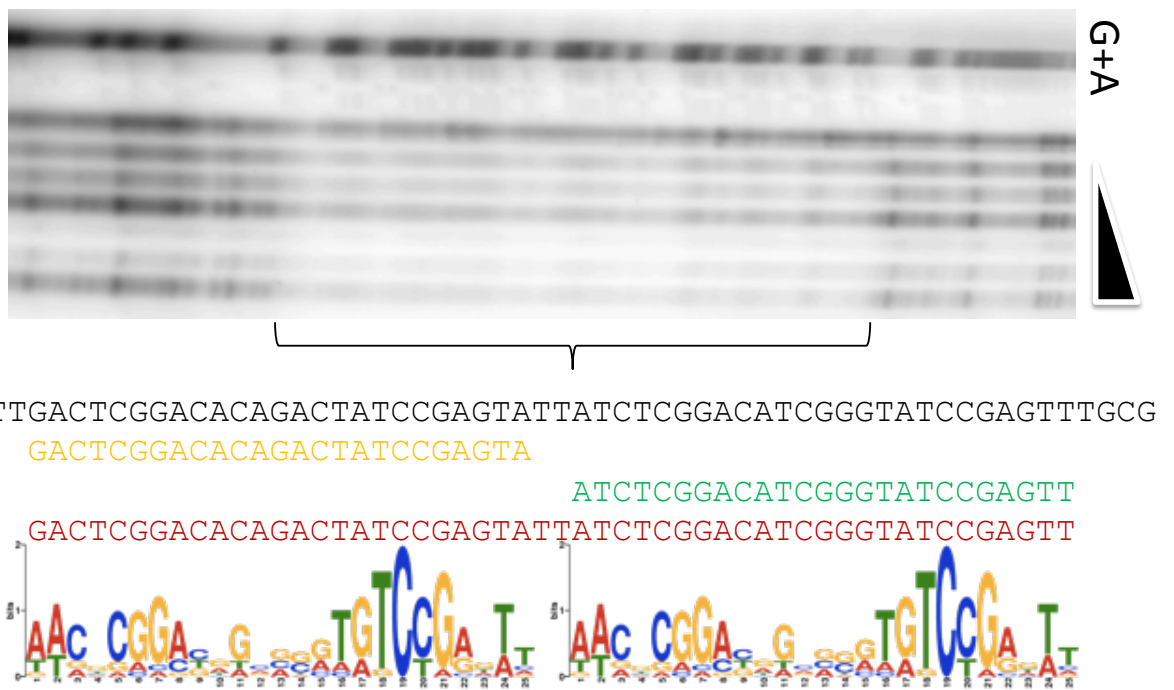
Figure 4-19 RsrR DNaseI footprinting of *nmrA-rsrR* binding sequences. Footprinting was carried out using primers JTM135 + JTM136 with JTM136 being labelled with P^{32}. Following G/A ladder and a space, we have sample reactions containing 0, 0,1, 0.25, 1, 2, 2, 2 µM protein with DNaseI digestion times of 10, 60, 90, 120, 150, 200 and 250 s respectively. The black text corresponds to the binding site highlighted by the zone of protection, the yellow site corresponds to the *nmrA* predicted binding sequence and green the *rsrR* prediction binding sequence. Red shows the merged sequence with no gaps and below is the MEME predicted consensus for the RsrR full binding site.

4.3 Discussion

From the initiation of this project various steps have been taken to further our understanding of the role of RsrR in *S. venezuelae.* Similar to the work described in the NsrR chapter, we see no developmental phenotype under our normal laboratory growth conditions when RsrR has been disrupted. However, unlike NsrR, RsrR appears to have an expansive regulon, based on ChIP-seq experimentation, with many genes linked to recycling of NAD(P)H to NAD(P) or with products predicted to bind NAD(P) suggesting a role in redox balance and genes involved in glutamate/glutamine metabolism. However, comparing gene expression of WT to and *rsrR::apr,* suggest that RsrR is not as strongly or specifically linked to these functions, with man of the strongest ChIP-seq targets looking to be unaffected by the loss of RsrR. RsrR has two binding sites between the loci *sven6562*-3, these genes corresponding to *nmrA* and *rsrR* respectively. Expression of *nmrA* is greatly increased in the *rsrR::apr* back ground.

4.3.1 RsrR disruption and *nmrA* expression.

Disrupting genes is not trivial. In many cases we are limited by the technology available, particularly in regards to the short falls each has. For *Streptomyces*, three major systems are now available, the Redirect PCR targeting route (Gust et al. 2006), the meganuclease system (Fernández-Martínez & Bibb 2014), and three independently produced systems relying on CRISPR/CAS9 (Tong et al. 2015; Cobb et al. 2014; Sampson et al. 2013). Each has its advantages and disadvantages however, for RsrR we utilised the redirect system, primarily because we had access to a cosmid library to facilitate its use and it was the only available system at the time.

As part of the redirect system, a complex disruption cassette is inserted into a target gene. The major features of this cassette include a resistance gene, an origin of transfer, FRT sites flanking the cassette for FLP recombinase recognition and a promoter driving the expression of apramycin resistance gene. Insertion of the resistance gene allows for easy identification of mutants, however, the promoters have been anecdotally linked to changes in expression of operonic/local genes. It is commonplace to "FLP out" the resistance cassette, using FLP recombinase to remove the cassette, leaving an ~90 bp scar, to avoid downstream expression modifications. However, for currently unknown reasons, doing so in *S. venezuelae* is a difficult and rarely successful endeavour (Dr Mahmoud Al-Bassam, University of

California and Matthew Bush, John Innes centre, personal communication). As a result, the *rsrR* mutant contains promoters potentially driving expression from the disruption cassette. There was concern that multiple promoters in the cassette could potentially be driving expression of *nmrA (sven6562)* however literature reviews and discussions (Dr David Widdick, John Innes Centre, Dr Bertolt Gust, University of Tübingen personal communications) have concluded this is most likely not the case. Within the disruption cassette, the only promoter is directed in the same orientation as the disrupted gene (Gust et al. 2006).

We have two hypotheses in regards to the RsrR *rsrR/nmrA* binding sites. Either, as is the case with the *E. coli iscR* promoter, which contains three experimentally derived binding sites (Giel et al. 2013), the two RsrR binding sites both control RsrR as part of a fine tuning mechanism, or each site regulates the expression of one of the divergent targets (*rsrR* and *nmrA*). At this point we have not distinguished which is the more likely option. Our co-conservation and expression data suggest that *rsrR* and *nmrA* are linked and based on this we suspect that both genes are divergently expressed and controlled by RsrR. The increase in expression of *nrmA* however provides us with confidence that the genes are linked functionally.

A similar series of experiments will be carried out to study *nmrA* as has been associated our *rsrR* work as well as investigating any interactions they have as multimers or by regulon overlap.

4.3.2   RsrR targets: a global reach with subtle implications?

Between the two data sets, ChIP-seq and RNA-seq, we identify two fairly distinct regulons. ChIP-seq, ultimately defining the sites RsrR binds to, does not necessarily correlate to direct genetic regulation. RNA-seq on the other hand is a direct representation of the RNA pool, and as a result a representation of the cellular response to losing RsrR. With the advent of Ribo-seq (Ingolia 2014; Ingolia et al. 2009) and the increasing awareness of translational level control, even this is not as robust as was once thought. The two data sets do correlate strongly and as a result have quite different putative targets.

When considering the ChIP-seq results, we see many examples of glutamate and glutamine associated targets. glutamate and glutamine are precursors for the production of mycothiol, the Actinobacterial equivalent of glutathione that acts as a cellular reducing agent. Mycothiol also acts as a cellular reserve of cysteine and in the detoxification of redox species and antibiotics (Newton & Fahey 2008).

Glutamate is important, as a non-essential amino acid, as it links nitrogen and carbon metabolism in bacteria (Berg et al. 2007). Additionally glutamate acts as a proton sink through its decarboxylation to GABA, which, especially under acidic conditions, favorably removes protons from the intracellular milieu (Feehily & Karatzas 2013).

Additionally, we see many targets associated with $NAD^+/NADH$, $NADP^+/NADPH$, either linked to its utilisation or regeneration. The pools of these resources are important stores for cellular redox poise, regulation of its utilisation is important under redox stress conditions. Having regulators such as NmrA involved in this regulon, could be a link to e.g. RsrR sensing a stress through its [2Fe-2S] cluster and NmrA, through binding of NAD cofactors, sensing various levels of nutrient and resource depletion throughout the response and recovery stages.

With the transcriptional data available, its general lack of correlation with the ChIP-seq results and no defined phenotype, it is difficult to rationalise how the changes in expression observed (and indeed the genes regulated) specifically effect the cell and how they are directly or indirectly related to the function of RsrR. Additionally with only one replicate of each of the RNA-seq experiments it is difficult to confirm that the observed fluctuations are biologically significant or relevant. With this is mind our primary aim for progressing this work will be carrying out a single and double disruption of *rsrR* and *nmrA* and carrying out a robust investigation to identify a phenotype. Recent access granted to equipment at the University of Strathclyde, specifically referencing their multiplate, 96 well, Biolog, a device capable of screening over a thousand different growth conditions simultaneously, we have high hopes of defining the role of *rsrR.* Once this has been defined further transcriptional analysis might result more significant results. With this in mind, we tentatively suggest that RsrR is involved in redox stress but this has not been confirmed.

4.3.3 The RsrR regulon: is *nmrA* the only true target?

We have seen clear links between RsrR and *nmrA.* When we consider the other targets however are these linked to the activity of RsrR or more directly, NmrA? Analysis of the dRNA-seq data using CLC genomics workbench shows significant changes in the expression profile of some target genes but many do not directly correlate with the RsrR ChIP-seq data. Specifically looking at *sven1847* and *sven3827*, we see little change. These along with *rsrA-nmrA* constitute three of the

strongest bound and highly enriched ChIP-targets. Rationally we would expect binding strength to be linked with either target expression or the requirement for tight control of a potentially toxic/wasteful target if poorly regulated. At this stage we are becoming increasingly convinced that *rsrR/nmrA* are linked its is possible that NmrA and RsrR have overlapping regulons, subject to different criteria of regulatory switches, both of which must be met to free expression of certain genes while having independent targets. From this point we suggest that RsrR plays a primary role, sensing the presence of redox stress and derepresses expression of NmrA which may detect the cellular redox poise, similar to Rex (Pagels et al. 2010), by interacting with the cellular NADH/NADPH pool. This would allow the detection of a stress, for example redox, through RsrR and monitoring of an on going process of damage, loss of cellular reducing power, through NmrA. With this in mind we suggest that a disruption of *rsrR* likely is not enough to see a full change in expression of these "regulons" but upon addition of a more wide effecting stress we will see a change in the regulon. Part of our issue in defining the function of RsrR is going to be in determining the possible multiple levels of regulation upon the target genes and defining function for targets identified as having unknown function.

It is difficult to fully characterise the regulon however we tentatively suggest RsrR/NmrA are likely linked to redox stress through sensing cellular redox poise by regulating a broad set of pathways to manage the use of molecular reducing power and accommodate regeneration of these sources.

4.3.3.1 dRNA-seq

dRNA-seq, described by Sharma et al. (2010), is a well utilised method. Unfortunately, due to a lack of communication, a modified method of this was carried out, as described in the results section. Additionally, as described much (~70%) of the Plus TAP treated data, containing information on transcriptional start site enrichment, resulted in very short reads which were removed during adapter trimming.

4.3.4 RsrR biochemistry

4.3.4.1 RsrR redox cycles like IscR but it doesn't appear to affect DNA-binding, as with IscR.

Previous work carried out on the *E. coli* IscR protein has shown that under oxidising and reducing conditions, the proteins [2Fe-2S] can redox cycle between this state

and its ($[2Fe-2S]^{2+}$) state (Fleischhacker et al. 2012). Our collaborators were also able to show this for purified RsrR. Utilising this protein under oxidising and reducing conditions (oxidised by exposure to air and reduced by exposure to dithionate) we could not detect changes in DNA-binding using the *rsrR/nmrA* sites. The same result was shown for IcsR. In both cases we are unsure if this redox cycling is a functional feature of the proteins or simply a property of the protein/cluster.

4.4 Summary

This chapter explores the role of RsrR in the emerging model organism, *Streptomyces venezuelae*. We have begun initial characterisation of both the regulon of the protein and the protein itself. A core regulon consisting of between 116-129 genes was defined with a much larger data set produced. This putative regulon contains a large portion of putative regulators and potential links to NAD and or glutamine metabolism. A 11-3-11 bp imperfect inverted repeat binding sequence was defined by MEME and confirmed by in vitro EMSA assay. A binding site corresponding to ~half the 11-3-11 sequence was also defined. However EMSA assays with using target sequences and artificial half sites suggest that each of the sites is likely a miss called full site. The TSS for many of the genes in the putative regulon were defined in addition to broadly mapping all of the transcripts expressed at 16h. Preliminary expression profiles for wildtype and *rsrR* mutant strains were produced indicating little change within the regulon as linked to the RsrR binding site. The RsrR protein has been shown to form a dimer, each monomer containing a [2Fe-2S] cluster ligated likely by three conserved cysteines, at least one of which is essential for retention of the cluster during purification.

5   Lipoprotein Signal peptidase in *Streptomyces coelicolor*

Initial work carried out towards this degree was involved in the characterisation of the suppressor mutations that allowed *S. coelicolor* to survive in the absence of *lsp.* Due to unforeseen developments in this project, work was began on NsrR (Chapter 3) and subsequently RsrR (Chapter 4) that formed the larger contribution to this work. Although distinctly different in nature to the other chapters, the work contained within and the experience associated contributed greatly to Ph.D. experience and as such is included below with a distinct introduction. The work in this chapter was published shortly following the original submission of this document (Munnoch, Widdick, et al. 2016).

5.1   Introduction

Bacteria are subject to stresses from the varying environments in which they live. They have, as a result, developed a multitude of sensory and response mechanisms to manage their internal environment facilitating survival. All major branches of the bacterial kingdom contain a group of membrane bound proteins known as lipoproteins. This chapter will focus on my work studying a component of the lipoprotein biogenesis machinery, *lsp* (lipoprotein signal peptidase) that was thought to be essential in *S. coelicolor* following previous work in the Hutchings laboratory.

5.1.1   Lipoprotein biogenesis

Biogenesis of lipoproteins is a multi-step pathway involving a translocation step (carried out by the Sec or Tat machinery) to guide proteins to and across the cytoplasmic membrane and several genes *(lgt*, *lsp*, *lnt* and *lol*) involved in lipoprotein maturation. The pathway varies in complexity depending on the complexity of the cell envelope (e.g. Gram positive or negative) as illustrated in Figure 5.1.

The pathway, which will be discussed below, is unique to bacteria (Babu et al. 2006), even though lipid modified proteins exist in eukaryotes and archae (Berg et al. 2007), the routes of production differ. The first discovered lipoprotein, named Braun's lipoprotein (Lpp), was discovered in 1969 in *E. coli* (Braun & Rehn 1969) and ultimately the vast majority of work on lipoprotein biogenesis has been carried out in *E. coli*. Some other major strains investigated include: *Salmonella enterica* (Feldman et al. 1981; Gupta et al. 1993)*, Stapholococcus aureus* (Kurokawa et al.

2012; Kurokawa et al. 2009), *Lactococcus lactis* (Banerjee & Sankaran 2013), *B. subtilis* (Hayashi et al. 1985), *M. xanthus* (Xiao & Wall 2014), *M. smegmatis* (Tschumi et al. 2012), *M. tuberculosis* (Sassetti et al. 2003; Tschumi et al. 2012; Sander et al. 2004; Rampini et al. 2008; Tschumi et al. 2009), *M. bovis* (Brülle et al. 2013), *C. glutamicum* (Mohiman et al. 2012), *S. coelicolor* (Thompson et al. 2010; Widdick et al. 2011; Córdova-Dávalos et al. 2014), *S. scabies* (Widdick et al. 2011) and *S. lividans* (Gullón et al. 2013). Protein translocation followed by each enzymatic step in lipoprotein biogenesis will be discussed below.
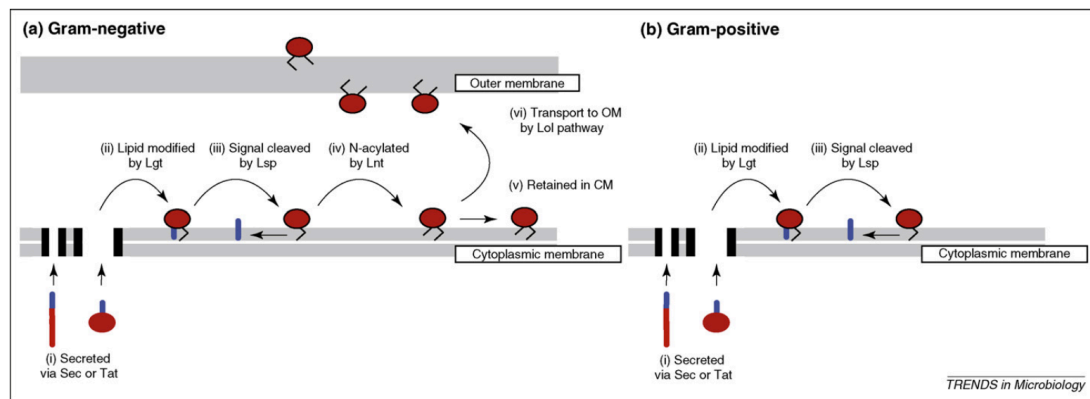
Figure 5-1 A graphical representation of the Gram-negative and Gram-positive lipoprotein biogenesis pathways Taken from Hutchings et al. 2009. (a) Represents the Gram-negative pathway including translocation across the membrane by Sec or Tat, followed by insertion of the signal peptide into the membrane, addition of the diacylglyceride moiety by Lgt, cleavage of the signal peptide by Lsp additional N-acetylation of the protein by Lnt and either retention on the inner membrane or passage to the outer membrane by the Lol system. (b) The Gram-positive system can be considered a simpler version of the Gram-negative pathway, having the same components, Sec and Tat translocation, Lgt and Lsp and, in actinomycetes, Lnt.

5.1.1.1 Overview of protein translocation

The phospholipid membrane bilayer, common to all cellular life, is a universal barrier utilised, in-conjunction with varied types of embedded and anchored proteins, to maintain the ionic integrity of the cytoplasm as well being a functional platform for a host of molecular processes. This membrane prevents the uncontrolled passage of macromolecules and proteins from their cytoplasmic site of synthesis to the extracytoplasmic environment. However many proteins and macromolecules have key metabolic or structural roles, outside of the cell, free or

bound to the membrane or cell wall. Various protein translocation systems have been identified, general and dedicated, which control and regulate the secretion of proteins. Only two of these systems can be considered as general translocation systems with a third now recognised as a Gram-positive specific general secretion pathway (all of which have been identified in *Streptomyces* spp.), the general **Sec**retary (**S**ec) pathway and the **T**win **A**rginine **T**ransport (TAT) pathway (Zückert 2014) and the Esx or Type VII secretion systems (Houben et al. 2014), the latter of which will not be discussed in detail. The dedicated Gram-negative systems (Costa et al. 2015), known as the Type I-VI secretion systems (T1-6SS) (Masi & Wandersman 2010; Cianciotto 2005; Cornelis 2006; Henderson et al. 2004; Filloux et al. 2008) have specific functions within prokaryotes and types III, IV, VI and VII are also key virulence factors. The type VII secretion system was initially found in *M. tuberculosis* where disruption of this system caused virulence attenuation. This is potentially important because a naturally attenuated strain, *M. bovis* BCG, resulted in the BCG vaccine (Hsu et al. 2003; Pym et al. 2003; Gey Van Pittius et al. 2001; Stanley et al. 2003). *M. tuberculosis* has 5 of these secretion systems (Houben et al. 2014). However, the type VII secretion system has also been found in non-pathogenic organisms and so is hypothesised to have additional functions in addition to virulence (Chater et al. 2010).

5.1.1.2 Translocation machinery: Sec and Tat

Of the secretion pathways mentioned, Sec (Beckwith 2013; Chatzi et al. 2013; Kudva et al. 2013) and Tat (Palmer & Berks 2012; Kudva et al. 2013; Goosens et al. 2014; Costa et al. 2015; Cline 2015) can be considered the housekeeping systems and will each be discussed below.

5.1.1.3 The Sec Pathway

The Sec pathway, also known as the main terminal branch of protein secretion, is found in bacteria, within the endoplasmic reticulum (ER) of eukaryotes (Muñiz et al. 2001), the thylakoid membranes of plants chloroplasts (Natale et al. 2008) and in Archaea (Bolhuis 2004). The Sec machinery specialises in the translocation of unfolded, nascent proteins into and across the cytoplasmic membrane.

The primary Sec machinery effectively consists of three key components: the protein conducting channel (PCC) through the membrane (formed by SecYEG), an ATPase motor and slippage system driving secretion in one direction (SecA and SecDFYajC) and a protein targeting system, utilising signal sequences (which will

be discussed along with the Tat signal sequences) along with chaperone molecules. The PCC consists of a heterotrimeric complex of SecY, SecE and SecG, which, depending on the process of translational passage to the Sec machinery, associates either with FtsY or with SecA, the ATPase containing motor protein that drives protein translocation. SecY, the transmembrane pore component of the SecYEG PCC complex, has ten transmembrane domains (TMD 1-5 and TMD 1-6 mirroring each other in conformation to form either side of the pore) with both the N and C termini residing on the cytosolic side (Osborne et al. 2005; Natale et al. 2008). Reviews of SecYEG (Osborne et al. 2005; Natale et al. 2008; Kudva et al. 2013) in various backgrounds suggests an array of mechanisms, utilising biochemical and crystal data, to explain how the pore opens and closes to allow protein translocation and how the system drives unidirectional transport (to prevent protein "slipping" back into the cytosol). The mechanisms rely, at least in *E. coli*, on an essential plug mechanism, which can be provide by either SecY or E as shown by truncation studies, however, a complete loss of either results in cell death. SecE acts as a locking clamp, holding the two sides of the SecY protein together, by an amphipathic cytoplasmic loop binding the second and third TMD (Natale et al. 2008). SecG has been implicated in assisting SecA functionally. SecG has been reported to cycle within the membrane inverting its two TMD from one side of the membrane to the other during translocation. SecG is not essential for function when bound in one conformation (Driessen & Nouwen 2008), although translocation efficiency is reduced in its absence (Beckwith 2013).

Additionally the PCC complex associates with another heterotrimeric membrane complex comprised of SecD, SecF and YajC (Komar et al. 2015). The SecDFYajC complex has been implicated in facilitating preprotein translocation (du Plessis et al. 2011) as well as complex formation of SecYEG, catalytic regulation of SecA (Driessen & Nouwen 2008) and virulence in *S. aureus* (Quiblier et al. 2013) while two copies of SecDF, separate and fused have been identified in *S. coelicolor* (Zhou et al. 2014). SecDF is thought to improve efficiency by preventing the backwards slippage of the translocating proteins through the SecY pore. SecB is a 17 kDa tetrameric *E. coli* chaperone protein that binds to basic regions on Sec-targeted proteins as they are synthesised by the ribosome and prevents them folding before they reach the SecYEG machinery (Weiss et al. 1988; Randall et al. 1998). SecB has been shown to interact with SecA, the ATPase component, however if a protein does not require SecB, it is recognised by SecA alone. SecA, using ATP as an energy source, drives translocation of SecYEG substrates across the membrane.

Finally the protein YidC (discussed in detail by du Plessis et al. (2011)) has been implicated in having a chaperone function with the Sec translocase from insertion of SecE to interaction with SecD and SecF during inner membrane protein insertion of for example, respiratory complex assembly of cytochrome bo_3 quinol oxidase and ATP synthase (Kol et al. 2008). From all the specific proteins involved in the Sec pathway; SecY, SecE and SecA are essential (Natale et al. 2008). However, disruption of SecDF results in pleiotropic secretion defects (Pogliano & Beckwith 1994) that ultimately are thought to be due to slippage within the pore. Interestingly, species which are naturally lacking SecDF have improved secretion function when complemented with heterologous SecDF (du Plessis et al. 2011) suggesting that acquisition of SecDF was evolutionarily favourable and potentially an early evolutionary divergence event.

5.1.1.4 Routes of Sec mediated protein translocation

Sec substrates are specifically targeted to the Sec machinery by an N-terminal signal sequence and depending on the signal sequence are targeted to the Sec translocase in two ways: Co-translationally, which is a more generally utilised process responsible for transmembrane protein insertion, and post-translationally, which utilises additional, more specific chaperone proteins to keep the substrate unfolded (summarised in the Figure 5.2). Both routes are utilised in *Streptomyces* spp. (Chater et al. 2010).
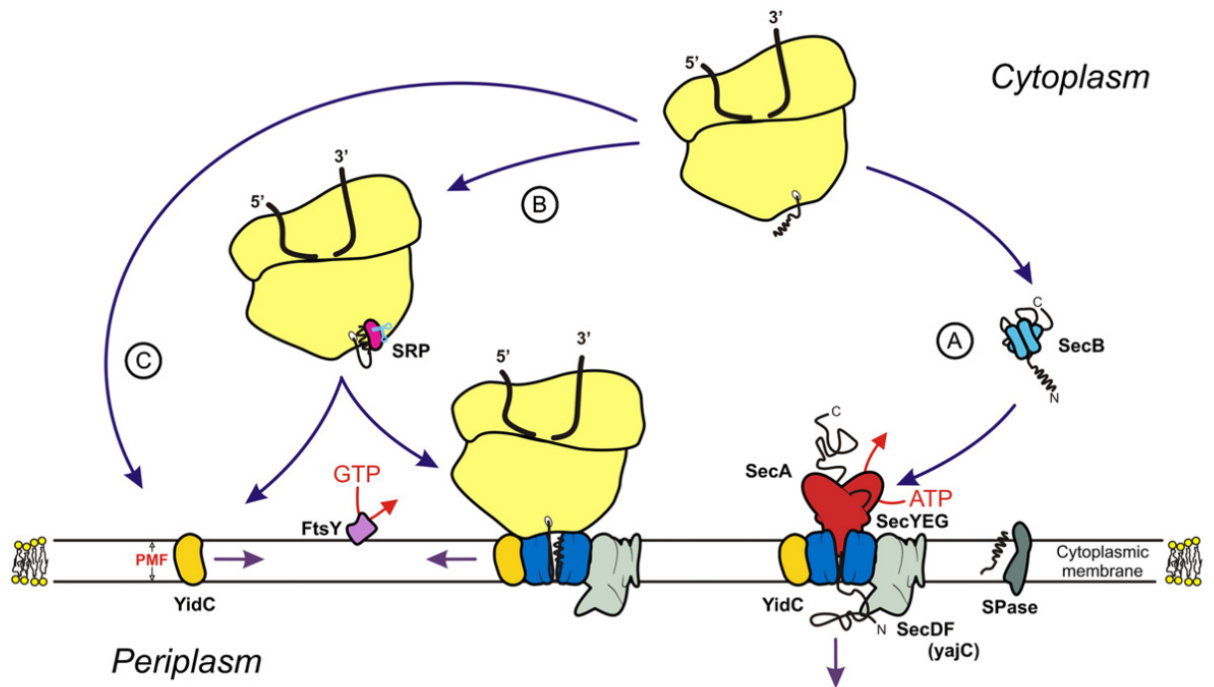
Figure 5-2 Taken from du Plessis et al., 2011, this is an overview of the Sec pathway. (A) Represents the Post-translational pathway, (B) represents the Co-translational pathway where YidC is variably associated and (C) representing that some proteins are inserted into the membrane by YidC alone. Each of the individual components are labelled: Ribosome (pale yellow), SecB (cyan), SRP (pink), SecA (red), SecYEG (dark blue), SecDF (yajC) (light grey), FtsY (purple) and YidC (dark yellow).

5.1.2 Co-translational protein translocation

The co-translational translocation process requires the ribosome to act as a "major channel partner" where the Signal Recognition Particle (SRP) binds the signal sequence of protein during synthesis by the ribosome. This SRP targets the ribosome/protein complex to the membrane where SRP binds to the SRP receptor (SR), previously named FtsY (Albers et al. 2006), and the ribosome binds to the PCC. The SR catalyses GTP hydrolysis facilitating transfer of the nascent polypeptide chain to the PCC and translation of the remainder of the protein drives translocation. There is evidence that in mammalian systems that the SRP slows down polypeptide synthesis by blocking tRNA interaction/incorporation with the ribosome through two proteins SRP9 and SRP14 interacting with the Alu domain of the 7S RNA to insure that correct translation efficiently drives translocation (Wild et al. 2004).

5.1.2.1   Post-translational protein translocation

The post-translational translocation process requires an additional chaperone protein, which in *E. coli* is named SecB. The protein is completely translated by the ribosome; the signal sequence avoids binding to the SRP, which is hypothesised to be a result of a higher net negative charge. SecB then binds to basic regions of the mature protein rather than the signal sequence and targets the unfolded protein to SecA, which is bound to the PCC complex (Natale et al. 2008).   The protein is subsequently transferred to SecA that drives translocation across the membrane through hydrolysis of ATP to ADP. SecB is released from the complex, as it is not required for translocation. SecA interacts with most of the other Sec proteins through various domains and is thought to be active in its dimeric form. A theory among many of how SecA drives translocation is through a racheting mechanism where for each ATP hydrolysed to ADP and the energy generated is used to drive the polypeptide chain through the pore 20-30 amino acids at a time (Osborne et al. 2005).

5.1.2.2   Tat

The Tat system is used to export fully folded and cofactor containing proteins (Robinson & Bolhuis 2004) which commonly only consists of a small percentage of the secreted proteome (the vast majority do not require cytoplasmic folding and are exported by Sec). The Tat pathway, unlike Sec, is not ubiquitous, found in ~50% of sequenced bacteria, in plant chloroplasts and in most Archaea and is considered an ancient system because of its prevalence in prokaryotes and prokaryote derived organelles (Cline 2015). The Tat pathway has been shown to be functional with a relatively minimal set of components, consisting only of the TatA and TatC (6) and the proton motive force. The Tat system in *Streptomyces*, unlike in other bacteria, acts as a major secretion system, having a large number of associated cargo proteins (Widdick et al. 2006; Joshi et al. 2010; Schaerlaekens et al. 2004). Recent work in *Streptomyces* has suggested that Tat, alongside Sec, is involved in the localisation of the Cytochrome *bc1* complex (Hopkins et al. 2014).

5.1.2.3   The Tat translocation machinery

The Tat system consists of two major proteins, TatA and TatC (Yen et al. 2002) with the larger TatA family consisting of TatA, TatB and TatE (Fröbel et al. 2012). Figure 5.3a illustrates the Tat proteins. There are two major Tat translocase complex types either consisting of a two component translocase, TatA and TatC are

described as a minimal system, while there is also a more complex three component TatABC system (Natale et al. 2008). In Gram-negative bacteria it is common for a *tatD and tatE* to be found encoding a cytosolic protein and a TatA paralogue respectively. *tatE* is regarded as a duplication of the *tatA* gene as *tatABCD* form an operon and *tatE* is monocistronic in *E. coli* (Fröbel et al. 2012). In *Streptomyces*, TatABC are present and required for Tat transport to work (Goosens et al. 2014). It has been shown in *E. coli*, through gene disruption, that *tatD* and two homologs *ycfH* and *yijV* do not effect function or localisation Tat or its substrates and it has been suggested that *tatD* is only part of the operon due to shared regulatory requirements (Wexler et al. 2000). However, *in vivo* evidence shows that *tatD* is involved in degradation of misfolded Tat substrates (Matos et al. 2009). Within most Gram-positive bacteria however, only orthologues of TatA and TatC are expressed. *B. subtilis*, which interestingly has a low predicted number of Tat substrates (Widdick et al. 2006), expresses two paralogues of TatC, $TatC_d$ and $TatC_y$, and 3 paralogues of TatA, $TatA_d$ $TatA_y$ $TatA_c$ forming two systems $TatA_dC_d$ and $TatA_yC_y$ in which they have been, correctly although inaccurately described as substrate specific pathways for the proteins PhoD and YwbN, respectively (Jongbloed et al. 2004), due to conclusions drawn from the low number of *B. subtilis* Tat substrates (Barnett et al. 2009).

5.1.2.4 The Tat complex translocation process

The Tat complex effectively consists of two components: a docking complex (TatBC) and a pore complex (multimeric TatA) illustrated in Figure 5.3b (Goosens et al. 2014). As described previously there are TatABCDE present in *E. coli* but only TatABC are essential for function and TatB is only present/essential in Gram-negative bacteria and some Gram-positive bacteria (e.g. *Streptomyces* spp.). It is suggested that Tat substrates have no major cytosolic messenger chaperones to target the proteins, by their signal sequence, to the Tat machinery and instead it is suggested that signal sequence is recognised and bound instead by the Tat complex itself in the absence of chaperones (Fröbel et al. 2012). Due to a broad array of experimental cross linking proximity tests using *ΔtatC* mutants and Tat signal sequences with various mutations in the conserved motif fused to reporter enzymes it is widely regarded that TatC binds the signal sequence, inserting the protein in the membrane temporarily and acts as initial binding platform for Tat substrate translocation (Fröbel et al. 2012; Goosens et al. 2014). Following this, the protein interacts with TatB where TatB oligomerises to form, very similarly to TatA,

an encapsulating structure that binds to the Tat substrate in the absence of the PMF (proton motive force). This was shown by experimentation with the PMF inhibitor carbonyl cyanidem-chlorophenyl-hydrazone (CCCP) (Maurer et al. 2010; Fröbel et al. 2012).
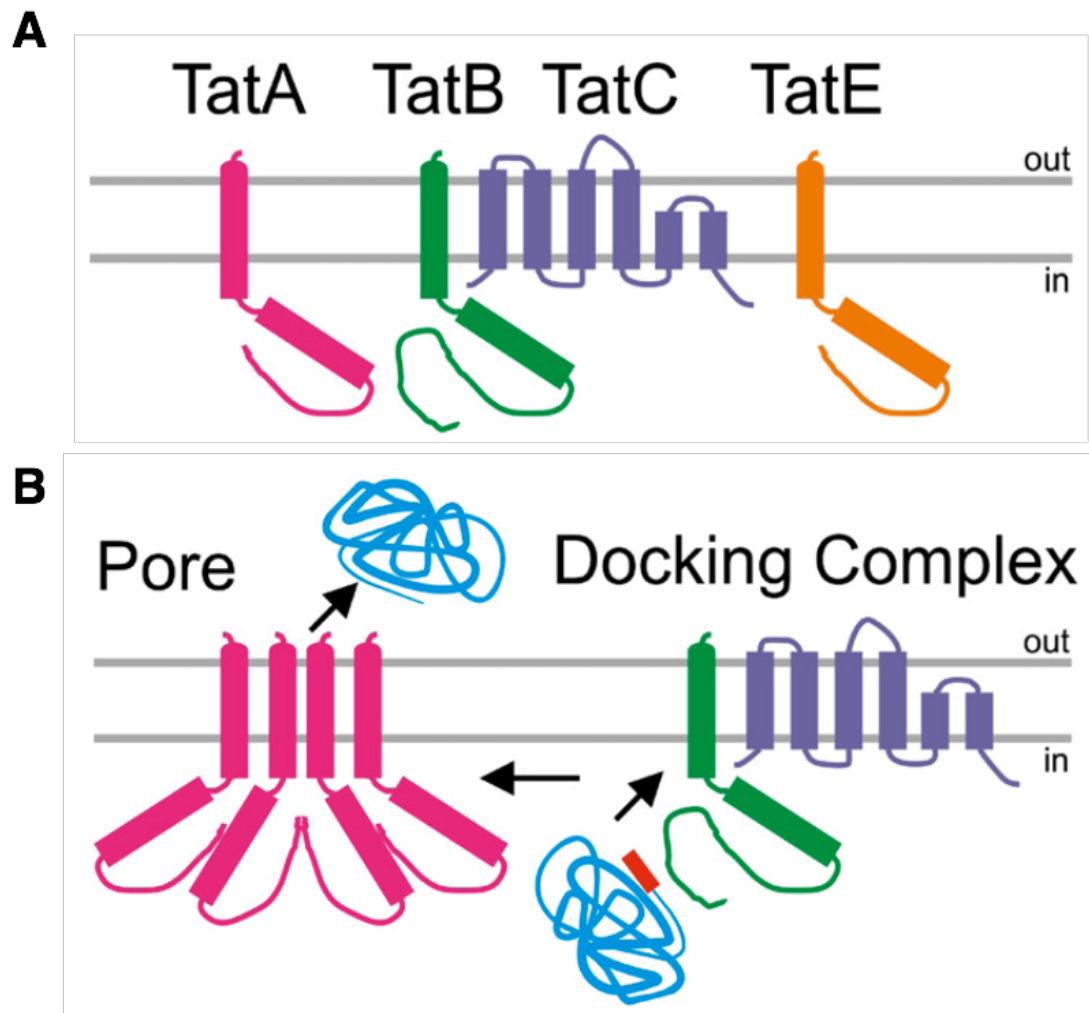
Figure 5-3 A schematic representation of the Tat proteins A-E showing their roles in the Gram-positive tat translocation system (Goosens et al. 2014). A) The universally confirmed components including TatA and TatA like monomers TatB and TatE. (B) Tat functional complexes based on a consensus model, showing the docking complex formed by TatC (purple) and TatB (green) showing the interaction of a Tat substrate (blue) signal sequence (red) before transfer to the Pore complex (formed by TatA (pink) multimers) and transportation across the membrane and out of the cell.

TatB has also been suggested to bind multiple substrates (Maurer et al. 2010) in this way which may act as a "caching" mechanism if there is a PMF limiting

condition or there is a build up of Tat substrates due to the rate of expression exceeding the rate of translocation. Following binding of substrates to the docking complex, TatA is recruited to form the pore complex, forming multimers of reportedly varying sizes, through the membrane facilitating passage of the Tat substrate (Goosens et al. 2014).

5.1.2.5 Signal sequences

The Sec and Tat signal sequences (Figure 5.4) have several aspects in common however they also have a key difference that facilitates their specificity. Interestingly however a study in *S. lividans* has shown that proteins exported by the Sec pathway can be ReDirected to the Tat pathway but not *vice versa* by switching the signal peptides (Gullón et al. 2015).

The signal sequence/peptide is essential for correct targeting of the proteins to the Sec machinery. Typically ranging between 15-30 amino acids, signal peptides are highly conserved and have a tripartite structure consisting of a positively charged N-terminus (n-region), a hydrophobic core (h-region) and a polar C-terminus (c-region) (Natale et al. 2008). Although there is high structural conservation the specific amino acid sequence is highly variable within the restriction of maintaining the overall charge characteristics. Sec signal sequences can be predicted from protein databases using the online tool SignalP (Emanuelsson et al. 2007).
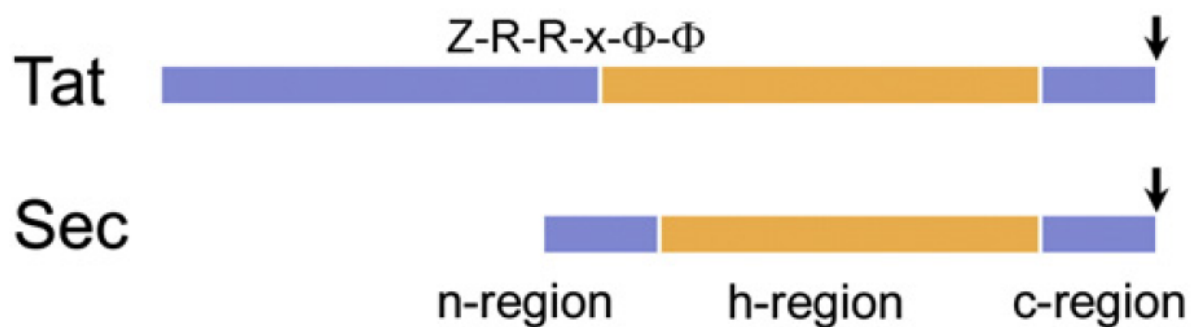
Figure 5-4 A schematic representation of the Sec and Tat signal sequences showing the three major regions, general size and also in the case for TAT, the location of the twin arginine (RR) motif. Copied from Natale et al., 2008.

The N-terminal signal sequence of Tat substrates has a very similar tripartite structure to that of Sec signal sequences. Although significantly larger the Tat and the Sec signal sequences share the n-region, h-region and c-regions previously mentioned. However unlike Sec, Tat substrates have a near invariant double

164

arginine (Arg) located at the interface between the n and h-region although examples have been reported where one Arg is replaced by a Lys, Asn or Gln (Fröbel et al. 2012). The original Tat motif was described as (S/T)-R-R-x-F-L-K, but the more recently revised motif consists of Z-R-R-x-φ-φ where Z represents any polar residue, x (in both) any residue and φ any hydrophobic residue (Berks 1996; Natale et al. 2008). Also more species specific motifs have been published such as that for *Streptomyces lividans* which is an amalgamation of the two described motifs (De Keersmaeker et al. 2007). Translocation efficiency is reduced in many cases if residues are changed within this conserved region (Berks et al. 2003). It has also been shown that an increase in n-region or h-region hydrophobicity can result in a misdirection of Tat substrates to the Sec pathway. Similarly within the c-region the overall basic nature of the residues prevents Tat substrate targeting to the Sec machinery. These three components of the signal sequence suggests a general Sec avoidance mechanism for Tat specific substrates (Fröbel et al. 2012).

5.1.3   Lipoprotein biogenesis pathway

5.1.3.1   Lgt, Lipoprotein diacylglyceryl transferase and the lipobox sequence

The lipoprotein biogenesis pathway is typically considered a linear system relying on the ordered processing of each protein by Lgt, Lsp, Lnt then Lol when applicable. Each lipoprotein contains a Sec or a Tat signal sequence to facilitate transport to the membrane. These signal sequences differ from standard Sec and Tat signals. Each Lipoprotein signal sequences contains, in addition to the standard components of the signal peptide, a conserved sequence known as the lipobox, consisting of: L_{-3}-[A/S/T]$_{-2}$-[G/A]$_{-1}$-C_{+1} (Hutchings et al. 2009). The key feature of the lipobox sequence is the invariant Cysteine residue (C_{+1}) containing the sulfhydryl group to which Lgt, the first unique step in the lipoprotein pathway (Tokunaga et al. 1982), will covalently attach a diacylglyceride moiety via a thioether linkage. The diacylglyceride comes from the lipid phosphotidylglycerol which, when catabolised for this reaction, forms the diacylglyceride and *sn*-glycerol-1-phosphate by-product (Lai et al. 1980; Sankaran & Wu 1994). Phosphotyidylglycerol is the only lipid donor utilised by Lgt (Buddelmeijer 2015). Experiments have also shown that the +2/+3/+4 residues are important for correct localisation within the membrane by the Lol system, referred to as the Lol sorting signals, however, this has not been shown to be universal (Zückert 2014).

Lgt was originally identified in *S. enterica* by screening for the accumulation of unmodified Lpp in the cytoplasm of temperature sensitive mutants (Williams et al. 1989; Bell-Pedersen et al. 1991; Feldman et al. 1981). Prior to this, the only work carried out on the as yet unidentified protein was through the use of crude membrane extracts (Tokunaga et al. 1984). Several regions and residues of *E. coli lgt* have been reported to be highly conserved including: H103-G108, Y26, Y235 of which H103 and Y235 have been shown to be essential (Sankaran et al. 1997) and G142-G154, forming a 12 amino acid region referred to as the Lgt signature motif, of which 2 (N146 and G154) have been shown to be essential (Pailler et al. 2012). Interestingly it appears that the essential H103 residue in many actinomycetes is actually a tryptophan residue while Y235, N146 and G154 however are each conserved (Thompson 2010). Structurally, Lgt is a large membrane bound protein with 7 transmembrane spanning domains (Pailler et al. 2012). Each of the three proteins, Lgt, Lsp and Lnt membrane topology is presented in Figure 5.5.

A common feature of *lgt* mutants in Firmicutes is the shedding of unprocessed lipoproteins into the supernatant (Leskelä et al. 1999) or those processed by unknown peptidase (Stoll et al. 2005; Denham et al. 2009) or Lsp (Baumgärtner et al. 2007; Henneke et al. 2008). Interestingly, *S. coelicolor lsp* mutants result in a similar phenotype, resulting in shedding of lipoproteins into the supernatant (Thompson et al. 2010). *S. coelicolor* contains two *lgt* genes, each can be disrupted independently without a loss of Lgt function but not simultaneously suggesting a functional redundancy and also that Lgt is essential (Thompson et al. 2010). *S scabies* contains one *lgt* gene, disruption mutants result in a release of lipoproteins into the supernatant which can be complemented and prevented with either of the *S. coelicolor* Lgt enzymes (Widdick et al. 2011). However, based on mass spec data of purified lipoproteins from complemented strains, it appears that there are differences in protein efficiency as some were not lipidated (Widdick et al. 2011). *M. tuberculosis* and *M. smegmatis* Lgt enzymes have recently been identified (Tschumi et al. 2012). One of the *M. smegmatis* genes (MSMEG_3222) has been shown to functionally transfer the diacylglycerol moiety to lipoproteins however the other has still to be investigated and, like *S. coelicolor* may play a redundant role. *M. tuberculosis lgt* could not be disrupted suggesting it is essential (Sassetti et al. 2003; Tschumi et al. 2012). Comparisons of *E. coli*, *B subtilis* and Actinomycete Lgt enzymes (Thompson 2010) indicated that the Actinomycete Lgt proteins have C-terminal extensions with low sequence conservation however relatively little research has been carried out to investigate their function.
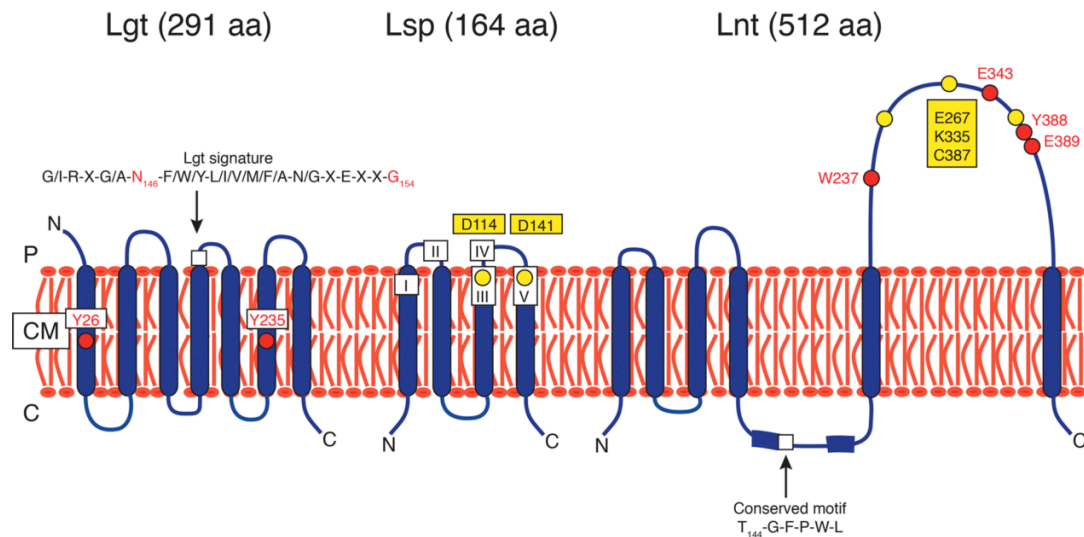
Figure 5-5 Membrane topology of the three lipoprotein biogenesis proteins, Lgt, Lsp and Lnt from *E. coli* showing key essential residues (red) and catalytic residues (yellow) important for function (Buddelmeijer 2015). C = cytoplasm, CM = cytoplasmic membrane, P = periplasm.

5.1.3.2 Lsp – Lipoprotein signal peptidase

Lsp, the lipoprotein signal peptidase is responsible for the cleavage of prolipoproteins, removing their signal peptide, leaving behind the C_{+1} residue at the amino terminal of the protein. In doing so the signal peptide no longer acts as the membrane anchor and the diacylglycerol moiety becomes solely responsible for this role. The Lsp protein consists of 4 transmembrane domains and contains 5 highly conserved regions (Muñoa et al. 1991; Tjalsma, Kontinen, et al. 1999) which, in *E. coli* (and *S. coelicolor*) correspond to amino acid residues: Region I, 23-27 (62-66), Region II, 53-58 (82-87), region III, 104-115 (128-139), region IV 121-123 (156-158) and region V 137-141 (174-178). For the most part these regions of conservation have not been investigated however in *E. coli* D114 and D141 compose the active site of Lsp (Tjalsma, Kontinen, et al. 1999; Tjalsma, Zanen, et al. 1999). At least in some actinomycetes, Lsp enzymes have varying lengths of N-terminal extensions (Thompson 2010) which, unlike the Lgt C-terminal extensions, are generally more conserved. Truncation of the N-terminal of the ScoLsp enzyme removing 20 and 40 amino acids resulted in a loss of function as did mutation of the conserved aspartate residues however truncations of 10 and 30 amino acids did not effect Lsp function (Thompson et al. 2010). The implication here is that the truncation either makes the enzyme more sensitive to degradation or the N-terminus of Actinomycete Lsp is necessary for function. Lsp does not act on unmodified lipoproteins (Tjalsma,

Kontinen, et al. 1999; Tjalsma, Zanen, et al. 1999) suggesting that it recognises the lipid moiety; a notable exception however is *L. monocytogenes* Lsp (Baumgärtner et al. 2007). Lsp is thought to be essential in *E. coli* (Tokunaga et al. 1982; Yamagata et al. 1983) and *S. coelicolor* (Thompson et al. 2010).

A key tool in the investigation of Lsp has been the antibiotic, globomycin, first identified in *Streptomyces globeosporus* as causing spheroplast-forming activity on *E. coli* (Inukai M., Nakajima M., Osawa M. 1978). Globomycin is a non-competitive inhibitor of Lsp, preventing the cleavage of the signal peptide (Dev et al. 1985), it is lethal in *E. coli* (MIC 0.4 µg/ml) and has been shown to be less active against *S. aureus* and *B. subtilis* (MIC > 100 µg/ml) (Inukai M., Nakajima M., Osawa M. 1978). Although there are no Lsp homologues in eukaryotes and archaea, it has been shown to inhibit signal cleavage and is therefore too toxic to be used clinically (Tjalsma, Kontinen, et al. 1999; Tjalsma, Zanen, et al. 1999; Giménez et al. 2007). Lsp over production has been shown to raise the MIC of globomycin *in vivo* (Geukens et al. 2006; Paitan et al. 1999; Prágai et al. 1997; de Greeff et al. 2003). In addition to globomycin, antibiotic TA produced by *Mycococcus xanthus* DK1622 has recently been shown to be an inhibitor of type II signal peptidases, of which, Lsp is a member (Xiao et al. 2012).

Although information is known about their membrane topology based on predictions, there are no structural studies of any of the biosynthetic genes of the lipoprotein pathway, including Lsp. However several other peptidases have been investigated including signal peptidase I/leader peptidase and type IV pilin signal peptidase/prepilin peptidase that have allowed for some inferred properties of how Lsp would work. Much of the work for leader peptidases in to the Ser/Lys active site (Wang et al. 2008) and catalytic mechanism (Paetzel 2014) have been reported, often relying on truncated proteins lacking hydrophobic domains. NMR work has implicated these proteins in carrying out cleavage on the membrane surface (De Bona et al. 2012) or deep within the membrane bilayer (Paetzel 2014). Other peptidases such as prepilin peptidase (Erez et al. 2009) an aspartyl peptidase like Lsp and other aspartyl proteases like FlaK (of which the crystal structure has been solved (Hu et al. 2011) provide putative evidence into how Lsp would work however, lacking direct Lsp data is a hindrance within the field.

5.1.3.3 Lnt – Lipoprotein N-acyl transferase

Lnt, first discovered in *S. enterica* by screening for temperature sensitive mutants, as with Lgt (Gupta et al. 1993), carrying out the final enzymatic modification to mature lipoproteins, functions by N-acylating the amino group of the C_{+1} cysteine of prolipoproteins forming an *N*-acyl-S-diacylglyceryl-cysteine linked protein. Similarly to Lgt and Lsp, Lnt is a membrane bound protein (Gupta et al. 1991), it contains 5 transmembrane domains, a EKC catalytic triad domain (Vidal-Ingigliardi et al. 2007) consisting of E267, K335 and C387 (in *S. coelicolor* Lnt1, E268, K355 and C405 and Lnt 2 E261, K322 and C374 respectively) (Thompson 2010). An additionally 4 conserved residues of interest are found within the periplasm (W237, E343 Y388 and E389) (Vidal-Ingigliardi et al. 2007). (In the case of *M. bovis*, the cysteine residue has been changed to a serine (Brülle et al. 2013). To date the only reported Lnt or Lnt orthologues that have been reported as essential exist in *E. coli* (Gupta et al. 1991; Rogers et al. 1991; Gupta et al. 1993; Robichon et al. 2005) *and S. enterica* (Gupta et al. 1993) from the strains previously mentioned*. Deletion mutants in proteobacteria cannot be complemented by homologues in actinobacteria (Vidal-Ingigliardi et al. 2007). The Lnt enzymatic reaction involves two stages: First, the Lnt cysteine of the catalytic domain reacts with phosphotidylethanolamine, forming a lysophopholipd by-product and a thioesteracyl-enzyme intermediate (Jackowski & Rock 1986; Buddelmeijer & Young 2010; Hillmann et al. 2011), which is then secondly transferred to a target prolipoprotein forming a mature triacylated lipoprotein (Buddelmeijer & Young 2010). A kinetic study of each enzyme in the pathway has led to similar K_m values for each respective reaction (Bishop et al, 2000, Dev et al, 1985 and Hillman et al 2011).

Streptomyces species have been reported to always contain two *lnt* genes. Studies in *S. scabies* have shown efficient N-acylation is reliant on both Lnt1 and Lnt2 (Widdick et al. 2011). They showed that both enzymes are independently capable of N acylation however efficient and complete processing requires both genes, which following work in Firmicutes, particularly late exponential growth studies at low pH grown *S. aureus* (Kurokawa et al. 2012) it was suggested that multiple enzymes potentially allow for the compensation of lipid availability and fluctuation under different growth conditions (Buddelmeijer 2015).

M. smegmatis work using Lnt deletions, the first reported for actinomycetes, directly showed Lnt N-acyltransferase activity by the accumulation of diacylated

lipoproteins in the *lnt* mutant background (Tschumi et al. 2009; Brülle et al. 2010). Mycobacterial enzymes have been shown to have a different substrate specificity than, for example, *Streptomyces* and proteobacteria by failed complementation studies, instead utilising the substrates palmitate generally and tuberculostearic acid, specifically in *M. bovis* (Brülle et al. 2013). The major conserved residues in Lnt: W237 and Y388 are poorly conserved in actinomycetes (Thompson 2010; Vidal-Ingigliardi et al. 2007; Tschumi et al. 2009).

Often studies have investigated the conservation of the proteins by attempting to complement mutants by heterologous expression of proteins. Some of those that have been successful include: *S. aureus lgt* complementing an *E. coli* mutant (Qi et al. 1995), Firmicutes *lsp* complementing *E. coli* mutants (Prágai et al. 1997), M. xanthus *lsp* can complement an *E. coli* Lsp depleted strain, but only in the absence of Lpp (Xiao & Wall 2014). Often actinobacterial enzymes will not complement non-actinobacteria.

5.1.4   Lol – Lipoprotein outer membrane localization pathway

The Lol system, although only relevant to Gram-negative bacteria, is an important system for the transfer of lipoproteins from the inner to the outer membrane (Figure 5.6). The Lol system has been extensively reviewed (Zückert 2014; Tokuda et al. 2014) however it has been included briefly here because of its links to the biosynthetic genes, but excluded from more thorough investigation because of its absence in Gram-positive bacteria, including *Streptomyces*, on which this work will focus.

Consisting of the proteins LolABCDE, each of which plays an important function in the system. LolE, a cytoplasmic membrane protein, is the first to interact with mature lipoproteins, recognising conserved signal residues located at the C-terminal of the C_{+1} cysteine (Okuda & Tokuda 2009). LolE passages the protein to LolC, a second cytoplasmic membrane protein, which interacts with periplasmic chaperone LolA and following ATP hydrolysis by LolD, a cytoplasmic ATPase, fully passages the lipoprotein to LolA. LolA transfers the protein to LolB, the outer membrane receptor protein, which inserts the mature lipoprotein into the membrane. The driving force of the mechanism is that increasing strength of binding to subsequent protein steps e.g., in order of decreasing strength of binding LolBACE which facilitates forward motion through each step (Taniguchi et al. 2005). N-acylation of lipoprotein by Lnt has been shown to be essential for interaction with the Lol

system. (Fukuda et al. 2002) however the mechanism for the LolCDE interaction is unknown (Buddelmeijer 2015). Overexpression of LolCDE can negate the need for N-acylation (Narita & Tokuda 2011).

Structural (Takeda et al. 2003) and functional (Remans et al. 2010) studies for LolA and B have been carried out. A large hydrophobic region of LolA has been implicated in function as a reduction in hydrophobicity results in a loss of Lol functionality (Remans et al. 2010). LolA has been shown to interact specifically with LolC but not LolE. Interestingly both LolE and C (as well as B) show sequence similarity in their cytoplasmic domains (Okuda & Tokuda 2009). The Lol pathway is not totally conserved in proteobacteria or diderm actinomycetes, where a functionally similar but currently unknown system to Lol has been suggested to transfer lipoproteins to the mycobacterial outer membrane (Okuda & Tokuda 2011).
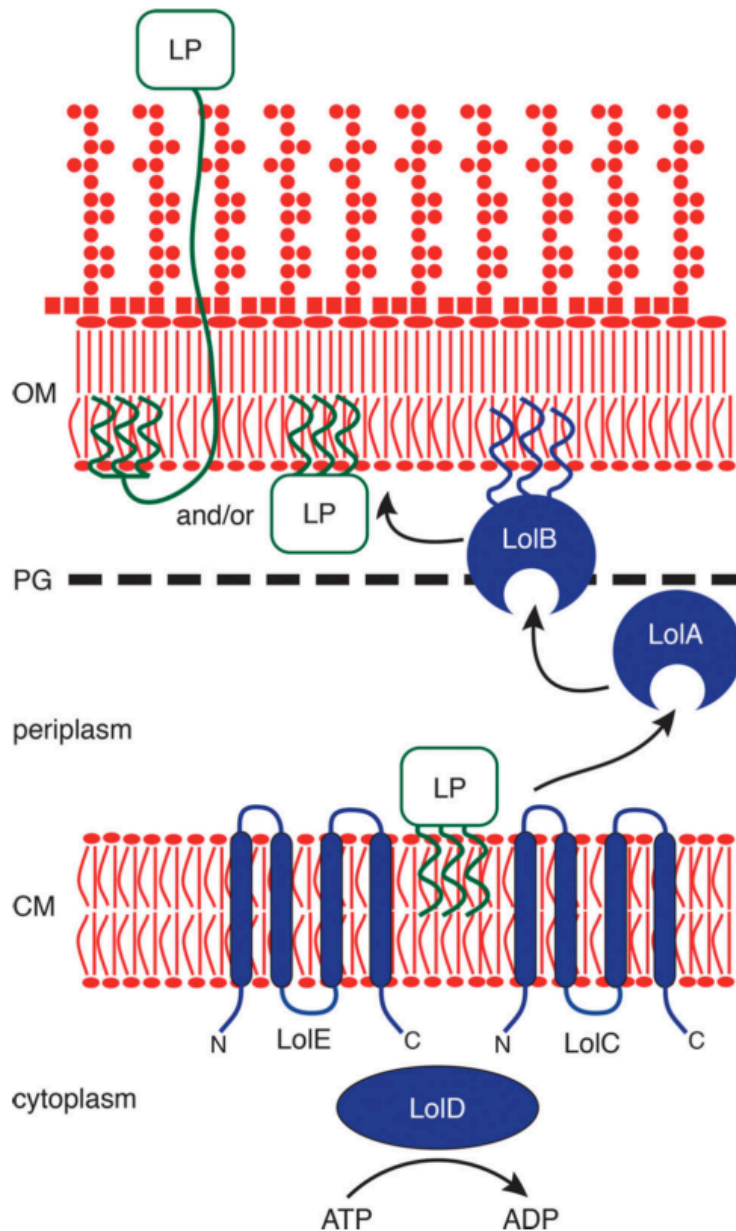
Figure 5-6 The Lol system in proteobacteria. Consisting of LolABCDE proteins, LolCDE form the cytoplasmic membrane ABC transporter component (CE) and the ATPase (D). LolA is a periplasmic chaperone which, following binding to target lipoproteins and interaction with LolC and ATP hydrolysis by LolD, transfers the lipoprotein to LolB, the receptor protein. LolB inserts the lipoprotein into the inner face of the outer membrane, while the mechanism for transfer to the outer face is yet unknown.

5.1.5 Lipoprotein functions

The bacterial lipoproteins have a range of functions within the cell. The most abundant and longest studied lipoprotein is Lpp and, as previously mentioned, was

the first lipoprotein discovered in *E. coli*. Lpp is involved in cell envelope homeostasis and 1/3rd exists covalently linked to murein of the peptidoglycan cell wall (Braun & Rehn 1969; Inouye et al. 1972). Lpp exists in an equilibrium (Hiemstra et al. 1986; Hiemstra et al. 1987) with 2/3rds of the protein being free within the membrane which, over time, is degraded and/or interchanged and linked to the peptidoglycan by YbiS, primarily, but also ErfK and YcfS in *E. coli* (Magnet et al. 2008). The *P. aeruginosa* homologue on the other hand has been reported to exist only in its free form (Mizuno & Kageyama 1979; Cornelis et al. 1989). Disruption strains of Lpp result in the leaking of periplasmic proteins, implicating a cell envelope destabilisation and an increased sensitivity to a range of stresses including: EDTA, cationic dyes and detergents (Hirota et al. 1977; Yem & Wu 1978). More recent studies have shown that Lpp is under the indirect regulation of Sigma E (Guo et al. 2014). Sigma E regulates a range of sRNAs which ultimately down regulate Lpp translation and increase the degradation of *lpp* mRNA. Lipoproteins have often been reported to play roles in the correct localisation of OM membrane proteins or are involved in complex cell envelope/extracellular processes. Two lipoproteins, LpoA and LpoB are essential for PBP1A and PBP1B roles in peptidoglycan synthesis (Paradis-bleau et al. 2011; Typas et al. 2010). Lipopolysaccahride (LPS) transport across the membrane by the Lpt pathway (Ruiz et al. 2009) requires LptE, an essential outer membrane lipoprotein involved in the stabilisation of the LPS OM channel formed by LptD (Bos et al. 2004; Chng et al. 2010). LptE stabilises the complex by forming a plug structure within the pore preventing excess cellular leakage (Freinkman et al. 2011).

Much of the work carried out to date on bacterial lipoproteins has been with a focus on their contribution to virulence, many specifically utilising Lgt and Lsp disruption strains to investigate the role of lipoproteins on virulence modulation. The host immune system is capable of detecting lipoproteins through interaction of diacylated proteins with Toll-like receptor (TLR) 2-1 or triacylated proteins view TLR2-6 (Lee et al. 2007; Kang et al. 2009)

A decreased virulence has been reported in *lgt* disruption strains of *S. pneumoniae* (Petit et al. 2001; Chimalapati et al. 2012) and *L. monocytogenes* (Machata et al. 2008)*,* with *L. monocytogenes* showing a decreased survival in macrophages (Baumgärtner et al. 2007). A general reduction in germination rate of *B. anthracis* due to the loss of Lgt function resulted in attenuated virulence (Okugawa et al. 2012). *Streptococcus* mutants have a more mixed effect showing *S. suis lgt*

mutants, although not attenuated in virulence (de Greeff et al. 2003) resulted in a reduced innate immune response (Schreur et al. 2011). *S. equi* has no significant effect (Hamilton et al. 2006; Das et al. 2009) while *S. sanguinis* however results in a minor reduction in endocarditis virulence (Das et al. 2009). More significantly hyper virulent strains have been reported as a result of *lgt* loss as is the case with *S. agalactiae* (Henneke et al. 2008). *Lsp* mutants have shown attenuated virulence in *L. monocytogenes* and *M. tuberculosis.*

5.2 Project background

This chapter focuses on investigating the lipoprotein biosynthetic pathway from *S. coelicolor* and progressing the work started in the Hutchings lab on *lsp* (Thompson et al. 2010). Following disruption of *sco2074,* the lipoproteins signal peptidase (*lsp*) gene, it was clear that secondary mutations had accumulated in strain BJT1000 as suggested by the authors. Morphologically BJT1000 had a small, flat poorly sporulating colony phenotype (Figure 5.7) in addition to surface lipoproteins being released into the supernatant. The release of the lipoproteins is hypothesised to be in response to the inappropriately processed lipoproteins retaining their signal peptide and causing some level of membrane instability. Strains BJT1004 and BJT1005 were produced by complementing the loss of *lsp* both *in cis* and *in trans* respectively*.* This led to a functional complementation of Lsp function and lipoprotein retention in the membrane but it did not complement the colony morphology defects. This lack of full complementation led to the assumption that secondary mutations had spontaneously occurred due to *lsp* being essential. In addition to this, single mutants were produced for the lipoprotein diacyglycerol transferase (*lgt*) genes, *lgt1* and *lgt2*. However, it has not previously been achieved to produce a double disruption mutant. It is hypothesised, based on these results that the lipoprotein synthetic pathway is essential in *S. coelicolor* unlike in other Gram-positive bacteria.

5.3 Aims

The aims of this work were four fold: 1) Identify any secondary mutations that occurred during the disruption of *lsp*. This was done by genome sequencing and bioinformatics analysis. 2) Attempt to determine the role of the secondary mutations by reintroducing WT DNA and carrying out genetic disruptions of loci of interest and investigating phenotypes. 3) Identify the source of the secondary mutations (i.e. at what point they were incurred). 4) Attempt to produce a clean *lsp*

strain that can be functionally and phenotypically complemented in lieu of identifying any suppressor mutations. We hypothesis that the disruption of *lsp* causes secondary mutations to occur, including suppressor mutations that prevent cell death.
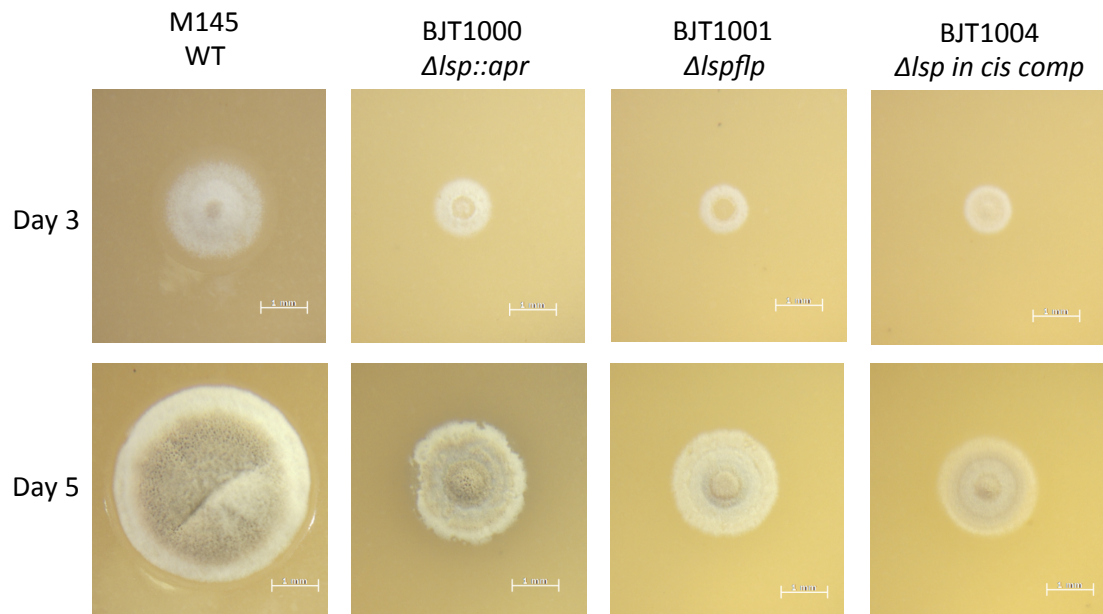
Figure 5-7 Light microscopy images of WT *S. coelicolor* M145, BJT1000, BJT1001 and BJT1004 at 3 and 5 days growth, grown on MS media at 30°C.

5.4 Results

5.4.1 The search for secondary mutations: large and small

5.4.1.1 Identification of SNP mutations and a transposon event

To identify the accumulated secondary mutations we carried out genome sequencing of the BJT1001 (Δlsp) strain and its parent strain M145. Two separate companies, TGAC (The Genome Analysis Centre) and GATC (GATC BIOTECH) carried out Illumina sequencing and the results were compared and mutations identified by Dr Govind Chandra at the John Innes Centre. From these results (summarised in Table 5-1) we were able to identify 51 SNPs, 13 found in only one of the BJT1004 sequences with 4 residing inside coding regions. However of all 51 SNPs only 1 of these occurs in both BJT1004 sequences and is found in the intergeneic space between *sco5331* and *sco5332*. In addition to this we were able to identify the insertion of an IS21 transposable element (*sco6393-4*), of the IS21/IS1162 family of transposases, between genes *sco6808-9* (Figure 5.8 a-b).

Confirmation of this was carried out by PCR amplification using JM0093+94 and both WT and BJT1001 genomic DNA as templates for the PCR (Figure 5.8 c). Supplementary data S5.1 and S5.2 contain the *sco6811-08* loci for the M145 and BJT1004 respectively.

	SNP			GATC[§]		TGAC[§]	
No.	Position[*]	Gene[†]	Change[‡]	M145	BJT1004	M145	BJT1004
1	104233	*sco0124*	---	Y	N	N	N
2	563185	*sco0531*	E128G	Y	N	N	N
3	619172	*sco0577*	---	Y	Y	N	N
4	657081	*sco0617*	E564D	Y	Y	Y	Y
5	1415134	*sco1337*	G15A	Y	Y	Y	N
6	1622503	*sco1516*	Y72N	Y	N	Y	Y
7	1625888	*sco1520*	R13C	Y	Y	Y	Y
8	1634749	*sco1529*	---	Y	Y	Y	Y
9	1642021	-	N/A	N	N	N	Y
10	1644238	*sco1536*	H262Q	Y	N	N	N
11[#]	1644332	*sco1536*	R230P	N	N	Y	Y
12[#]	1644333	*sco1536*	R230G	N	N	Y	Y
13	1649515	-	N/A	N	Y	N	N
14	1740776	-	N/A	N	N	Y	Y
15	2065371	-	N/A	N	N	N	Y
16	2065372	-	N/A	N	N	N	Y
17	2173610	*sco2026*	---	Y	N	N	N
18	2301530	*sco2139*	A249G	N	Y	N	N

19	2976040	*sco2729*	C112G	Y	N	N	N
20	3141224	*sco2886*	W84G	N	Y	N	N
21	3348737	*sco3054*	P36L	Y	Y	N	N
22	3349264	*sco3055*	G101R	Y	Y	Y	N
23	3958102	-	N/A	N	N	Y	N
24	4140160	-	N/A	N	Y	N	N
25	4140161	-	N/A	N	Y	N	N
26	4245354	*sco3860*	V261G	Y	N	N	N
27	4823788	*sco4405*	T219P	Y	N	N	N
28	4863506	-	N/A	Y	N	N	N
29	4863507	-	N/A	Y	N	N	N
30	4868999	-	N/A	Y	N	N	N
31	5016937	*sco4594*	---	N	N	N	Y
32	5018303	*sco4595*	---	N	N	Y	N
33	5044225	*sco4620*	---	N	N	N	Y
34	5095649	*sco4665*	E206A	Y	N	N	N
35	5633841	*sco5182*	R7G	N	N	Y	Y
36	5805609	-	N/A	N	N	N	Y
37	5805610	-	N/A	N	Y	N	Y
38	6367261	-	N/A	N	Y	N	N
39	6890540	*sco6265*	---	N	Y	N	N
40	7044234	*sco6381*	---	Y	Y	Y	Y
41	7118314	*sco6436*	V21L	Y	Y	Y	N

42	7146838	*sco6458*	S224G	Y	Y	N	N
43	7180216	*sco6487*	G263A	Y	Y	Y	Y
44	7213371	*sco6522*	A442G	Y	N	N	N
45	7261050	*sco6560*	L123V	N	N	Y	N
46	7284887	*sco6578*	F83V	Y	Y	N	N
47	7364061	*sco6635*	---	Y	N	Y	N
48	7393003	*sco6657*	R378S	Y	Y	N	N
49	8044092	*sco7236*	---	Y	Y	Y	N
50	8594044	*sco7763*	---	Y	Y	Y	N
51	8612266	-	N/A	Y	N	N	Y

[*] – SNP base positions within the genome *S. coelicolor* M145 genome.

[†] – Annotated *sco* gene number with "-" indicated an intergenic SNP.

[‡] – Amino acid change with "---" indicating no change and N/A for intergenic site.

[§] – The sequencing round responsible for each result.

[#] – The combination of SNP 11 and 12 would result in a R230A change.

Table 5-1 Identified secondary mutations in BJT1004. Re-sequencing and comparison of the parent strain M145 and the *cis* complemented *lsp* mutant BJT1004 revealed 51 putative single nucleotide polymorphisms (SNPs) between all 4 sequenced samples, of which 13 are unique SNPs detected in at least 1 of the BJT1004 sequences, 4 of these are within coding regions but only 1 SNP was detected in both BJT1004 sequences and is intergenic. One chromosomal rearrangement was detected (see also Fig 5-8).
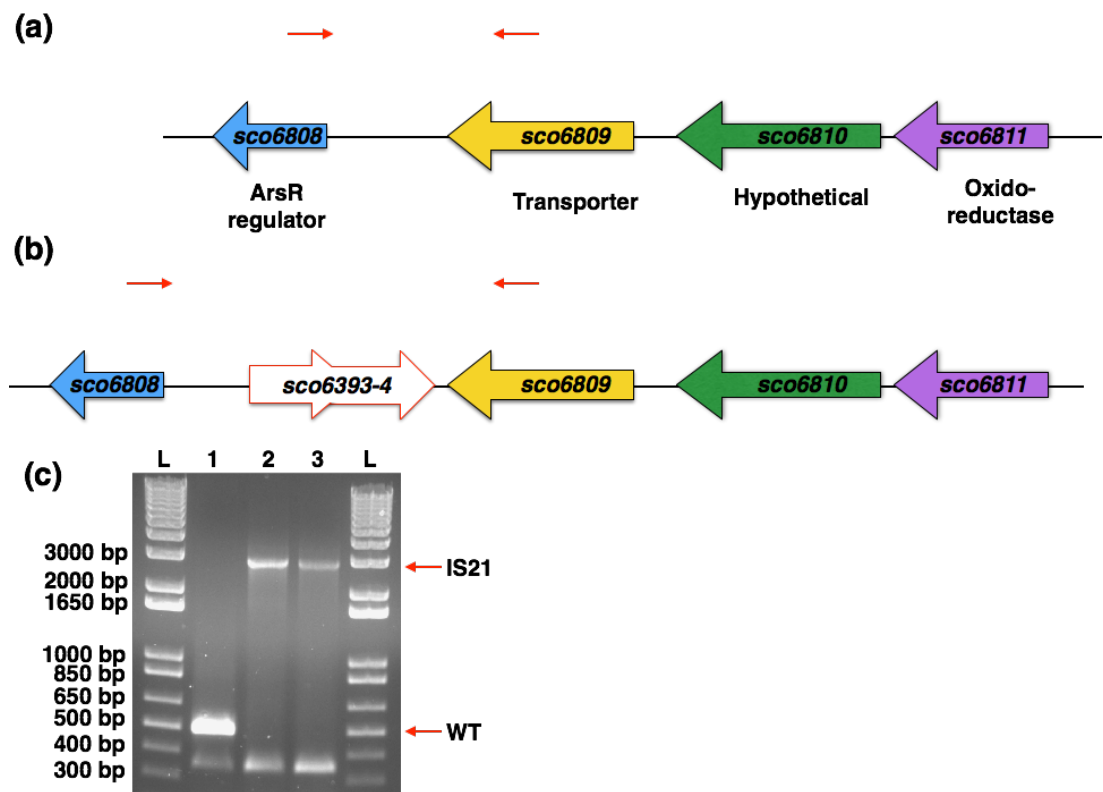
Figure 5-8 *IS21* insertion between *sco6808* and *sco6809.* The *sco6811-08* region of the *S. coelicolor* M145 genome contains *sco6811* - purple*, sco6810* - green *, sco6809* - yellow and *sco6809* - blue. The WT loci (a) and BJT1004 (b) indicate where an IS21 element (*sco6393* and sco6394) has inserted. PCR verification of this loci with primers JM0093 and JM0094 (small red arrows) was carried out (C) with template genomic DNA from M145 (lane 1), BJ1001 (lane 2) and BJT1004 (lane 3). Lanes marked L contain the size ladders (Invitrogen 1kb plus DNA ladder), lane 1 contains the PCR product using WT M145 DNA (514 bp), lane 2 contains the PCR product using Δ*lsp* strain BJT1001 DNA and lane 3 contains the PCR product using genomic DNA from the cis complemented Δ*lsp* strain BJT1004 (both 2884 bp).

All of the point mutations identified occur in only one set of sequencing data, resulting in conservative AA changes or fell within intergenic regions of DNA. This leads us to suggest that the IS21 element, its presence confirmed by PCR, is important in the regards to complementation issues of BJT1001. To determine what affect the IS21 element had, we investigated the putative operon of *sco6811-08* further.

5.4.1.2 Investigating the *sco6811-08* putative operon and IS21 elements in *Streptomyces*

To investigate the putative operon of *sco6811-08* we first took a brief bioinformatics approach to identify any functional characteristics of the genes already identified to attempt to link them to the phenotype. Previous work carried out on *sco6808* showed that a disruption mutant, when grown in minimal media results in an over production of Act and RED (Yang et al. 2008). Additionally, work carried out to investigate *sco6809* identified it as a AbrC, an atypical two-component system, target (Rico et al. 2014). The entire region is not well conserved among the species represented on StrepDB. Only *sco6809* and *sco6811* are conserved together with homologs in *S. scabies* and *S. venezuelae* (Figure 5.9). Interestingly upon investigating *S scabies* genome, we identified *scab23721-2* down stream of *scab23723* (the *sco6809* ortholog) another IS transposable element (however not of the IS21/IS1162 family). This is interesting in regards to the putative suppressor mutation hypothesis as disruption of *S. scabies lsp* resulted in no major morphological phenotypes (Widdick et al. 2011). Potentially this region is an insertion event hot spot and localisation to here results in suppression of a lethal *lsp* disruption phenotype. Although conjecture at this point, we wished to investigate further.
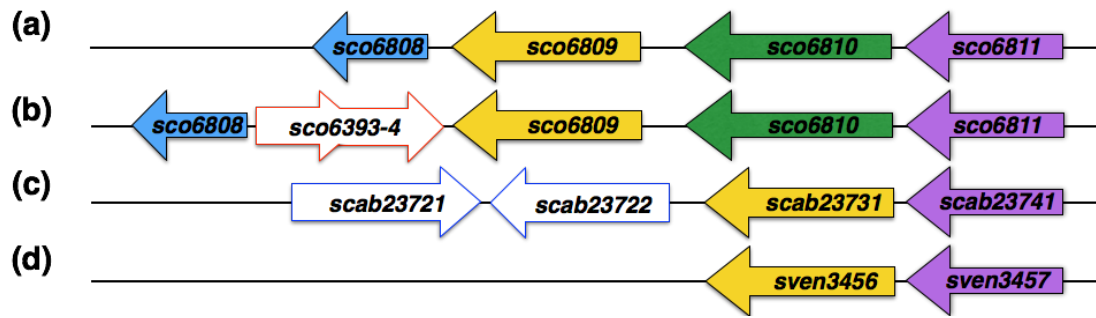
Figure 5-9 Homologous loci to the *sco6811-08* region for M145 (a), BJT1004 (b), *S. scabies* (c) and *S. venezuelae* (d). White arrows with red or blue outlines correspond to transposable elements.

5.4.2   Identification of *scr6809*: an unexpected gem

5.4.2.1   dRNA-seq analysis of the *sco6811-08* loci identified a sRNA, *scr6809*

Following this we began carrying out a mutational analysis of the *sco6811-08* operon while in parallel, utilising dRNA-seq data (GSM1121652 and GSM1121655), we were able to identify that the region contains three TSS driving genes *sco6811-10*, *sco6809* and *sco6808* respectively. In addition to this we also identified a putative sRNA, *scr6809*, between genes *sco6808* and overlapping a short portion of the N-terminus of *sco6809* (Figure 5.10). The sequence for *scr6809* can be found in Supplementary data S5.3. Following a BLAST analysis of the sequence we were unable to identify a homologous sequence in the other StrepDB representative strains (including *S. scabies* containing the IS element), with no reference to sRNA returned from a general BLAST analysis. At this stage we can infer that *scr6809* because of its homology/over lap with *sco6809* it may play a role in regulating *sco6809*.
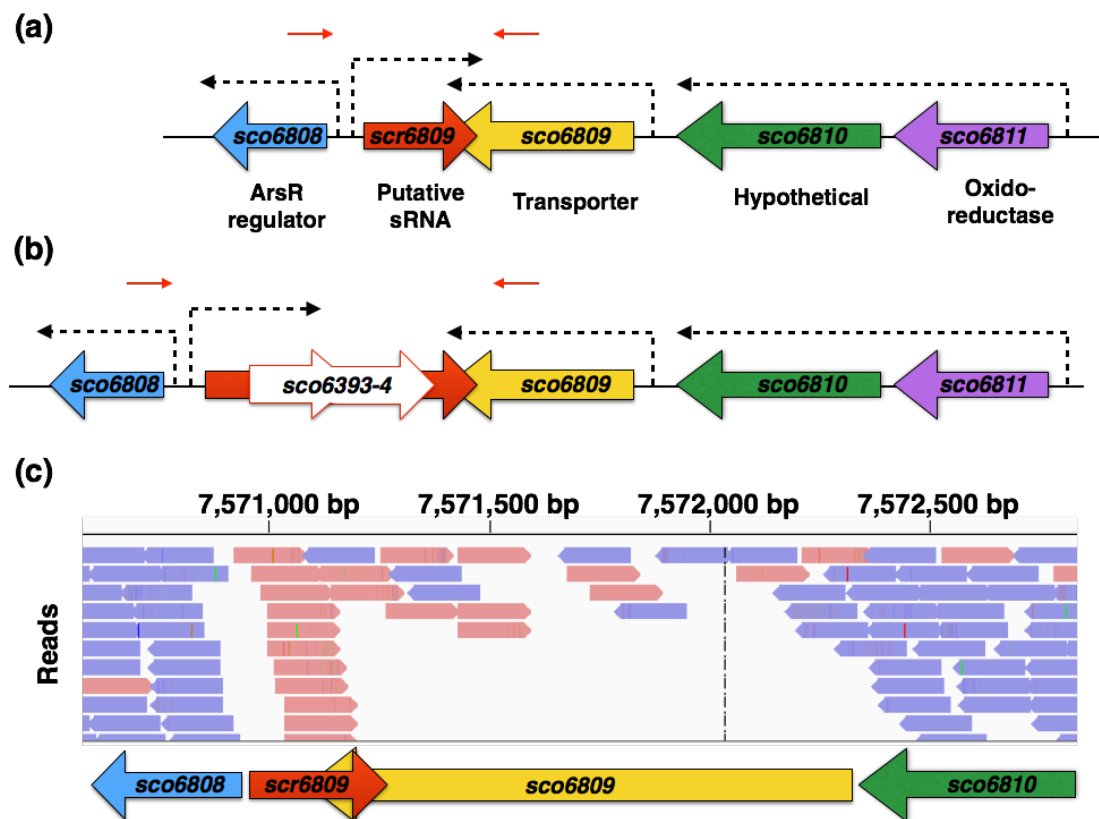
Figure 5-10 Updated graphical representation of the *sco6811-08* operon following analysis using RNA-seq data. The identified sRNA *scr6809* is highlighted in red and the 3 identified promoters shown in broken line arrows with the WT loci (a) and the BJT1001 loci (b) containing the IS21 element (*sco6393-4*). Also shown are the raw dRNA-seq reads (c) used to identify *scr6809* (GSM1121652) with the genes indicated below showing a genomic window spanning *sco6808 to part of sco6810*. Reads are represented as purple and red for forward and reverse sequence reads respectively.

5.4.2.2 Mutational analysis of the *sco6811-08* loci indicates the importance of *scr6809* in cell stability

To further understand the gene set, which we originally thought to be a putative operon, we produced four disruption constructs: 1) *sco6808::apr*, (2) *sco6811::apr*, (3), *sco6811-08::apr* and (4) *scr6809::apr*. All of these constructs were produced as described utilising cosmid St1A2 (Bentley et al. 2002) and confirmed by PCR (Figures 5.11-5.14). Following production, each construct was in into both WT *S. coelicolor* M145 and mutant BJT1004 strains. The major reasons for this is due to the system relying on homologous recombination there was a good chance that

recovery of *scr6809* would occur simultaneously with the first two disruption constructs.
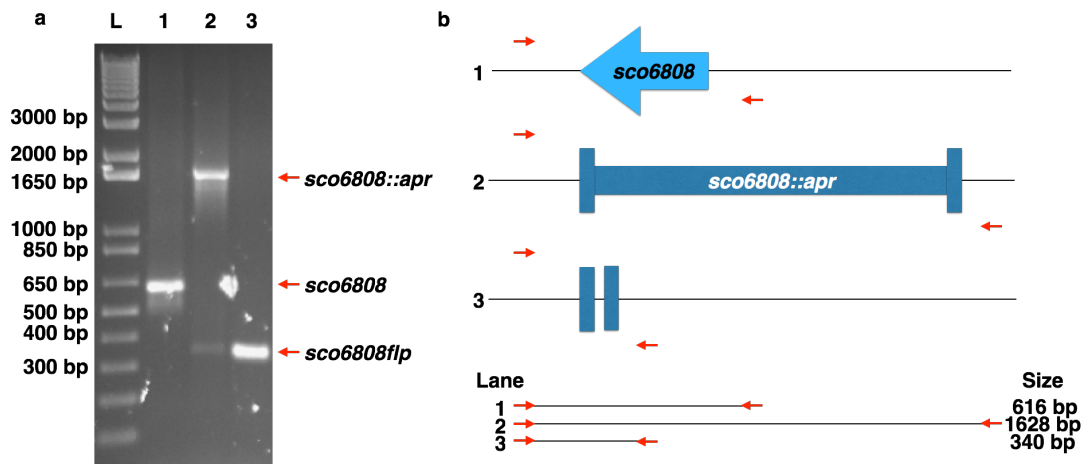
Figure 5-11 Agarose gel electrophoresis of PCRs using primers JM0085+86 to amplify the *sco6808* loci from the St1A2 disruption cosmids. (A) showing the PCR results of Lane 1) WT *sco6808* (616 bp) amplified from the unmodified St1A2 cosmid, lane (2) *sco6808::apr* (1628 bp) amplified from the marked disruption cosmid and lane (3) *sco6808flp* (340 bp) amplified from the flp cosmid. (B) illustrates the genotype of the WT, *sco6808::apr* and *sco6808flp* mutants and the expected PCR products along with the predicted sizes. Invitrogen 1 kb plus DNA ladder (L) was used.
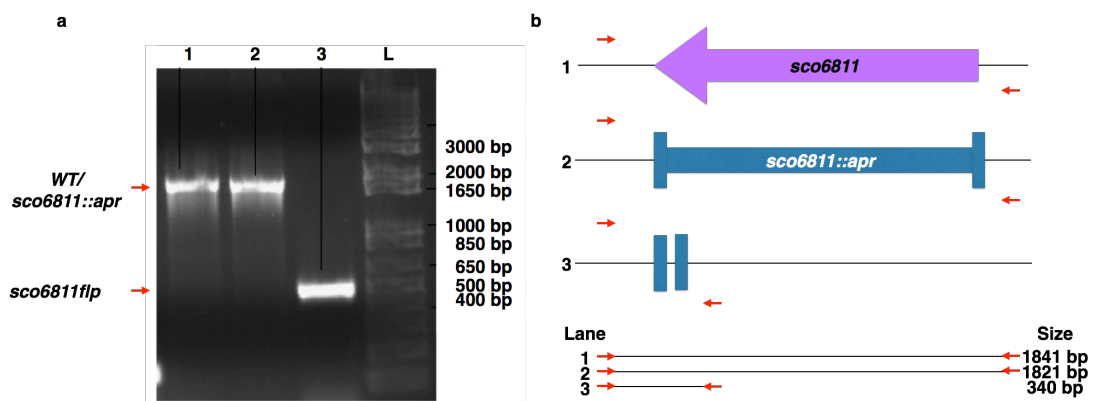
Figure 5-12 Agarose gel electrophoresis of PCRs using primers JM0089+90 to amplify the *sco6811* loci from the St1A2 disruption cosmids. Lane 1) WT *sco6811* (*1841* bp) amplified from the unmodified St1A2 cosmid, lane (2) *sco6811::apr* (*1821* bp) amplified from the marked disruption cosmid and lane (3) *sco6811flp* (533 bp) amplified from the flp cosmid. (B) Illustrates the genotype of the WT, *sco6808::apr* and *sco6808flp* mutants and the expected PCR products along with the predicted sizes. Invitrogen 1 kb plus DNA ladder (L) was used.
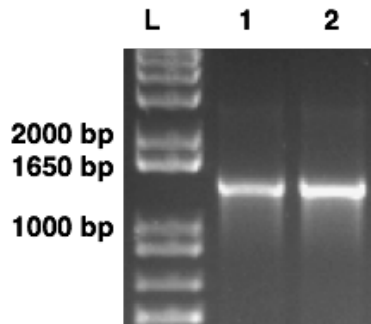
Figure 5-13 Agarose gel electrophoresis of PCRs amplifying the *sco6811-08* loci from the St1A2 disruption cosmids. Lane 1) JM0090+81 primer set used (1616 bp) amplified from *sco6811-08* cosmid, lane (2) JM0085+82 (1628 bp). These primers correspond to *sco6811* and *sco6808* specific primers and cassette specific primers. Invitrogen 1 kb plus DNA ladder (L) was used.
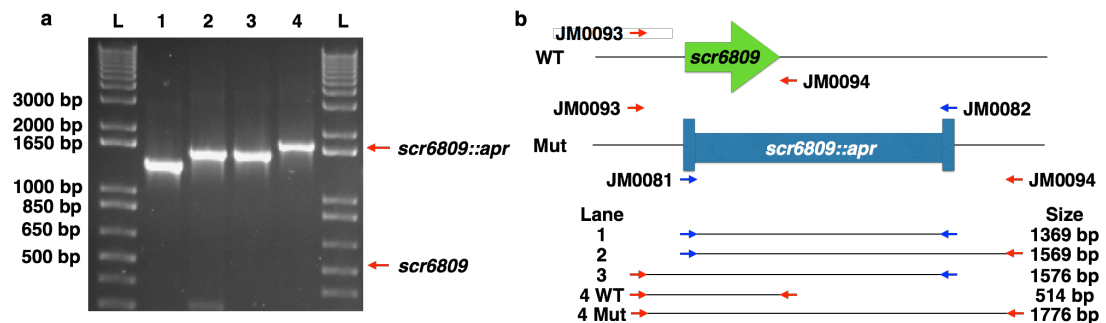
Figure 5-14 Agarose gel electrophoresis of PCRs using primers JM0093+94 and JM0081+82 in combination to amplify the *scr6809* loci from the St1A2 *sco6809::apr* disruption cosmid. Lane 1) disruption cassette amplified using primers JM0081+ JM0082 (1369 bp), lane (2) JM0081 and JM0094 (1569 bp) amplifying the disruption cassette using flank DNA, lane (3) JM0082 and JM0093 (1576 bp) amplified the disruption cassette using flank DNA and lane (4) JM0093+94 (1776 bp) amplifying the disruption cassette and flanking DNA confirming location where the WT band would be at 514 bp. (B) Illustrates the genotype of the WT and *scr6809::apr* mutants and the expected PCR products from each primer set along with the predicted sizes. Invitrogen 1 kb plus DNA ladder (L) was used.

Disruption mutants of each were produced, *sco6811::apr* and *sco6808::apr* strains were confirmed by PCR (Figures 5.15 and 5.16) and light microscopy carried out to investigate colony morphology under standard growth conditions (Figure 5.17). *sco6811::apr* strains were confirmed by sequencing because the disruption product and WT gene are very similar in size (Figure S2). Disruptions of either *sco6808* or

sco6811 in the WT background had no obvious phenotype, colony morphology and antibiotic production looked similar to WT M145.
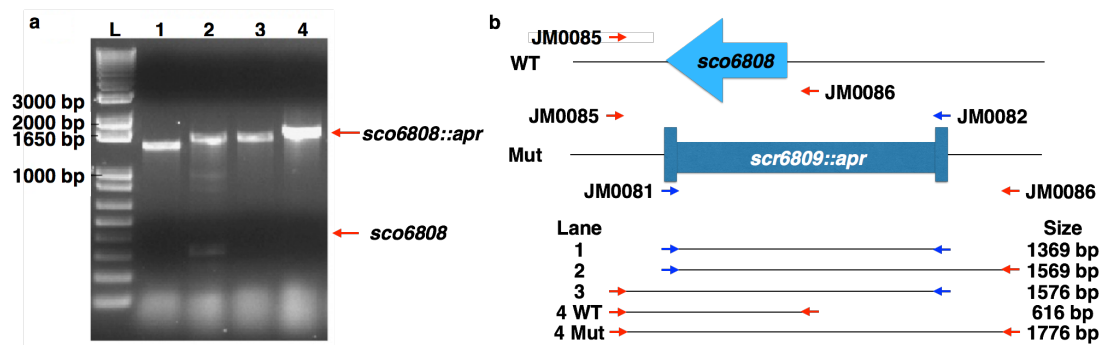
Figure 5-15 Agarose gel electrophoresis of PCRs using primers JM0085+86 to amplify the *sco6808* loci from M145 *sco6808::apr* genomic DNA. Lane 1) disruption cassette amplified using primers JM0081+ JM0082 (1369 bp), lane (2) JM0081 and JM0086 (1569 bp) amplifying the disruption cassette using flank DNA, lane (3) JM0082 and JM0085 (1576 bp) amplified the disruption cassette using flank DNA and lane (4) JM0093+94 (WT= 616 bp *sco6808::apr* = 1776 bp) amplifying the disruption cassette and flanking DNA confirming location. (B) Illustrates the genotype of the WT and *scr6809::apr* mutants and the expected PCR products from each primer set along with the predicted sizes. Invitrogen 1 kb plus DNA ladder (L) was used.
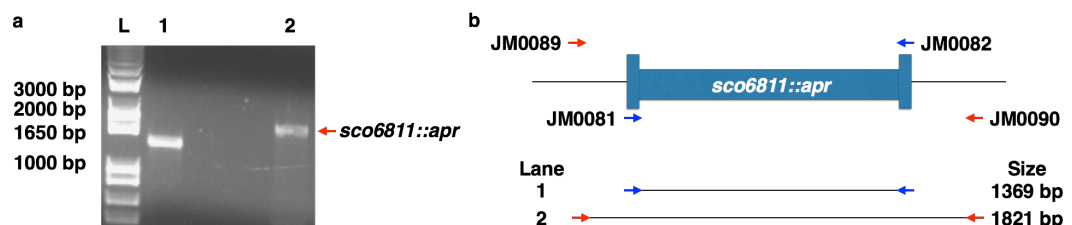
Figure 5-16 Agarose gel electrophoresis of PCRs using primers JM0089+90 to amplify the *sco6811* loci from M145 *sco6811::apr* genomic DNA. Lane 1) disruption cassette amplified using primers JM0081+ JM0082 (1369 bp), lane (2) JM0089+90 (1821 bp) amplifying the disruption cassette and flanking DNA confirming location. (B) Illustrates the genotype of the WT and *sco6811::apr* mutants and the expected PCR products from each primer set along with the predicted sizes. Invitrogen 1 kb plus DNA ladder (L) was used.

185
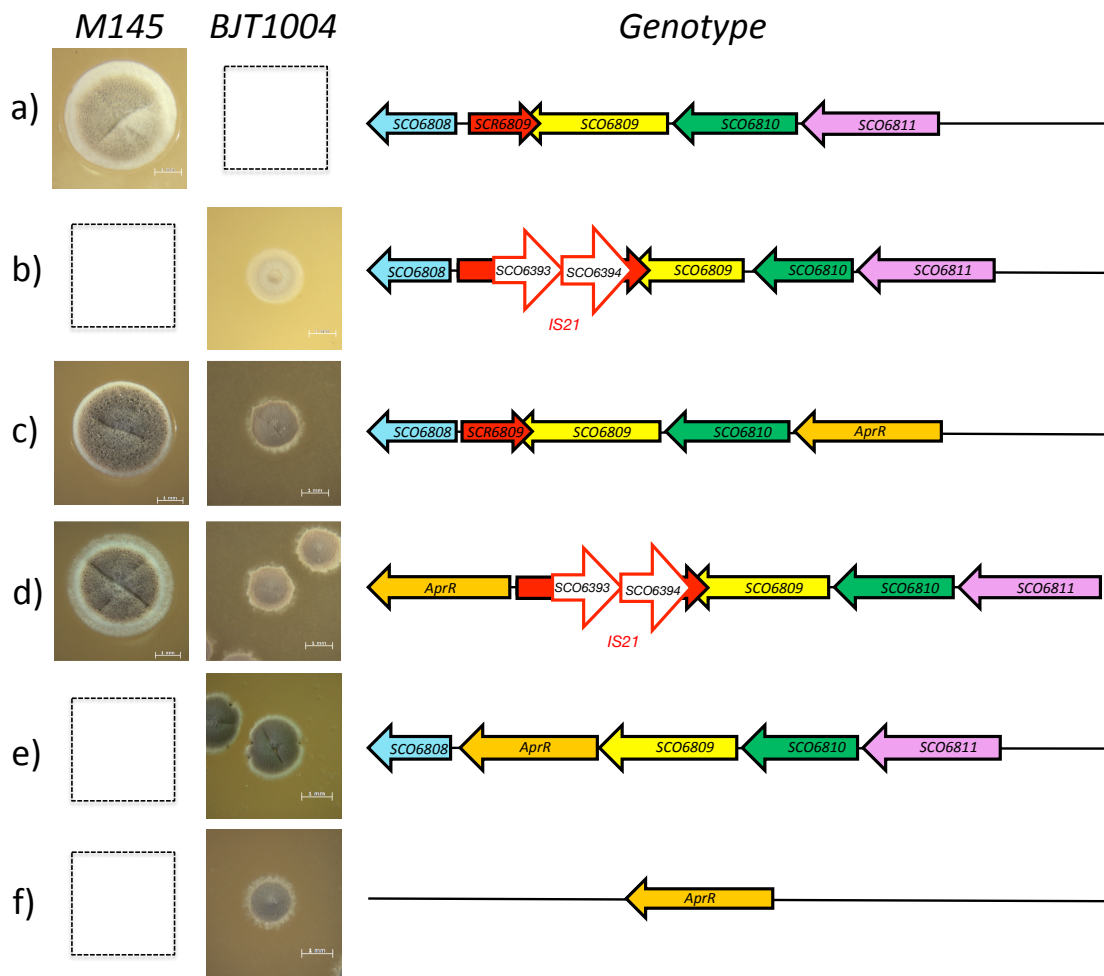
Figure 5-17 Phenotype (light microscopy images) and genotype (graphical representations) of WT M145 and BJT1004 strains with a combination of *sco6811-08* region disruption mutations. (a) M145, (b) BJT1004 with the IS21 element inserted into *scr*6809, (c) *sco6811::apr* disruptions of both strains, (d) *sco6808::apr* disruption of both strains, (e) *scr6809::apr* disruption in BJT1004 and (f) *sco6811-08::apr* disruption of BJT1004. Broken line boxes in (a) and (b) represent invalid strains and in (e) and (f) represent strains where a single phenotype was not representative of the disruption process.

Disruption mutants of *sco6811-08::apr* and *scr6809::apr* in the BJT1004 background both appeared to have partially recovered sporulation over the BJT1004 parent strain but the *sco6811-08::apr* had less than either *sco6811::apr* or *sco6808::apr* (quantified, based on spore preparation observations). However, this is strikingly different to the phenotype observed for disruptions in the WT M145 parent strain. Disruptions of both *sco6811-08* and *scr6809* resulted in similar phenotypes (Figure 5.19). Following replica-plating to identify double crossover candidates, colonies with WT characteristics were streaked forward. Within the first generation severe and broad ranging morphological phenotypes were observed. Representatives of the major phenotypes were all identified: bald, white, a range small-normal colony size, over producers of antibiotics (ACT specifically). Streak purifying mutated, non-WT looking colonies resulted in no further mutations within the next generation. Streak purifying WT looking colonies resulted in repeated development of the phenotypes seen with the original streak purification.
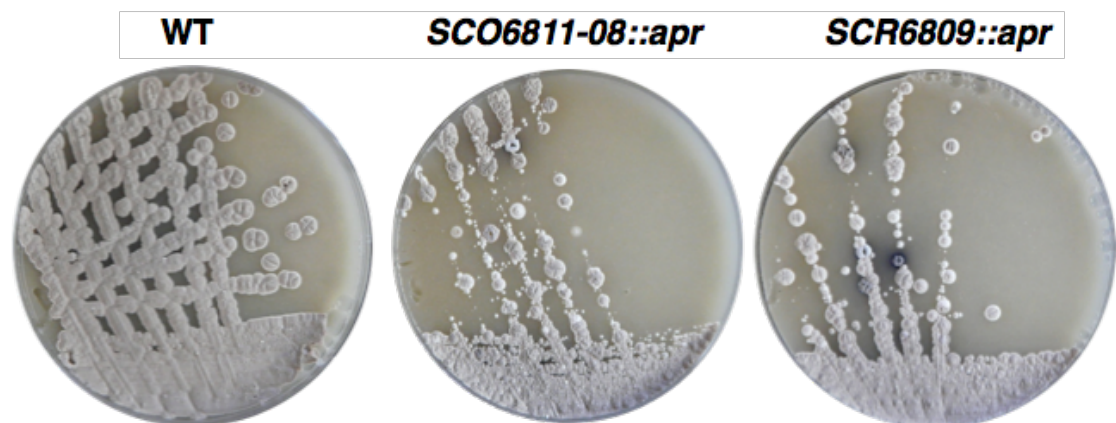
Figure 5-18 Photographs containing streak purified WT *S. coelicolor* M145, M145 *sco6811-08::apr* and M145 *scr6809::apr* double crossover mutants.

5.4.3 Not a suppressor story but a cautionary tale: the risks of recombineering

5.4.3.1 IS21 insertion into *scr6809* is not an *lsp* suppressor mutation

Following the identification of *scr6809* and the important role it plays in cell stability links to the *lsp* phenotype were investigated. This was carried out by repeated knockouts of *lsp* and looked for disruption of the *scr6809* loci, particularly with the IS21 element. The *lsp::apr* cosmid was produced using St4A10 as was carried out

originally using the same primers and confirmed it by PCR (Figure 5.20) and introduced it in to M145. A range of putative double crossover strains (based on replica-plate results) were isolated, spore stocks produced for each strain and PCRs ran to confirm the crossover state and *scr6809* genotype (Figure 5.21). Δ*lsp* strains produced included both single and double crossover mutant (n=27) with 8 used for analysis (single n=2 and double n=6). It was clear from our analysis that the insertion of the IS21 element at the *scr6809* loci was not a common occurrence, at least not in regards to acting as an *lsp* suppressor mutation. We suspect that the launch of the element is in response to whatever is causing the secondary mutations.
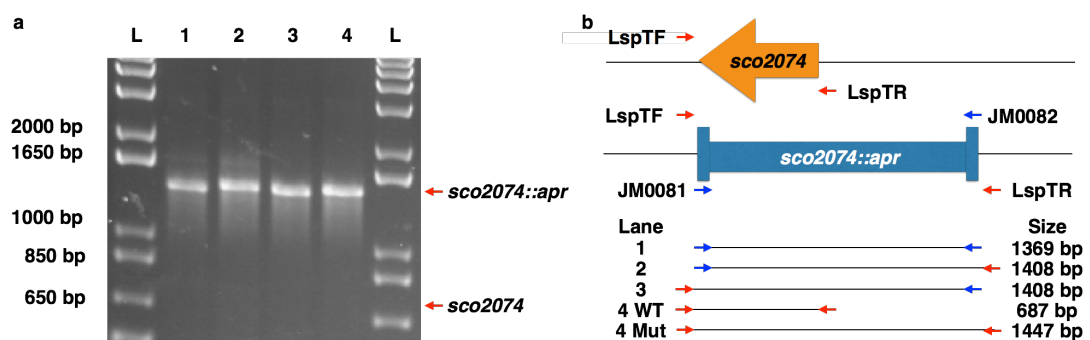
Figure 5-19 Agarose gel electrophoresis of PCRs using primers LspTF+LspTR (Thompson et al. 2010) and JM0081+JM0082 in combination to amplify the *sco2074* loci from the St4A10 *sco2074::apr* disruption cosmid. Lane 1) disruption cassette amplified using primers P1+P2 (1369 bp), lane (2) JM0081 and LspTR (1408 bp) amplifying the disruption cassette using flank DNA, lane (3) JM0082 and LspTF (1408 bp) amplified the disruption cassette using flank DNA and lane (4) LspTF+LspTR (1447 bp) amplifying the disruption cassette and flanking DNA confirming location where the WT band would be at 687 bp. (B) illustrates the genotype of the WT and *sc02074::apr* mutant and the expected PCR products from each primer set along with the predicted sizes. Invitrogen 1 kb plus DNA ladder (L) was used.
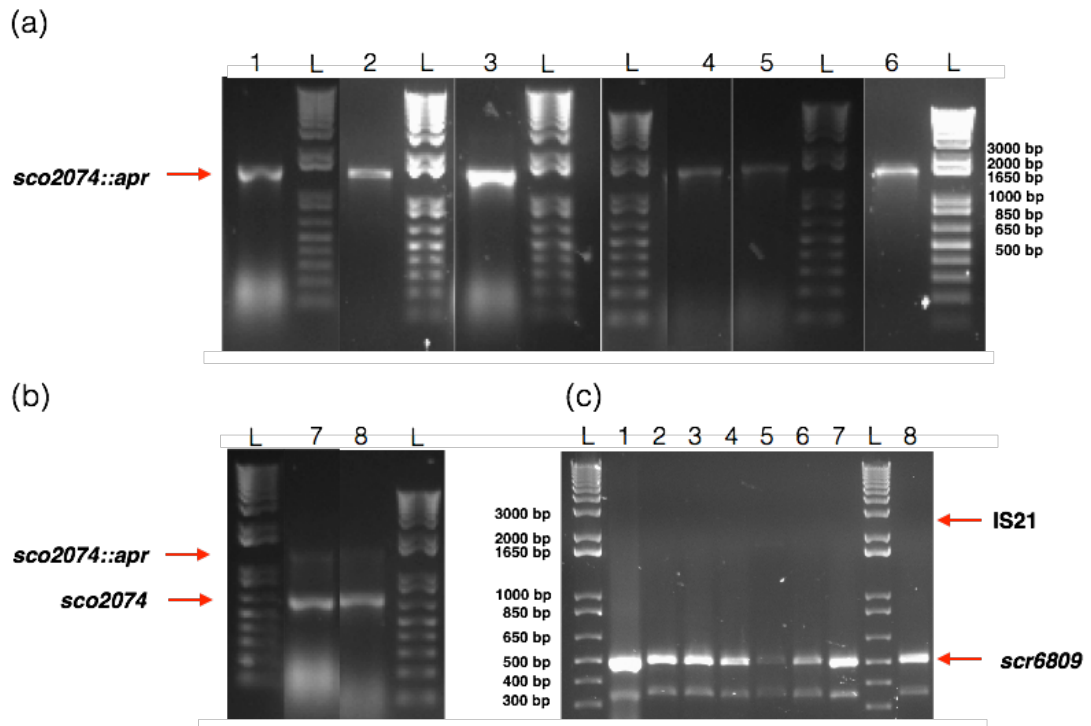
Figure 5-20 Agarose gel electrophoresis of PCRs using primers LspTF +LspTR (a) and (b) and JM0093+94 (c) to correlate *lsp* disruption and frequency of IS21 insertion into *scr6809*. (a) Double crossover mutants (1-6) of M145 *sco2074::apr* showing disruption (1447 bp) band, (b) single crossover mutants (7-8) with both WT (687 bp) and disruption band and (c) PCR amplification of *scr6809* loci showing the WT (514 bp) product with the predicted size highlighted if an IS21 element had transposed. Invitrogen 1 kb plus DNA ladder (L) was used.

5.4.3.2 Introduction of the *lsp* containing cosmid, St4A10, causes secondary mutations

While carrying out the frequency experiment it was observed that not only double crossover strains, but single crossovers (those that have fully functional *lsp* genes) also had abnormal phenotypes. This led us to hypothesis that *lsp* is not essential but the processing of the disruption, at some stage likely unrelated to the loss of *lsp*, is disturbing the organism. In an attempt to elucidate what this might be we first introduced cosmid St4A10 *bla::hyg*, which has had an *oriT* introduced to the superCos backbone to facilitate conjugation. This was confirmed by PCR (Figure 5.22). Following the conjugation we observed that almost all colonies had suffered a similar morphological change (Figure 5.23). This highlighted that introduction of the cosmid itself was the cause of the phenotypes seen and not solely the loss of *lsp*. Independent confirmation of this was received from another lab actively using the 4A10 cosmid (Gillespie and Kelemen, personal communication).

5.4.3.3 Analysis of the St4A10 loci highlights important cell division genes

Following this, investigation into why introduction of this cosmid specifically would be causing such toxic effect to the organism was undertaken. Bioinformatics analysis of the genes contained within the cosmid (Table 5.2). We quickly realised that the cosmid contains many of the genes associated with cell wall biogenesis and cell division. Many of these genes are likely to be closely regulated and duplication appears to be detrimental to the organism. This led us to produce a disruption construct for *lsp* that would have limited to no effect on surrounding genes. This hypothesis will be discussed further in section.
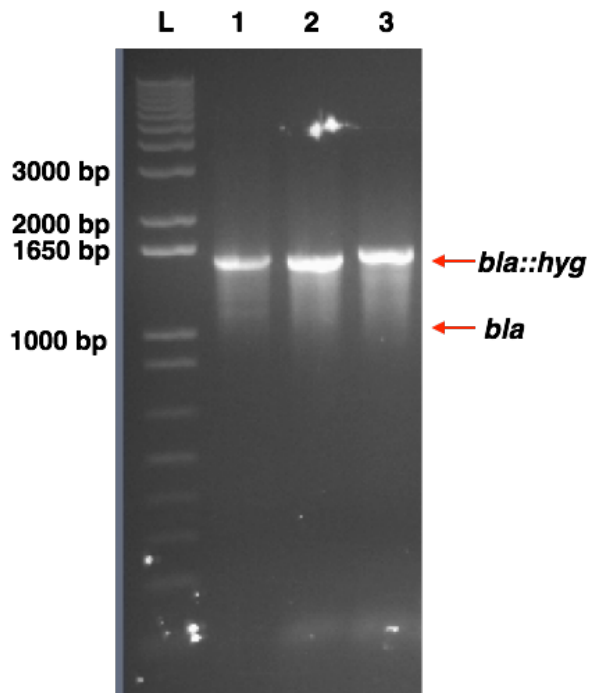
Figure 5-21 Agarose gel electrophoresis of PCRs using primers JM0097+98 on St4A10 *bla::hyg* cosmids. Confirmation of successful replacement of *bla* gene (1175 bp) with *hyg* cassette (1770 bp) containing an *oriT*.
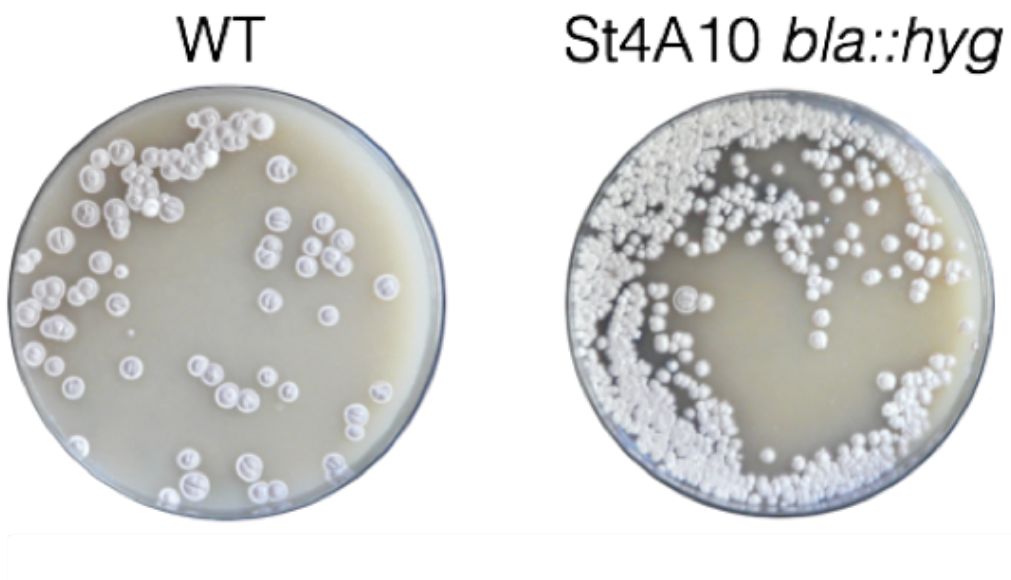
Figure 5-22 Photograph of plates containing M145 *lgt1::apr* exconjugants (WT representative) and M145 4A10 *bla::hyg* exconjugants.

Sco number	Gene name	Protein function
sco2068		Hypothetical
sco2069		Hypothetical
sco2070		Putative membrane protein
sco2071		Putative antiporter
sco2072		Hypothetical
sco2073		Possible ribosomal large subunit pseudouridine synthase
sco2074	lsp	Lipoprotein signal peptidase
sco2075		Putative DNA-binding protein
sco2076		Probable isoleucyl-tRNA synthetase
sco2077	divIVA	Division site selection protein
sco2078		Putative membrane protein
sco2079		Conserved hypothetical protein
sco2080		Conserved hypothetical protein
sco2081		Conserved hypothetical protein
sco2082	ftsZ	Cell division protein
sco2083	ftsQ	Required for efficient sporulation, but not growth and viability
sco2084	murG	Generates lipid II
sco2085	ftsW	Flippase for lipid II.
sco2086	murD	Adds second amino acid to the growing pentapeptide chain on cell wall precursor (in the cytoplasm)

sco2087	murX (mraY)	Generates lipid I
sco2088	murF	Adds D-Ala-D-Ala dipeptide to complete the pentapeptide chain on cell wall precursor (in the cytoplasm)
sco2089	murE	Adds third amino acid to the growing pentapeptide chain on cell wall precursor (in the cytoplasm)
sco2090	ftsI	PBP3 - transpeptidsase involved in cell division, interacts with FtsW
sco2091	ftsL	Possible membrane protein
sco2092		Conserved hypothetical protein
sco2093		Conserved hypothetical protein
sco2094		Transcription factor
sco2095		Putative membrane protein
sco2096		Putative membrane protein
sco2097		Putative membrane protein
sco2098		Possible methyltransferase
sco2099		Hypothetical
sco2100		Transcription factor
sco2101		Transposon
sco2102		Putative membrane protein
sco2103	metF	5,10-methylenetetrahydrofolate reductase
sco2104		Possible thiamin phosphate pyrophosphorylase

Table 5-2 A summary of genes contained within cosmid St4A10.

5.4.4   Disruption of *lsp*: a messy job made simple

5.4.4.1   Production of a *lsp* suicide, disruption vector pLAS

Following the identification of the recombineering issues associated with using the ReDirect system and the large insert cosmid clones, we instead produced an *lsp* suicide vector that would insert into (via single crossover) and disrupt the gene by inserting the entire vector within *lsp*.

To produce this vector we introduced a 411 bp fragment of central *lsp* with an N-terminal *Bam*HI site (N-terminal) into pGEM-T-EZ. The *Bam*HI site was then used to sub clone the *Bam*HI fragment from a pIJ773 digest, containing an *apr* disruption cassette. Following this the vector was then introduced into WT M145 and selected using apramycin resistance. Production of the vector was confirmed by PCR.

5.4.4.2 Disruption of *lsp* using pLAS

Following introduction of the pLAS vector into M145, two phenotypes were immediately observed (Figure 24): WT colony morphology and that of a small colony phenotype, overproducing actinorhodin, reminiscent of the original *lsp* mutants. Both strains were colony purified and PCRs were carried out to determine if *lsp* had been successfully disrupted (Figure 25). In each case tested the WT morphology was associated with a fully functional *lsp* gene and the second phenotype associated with pLAS insertion into *lsp*. An interesting recombination event had occurred in processing the strains. Following PCR and sequencing of the loci, we seen that most of the pLAS vector and almost the entirety of the *lsp* gene had been removed and all that remained was the apramycin resistance cassette (Supplementary data S5.3).

(a) M145

(b) M145 pLAS (1)   (c) M145 pLAS (2)
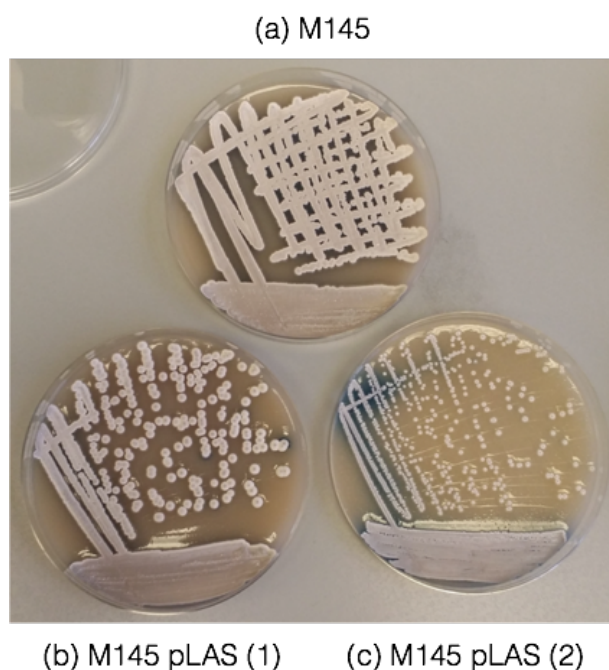
Figure 5-23 Photographs showing the two Morphologies associated with insertion of the pLAS vector into WT *S. coelicolor* M145. (a) Represents the M145 control, (b) the WT looking colonies, pLAS mutants (1) and (c) represents the small, antibiotic producing colonies, pLAS mutant (2).
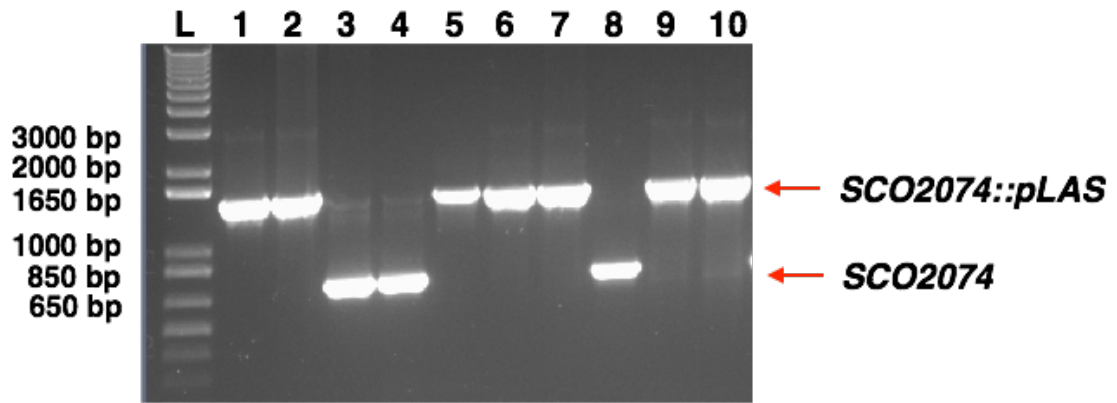
Figure 5-24 Agarose gel electrophoresis of PCRs to confirm disruption of *lsp* gene by insertion of the pLAS vector. Primers LspTF+TR were used on genomic DNA harvested from pLAS mutants n=10.

5.5 Summary

We have been able to determine that the *lsp* mutant phenotype is in part due to the insertion of the disruption cosmid, likely as a result of duplication of important genes including cell wall biosynthesis genes and in part due to the loss of Lsp function. *In trans* complementation of a suicide vector produced disruption almost fully complements the *lsp* disruption phenotype. Secondary mutations incurred as part of the original disruption, including the disruption of an sRNA by a transposable element, are not suppressor mutations or do not appear to be directly linked to the *lsp* phenotype. We identified a new, potentially important sRNA, *scr6809* that is clearly linked to a range of serious, developmental and survival phenotypes when disrupted. We have also highlight some key issues with recombineering technologies in *S. coelicolor* in a timely manner with the advent and release of two new systems utilising SceI meganuclease and the CRISPR/Cas9 systems.

5.6 Discussion

We initially hypothesised that *lsp* and indeed the lipoprotein pathway was essential in *S. coelicolor*. Following our work we are convinced that the lipoprotein biogenesis pathway is not essential in *S. coelicolor* however we suspect that there are key lipoproteins important for normal cellular development. The disruption process using the ReDirect system ultimately contributed to the initial reported *lsp* phenotype (Thompson et al. 2010). However, so far, we have not been able to completely salvage the WT phenotype in the BJT1004 background. Complementation studies of pLAS disruption strains will hopefully be informative in

this matter and embody our primary future work. We have inferred based on our data that part of the BJT1004 mutant phenotype is due to the duplication of important cell division genes. Many lipoproteins have been reported to be involved with homeostasis of the cell envelope and will also play a role in the phenotype. However introducing a second copy of *lsp*, which was shown to functionally complement Lsp activity, should result in a complete recovery of phenotype unless secondary mutations have occurred. We have seen a dramatic growth and developmental phenotype following the disruption of *scr6809* that was initially disrupted in the *lsp::apr* background. However we do not suggest that *scr6809* is linked to the phenotype directly but as a chance disruption by the IS21 launch based on the multiple *lsp* disruption mutants tested which lacked *scr6809* being disrupted. The implication here is that at least one of the point mutations is responsible in part for the BJT1004 phenotype unless there have been epigenetic effects (e.g. rearrangements) which have not been identified in the genome sequencing.

5.6.1 Why does St4A10 cause such an issue?

We wish to discuss the evidence we have as to why introduction of St4A10 would cause such a dramatic phenotype by duplication of the genes it carries. To do so we wish to discuss the roles of each gene within the cosmid and any phenotypic evidence associated with non-standard expression. Any phenotypic changes however arising from the transient duplication should be themselves, transient, unless secondary mutations have occurred which should have been identified during the sequencing. The implication here again, is that the SNPs identified (or an epigenetic modification) are playing a role.

5.6.1.1 *divIVA, ftsZQW* and *murDEFGX* – polar growth cell division and cell wall biogenesis.

Many of the genes discussed here are key players in growth, cell division and cell wall biosynthesis as illustrated in Figure 8-28. DivIVA, although linked to many functional roles, is responsible for the recruitment of additional proteins to the cellular poles (or hyphal tip) to fascinate cell division and growth (Flärdh et al. 2012). DivVIA is non-essential in Firmicutes but is essential in the high GC-rich Actinobacteria for polar growth. Depletion studies of DivIVA (10% of WT) showed irregular hyphae shape, with an increased number of branches forming near the hyphal tip (Flärdh 2003). Over-expression studies of DivIVA (25 fold increase on

197

WT) showed cells became swollen, shorter and thicker that, although formed branches, they appeared significantly rounder at their ends (Flärdh 2003). We have not investigated the hyphal tips of BJT1004 specifically however from SEM studies we do not appear to see the specific phenotypes referred to variations in DivIVA expression (Thompson et al. 2010).
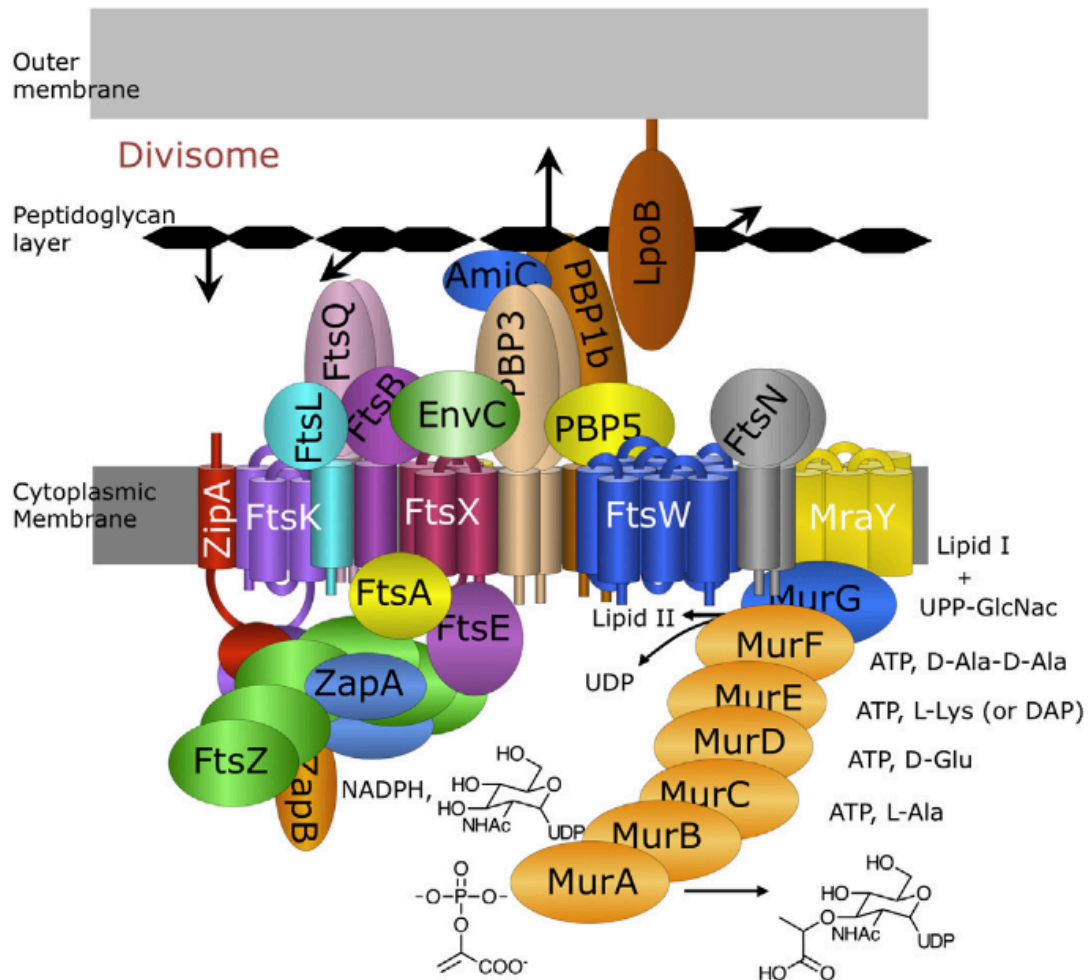
Figure 5-25 The divisome of *E. coli.* Many of the associated St4A10 cosmid genes are represented and the Mur pathway of Lipid I and II production is shown. Image taken from (Den Blaauwen et al. 2014).

FtsZ is involved in cell division in all bacteria, forms a Z-ring, which on constriction, facilitates cell division. Filamentous bacteria like *Streptomyces* don't require FtsZ for polar growth or indeed for survival (McCormick 2009). Typically FtsZ mutants are disrupted in growth in varying fashions due to a reduction in how robust hyphae are (lose of cross walls during vegetative growth means there is not a barrier

preventing complete cellular death from a single envelope fracture). FtsQ is involved in Z-ring stabilisation and is essential for efficient sporulation but is not essential growth (McCormick & Losick 1996). FtsW, reported as the lipid I flippase, is involved in septation and sporulation based on null-mutant studies depending on the media type with no other developmental issue (Bennett et al. 2009). We suspect that no lingering effects are caused by the duplication of these genes due to the completion of sporulation in BJT1004 based on the SEM data provided (Thompson et al. 2010).

The function of many of the Mur proteins can be summarised in Figure 5.30. These genes are involved in the synthesis of lipid I and II the precursors utilised during peptidoglycan biosynthesis (Kouidmi et al. 2014).
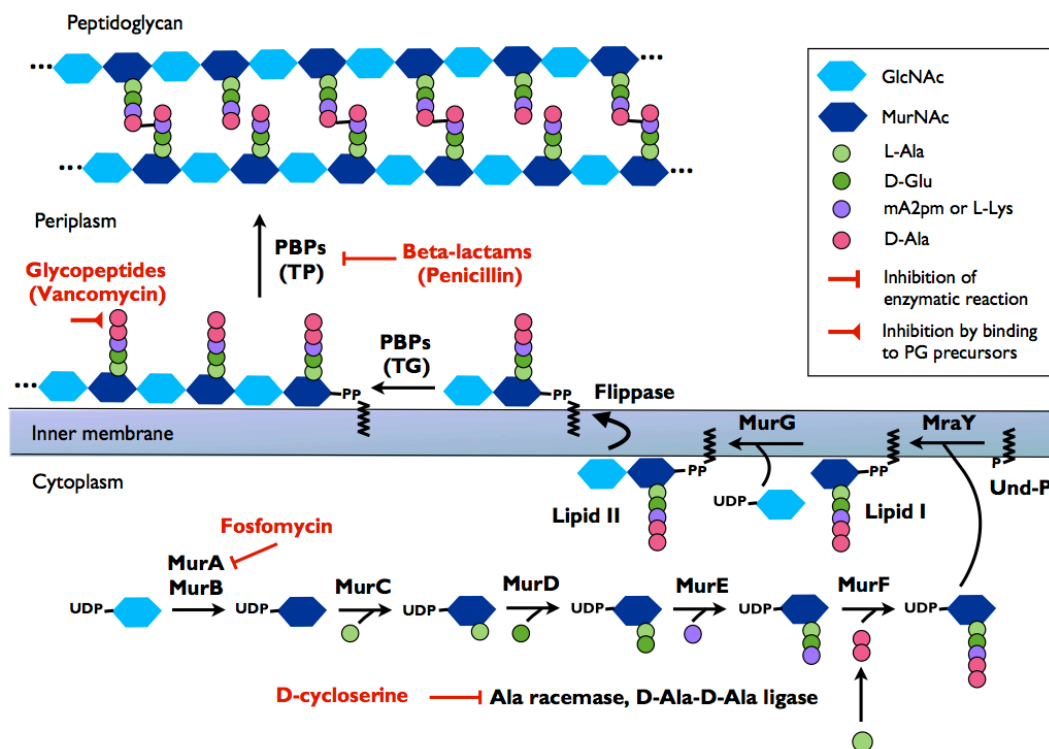
Figure 5-26 A summary of cell wall biosynthesis showing the major components, many of which are associated with St4A10. Image taken from (Kouidmi et al. 2014).

None of the St4A10 associated genes have reported phenotypes that result in a directly similar phenotype seen in BJT1004. At this point it is unclear why BJT1004 with the *sco6811::apr* disruption (unless it is due to the loss of *sco6811)* doesn't recover, assuming the uninvestigated SNPs are not involved. We have at least indicated that a frequent disruption of *scr6809* is not linked to the major *lsp* mutant

phenotype that arises from insertion of St4A10 or the loss of *lsp*. We haven't however indicated why BJT1004 won't recover (section about pLAS strain).

5.6.2 scr6809

Following our work we have uncovered two major interesting facts, based on our available data. Firstly, the disruption of *scr6809* results in a broad range of developmental phenotypes. Secondly, we have indicated under the conditions tested that overexpression of *scr6809* is not toxic to the cell nor does it cause any defects under our standard lab growth conditions. However we have not yet confirmed by northern blot that *scr6809* is indeed over or at all expressed from the vectors. We currently are unsure of the function of *scr6809*, however our data suggests it is involved in important cellular functions. To further investigate *scr6809* we wish to determine if the introduction of a second copy followed by the disruption of *scr6809* still yields the same phenotype.

5.6.3 IS21 element launching

Finally we would like to comment on the initial insertion of the IS21 element. It is well accept that the launch of a transposon will occur under stress conditions. It is possible that the insertion into *scr6809* was by chance, as the data suggests. It could be that in other strains, the transposon is inserting frequently but broadly throughout the genome. This could contribute to any fluctuations in the phenotype as observed during the frequency of *scr6809* disruption tests in section 5.4.3.1. To carry out a screening process to identify where the transposons insert, genome sequencing would be an invaluable tool. There are several options however the most economically efficient for both time and man hours, genome sequencing of each strain would allow for the identification of any transposon launch as well as identify any additional mutations. Recent advancements have allowed for exceptionally cheap genomic sequencing by MicrobesNG (http://microbesng.uk) who provide sequencing from £50 per sample. This alone would suffice in identifying the loci of insertion. Historically screening of shotgun libraries would be more economically viable however the reagent and time cost would ultimately result in higher costs.

6   Bibliography

Agar, J.N. et al., 2000. IscU as a scaffold for iron-sulfur cluster biosynthesis: Sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU. *Biochemistry*, 39(27), pp.7856–7862.

Ahn, S.J., Costa, J. & Emanuel, J.R., 1996. PicoGreen quantitation of DNA: Effective evaluation of samples pre- or post-PCR. *Nucleic Acids Research*, 24(13), pp.2623–2625.

Ahsan, M., 2015. Recent advances in the development of vaccines for tuberculosis. *Ther Adv Vaccines*, 3(3), pp.66–75.

Aínsa, J. et al., 2010. The complex *whiJ* locus mediates environmentally sensitive repression of development of *Streptomyces coelicolor* A3(2). *Antonie van Leeuwenhoek*, 98(2), pp.225–236.

Aínsa, J. a et al., 2000. WhiA, a protein of unknown function conserved among gram-positive bacteria, is essential for sporulation in *Streptomyces coelicolor* A3(2). *Journal of bacteriology*, 182(19), pp.5470–5478.

Al-Bassam, M.M. et al., 2014. Response regulator heterodimer formation controls a key stage in Streptomyces development. *PLoS genetics*, 10(8 e1004554).

Albers, S.-V., Szabó, Z. & Driessen, A.J.M., 2006. Protein secretion in the Archaea: multiple paths towards a unique cell surface. *Nature reviews. Microbiology*, 4(7), pp.537–547.

Amman, F. et al., 2014. TSSAR: TSS annotation regime for dRNA-seq data. *BMC bioinformatics*, 15(89), pp.1–11.

Anantharaman, V., Koonin, E. & Aravind, L., 2002. SPOUT: a class of methyltransferases that includes *spoU* and *trmD* RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J Mol Microbiol Biotechnol*, 4(1), pp.71–75.

Antelmann, H. & Helmann, J.D., 2011. Thiol-based redox switches and gene regulation. *Antioxidants & redox signaling*, 14(6), pp.1049–63.

Arita, Y. et al., 2006. Mitochondrial localization of catalase provides optimal protection from H_2O_2-induced cell death in lung epithelial cells. *American journal of physiology. Lung cellular and molecular physiology*, 290(5), pp.L978–L986.

Babu, M.M. et al., 2006. A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *Journal of Bacteriology*, 188(8), pp.2761–2773.

Bailey, T.L. et al., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, 37, pp.W202-8.

Banerjee, S. & Sankaran, K., 2013. First ever isolation of bacterial prolipoprotein diacylglyceryl transferase in single step from *Lactococcus lactis*. *Protein Expression and Purification*, 87(2), pp.120–128.

Barbieri, N.L. et al., 2014. FNR Regulates Expression of Important Virulence Factors Contributing to Pathogenicity of Uropathogenic *Escherichia coli*. *Infection and Immunity*, 82(12), pp.5086–5098.

Barnett, J.P. et al., 2009. The twin-arginine translocation (Tat) systems from *Bacillus subtilis* display a conserved mode of complex organization and similar substrate recognition requirements. *The FEBS journal*, 276(1), pp.232–243.

Barry, S.M. et al., 2012. Cytochrome P450–catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nature Chemical Biology*, 8(10), pp.814–816.

Barth, K.R., Isabella, V.M. & Clark, V.L., 2009. Biochemical and genomic analysis of the denitrification pathway within the genus Neisseria. *Microbiology*, 155, pp.4093–4103.

Bartolini, E. et al., 2006. Role of FNR and FNR-regulated, sugar fermentation genes in *Neisseria meningitidis* infection. *Molecular Microbiology*, 60(4), pp.963–972.

Baumgärtner, M. et al., 2007. Inactivation of Lgt allows systematic characterization of lipoproteins from *Listeria monocytogenes*. *Journal of Bacteriology*, 189(2), pp.313–324.

Beckwith, J., 2013. The Sec-dependent pathway. *Research in Microbiology*, 164(6), pp.497–504.

Beilschmidt, L.K. & Puccio, H.M., 2014. Mammalian Fe-S cluster biogenesis and its implication in disease. *Biochimie*, 100(1), pp.48–50.

Beinert, H., 2000. Iron-sulfur proteins: ancient structures, still full of surprises. *Journal of biological inorganic chemistry.* 5(1), pp.2–15.

Bell-Pedersen, D., Galloway Salvo, J.L. & Belfort, M., 1991. A transcription

terminator in the thymidylate synthase (*thyA*) structural gene of *Escherichia coli* and construction of a viable *thyA::Km(r)* deletion. *Journal of Bacteriology*, 173(3), pp.1193–1200.

Bennett, J. a. et al., 2009. Medium-dependent phenotypes of *Streptomyces coelicolor* with mutations in *ftsI* or *ftsW*. *Journal of Bacteriology*, 191(2), pp.661–664.

Bentley, S.D. et al., 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 417(6885), pp.141–147.

Berg, J., Tymocozko, J. & Stryer, L., 2007. *Biochemistry* W. Freeman, ed.,

Berg, J., Tymoczko, J. & Stryer, L., 2002. Aminoacyl-Transfer RNA Synthetases Read the Genetic Code. In *Biohemistry*. p. Section 29.2.

Berks, B.C., 1996. A common export pathway for proteins binding complex redox cofactors? *Molecular microbiology*, 22(3), pp.393–404.

Berks, B.C., Palmer, T. & Sargent, F., 2003. The Tat protein translocation pathway and its role in microbial physiology. *Advances in microbial physiology*, 47(0), pp.187–254.

Bibb, M.J., Molle, V. & Buttner, M.J., 2000. σBldN, an Extracytoplasmic Function RNA Polymerase Sigma Factor Required for Aerial Mycelium Formation in *Streptomyces coelicolor* A3(2). , 3(2), pp.4606–4616.

Bignell, D. et al., 2000. Study of the *bldG* locus suggests that an anti-anti-sigma factor and an anti-sigma factor may be involved in *Streptomyces coelicolor* antibiotic production and sporulation. *Microbiology*, 146, pp.2161–2173.

Den Blaauwen, T., Andreu, J.M. & Monasterio, O., 2014. Bacterial cell division proteins as antibiotic targets. *Bioorganic Chemistry*, 55, pp.27–38.

Blanc, B., Gerez, C. & Ollagnier de Choudens, S., 2015. Assembly of Fe/S proteins in bacterial systems. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853(6), pp.1436–1447.

Bodenmiller, D.M. & Spiro, S., 2006. The *yjeB* (*nsrR*) Gene of *Escherichia coli* Encodes a Nitric Oxide-Sensitive Transcriptional Regulator. *Journal of bacteriology*, 188(3), pp.874–881.

Bolhuis, A., 2004. The archaeal Sec-dependent protein translocation pathway. *Philosophical transactions of the Royal Society of London. Series B, Biological*

sciences, 359(1446), pp.919–27.

De Bona, P. et al., 2012. Structural studies of a signal peptide in complex with signal peptidase I cytoplasmic domain: The stabilizing effect of membrane-mimetics on the acquired fold. *Proteins: Structure, Function and Bioinformatics*, 80(3), pp.807–817.

Bonomi, F. et al., 2011. Facilitated transfer of IscU-[2Fe2S] clusters by chaperone-mediated ligand exchange. *Biochemistry*, 50(44), pp.9641–9650.

Bos, M.P. et al., 2004. Identification of an outer membrane protein required for the transport of lipopolysaccharide to the bacterial cell surface. *Proceedings of the National Academy of Sciences of the United States of America*, 101(25), pp.9417–9422.

Boyd, E.S. et al., 2015. Interplay between Oxygen and Fe−S Cluster Biogenesis: Insights from the Suf Pathway. *Biochemistry*, 53, pp.5834–5847.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry*, 72, pp.248–254.

Branchu, P. et al., 2014. NsrR, GadE, and GadX Interplay in Repressing Expression of the *Escherichia coli* O157:H7 LEE Pathogenicity Island in Response to Nitric Oxide. *PLoS Pathogens*, 10(1), p.e1003874.

Braun, V. & Rehn, K., 1969. Chemical characterization, spatial distribution and function of a lipoprotein (murein-lipoprotein) of the *E. coli* cell wall. The specific effect of trypsin on the membrane structure. *European journal of biochemistry / FEBS*, 10(3), pp.426–438.

Brekasis, D. & Paget, M.S.B., 2003. A novel sensor of NADH/NAD + redox poise in *Streptomyces coelicolor* A3 (2). *EMBO Journal*, 22(18), pp.4856–4865.

Brülle, J., Tschumi, A. & Sander, P., 2013. Lipoproteins of slow-growing Mycobacteria carry three fatty acids and are N-acylated by Apolipoprotein N-Acyltransferase BCG_2070c. *BMC Microbiology*, 13(1), p.1.

Brülle, J.K. et al., 2010. Cloning, expression and characterization of *Mycobacterium tuberculosis* lipoprotein LprF. *Biochemical and Biophysical Research Communications*, 391(1), pp.679–684.

Buddelmeijer, N., 2015. The molecular mechanism of bacterial lipoprotein modification--How, when and why? *FEMS Microbiology Reviews*, 39(2),

pp.246–261.

Buddelmeijer, N. & Young, R., 2010. The essential *Escherichia coli* apolipoprotein N-acyltransferase (Lnt) exists as an extracytoplasmic thioester acyl-enzyme intermediate. *Biochemistry*, 49(2), pp.341–346.

Burian, J. et al., 2012. The mycobacterial transcriptional regulator *whiB7* gene links redox homeostasis and intrinsic antibiotic resistance. *Journal of Biological Chemistry*, 287(1), pp.299–310.

Buttner, M.J., 2013. Genes Required for Aerial Growth , Cell Division , and Chromosome venezuelae. *mBio*, 4(5), pp.1–18.

Candelaresi, M. et al., 2013. A structural and dynamic investigation of the inhibition of catalase by nitric oxide. *Organic & biomolecular chemistry*, 11(44), pp.7778–88.

Chahal, H.K. & Outten, F.W., 2012. Separate FeS scaffold and carrier functions for SufB2C 2 and SufA during in vitro maturation of [2Fe2S] Fdx. *Journal of Inorganic Biochemistry*, 116, pp.126–134.

Chandra, G. & Chater, K.F., 2014. Developmental biology of Streptomyces from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiology Reviews*, 38(3), pp.345–379.

Chater, K. et al., 1989. The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility sigma factoer of *B. subtilis*. *Cell*, 59, pp.133–43.

Chater, K.F., 1972. A morphological and genetic mapping study of white colony mutants of *Streptomyces coelicolor*. *Journal of general microbiology*, 72(1), pp.9–28.

Chater, K.F., 1993. Genetics of Differentiion in Streptomcyes. *Annu. Rev. Microbial*, 47, pp.685–713.

Chater, K.F., 2001. Regulation of sporulation in *Streptomyces coelicolor* A3(2): a checkpoint multiplex? *Current opinion in microbiology*, 4(6), pp.667–73.

Chater, K.F. et al., 2010. The complex extracellular biology of *Streptomyces*. *FEMS microbiology reviews*, 34(2), pp.171–98.

Chater, K.F. & Chandra, G., 2006. The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiology Reviews*, 30(5),

pp.651–672.

Chater, K.F. & Chandra, G., 2006. The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS microbiology reviews*, 30(5), pp.651–72.

Chatzi, K.E. et al., 2013. Breaking on through to the other side: protein export through the bacterial Sec system. *The Biochemical journal*, 449(1), pp.25–37.

Chelikani, P. et al., 2005. Characterization of a large subunit catalase truncated by proteolytic cleavage. *Biochemistry*, 44(15), pp.5597–5605.

Chhabra, S. & Spiro, S., 2015. Inefficient translation of *nsrR* constrains behavior of the NsrR regulon in *Escherichia coli*. *Microbiology*, 161(10), pp.2029–2038.

Chimalapati, S. et al., 2012. Effects of deletion of the *Streptococcus pneumoniae* lipoprotein diacylglyceryl transferase gene Lgt on ABC transporter function and on growth in vivo. *PLoS ONE*, 7(7).

Chng, S.-S. et al., 2010. Characterization of the two-protein complex in *Escherichia coli* responsible for lipopolysaccharide assembly at the outer membrane. *Proceedings of the National Academy of Sciences of the United States of America*, 107(12), pp.5363–5368.

Choi, W.W. et al., 2009. The *whcA* gene plays a negative role in oxidative stress response of *Corynebacterium glutamicum*. *FEMS Microbiology Letters*, 290(1), pp.32–38.

Cianciotto, N.P., 2005. Type II secretion: A protein secretion system for all seasons. *Trends in Microbiology*, 13(12), pp.581–588.

Claessen, D. et al., 2003. A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in *Streptomyces coelicolor* by forming amyloid-like fibrils. *Genes and Development*, 17(14), pp.1714–1726.

Claessen, D. et al., 2006. Regulation of *Streptomyces* development: reach for the sky! *Trends in microbiology*, 14(7), pp.313–9.

Claessen, D. et al., 2004. The formation of the rodlet layer of streptomycetes is the result of the interplay between rodlins and chaplins. *Molecular Microbiology*, 53(2), pp.433–443.

Claessen, D. et al., 2002. Two novel homologous proteins of *Streptomyces coelicolor* and *Streptomyces lividans* are involved in the formation of the rodlet

layer and mediate attachment to a hydrophobic surface. *Molecular Microbiology*, 44(6), pp.1483–1492.

Cline, K., 2015. Mechanistic Aspects of Folded Protein Transport by the Twin Arginine Translocase (Tat). *Journal of Biological Chemistry*, 290(27), pp.16530–16538.

Cobb, R.E., Wang, Y. & Zhao, H., 2014. High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synthetic Biology*, 4(6), pp.723–8.

Constantinidou, C. et al., 2006. A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as *Escherichia coli* K12 adapts from aerobic to anaerobic growth. *Journal of Biological Chemistry*, 281(8), pp.4802–4815.

Córdova-Dávalos, L.E. et al., 2014. Lipoprotein N-acyl transferase (Lnt1) is dispensable for protein O-mannosylation by *Streptomyces coelicolor*. *FEMS Microbiology Letters*, 350(1), pp.72–82.

Cornelis, G.R., 2006. The type III secretion injectisome. *Nature reviews. Microbiology*, 4(11), pp.811–825.

Cornelis, P. et al., 1989. Cloning and analysis of the gene for the major outer membrane lipoprotein from *Pseudomonas aeruginosa*. *Molecular microbiology*, 3(3), pp.421–428.

Cortese, J.D. et al., 1982. Reactivation of D-beta-hydroxybutyrate dehydrogenase with short-chain lecithins: stoichiometry and kinetic mechanism. *Biochemistry*, 21(16), pp.3899–3908.

Costa, T.R.D. et al., 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology*, 13(6), pp.343–359.

Crack, J. et al., 2006. Detection of Sulfide Release from the Oxygen-sensing [4Fe-4S] Cluster of FNR. *Journal of Biological Chemistry*, 281(28), pp.18909–18913.

Crack, J. et al., 2015. NsrR from *Streptomyces coelicolor* is a Nitric Oxide-Sensing [4Fe-4S] Cluster Protein with a Specialized Regulatory Function. *Journal of Biological Chemistry*, 290(20), pp.12689–12704.

Crack, J.C. et al., 2012. Bacterial iron-sulfur regulatory proteins as biological sensor-switches. *Antioxidants & redox signaling*, 17(9), pp.1215–31.

Crack, J.C. et al., 2016. Differentiated, promoter-specific response of [4Fe-4S] NsrR DNA-binding to reaction with nitric oxide. *Journal of Biological Chemistry*, 291(16), pp.8663–8672.

Crack, J.C. et al., 2013. Mechanism of [4Fe-4S](Cys)4 cluster nitrosylation is conserved among NO-responsive regulators. *Journal of Biological Chemistry*, 288(16), pp.11492–11502.

Crack, J.C. et al., 2011. Mechanistic insight into the nitrosylation of the [4Fe-4S] cluster of WhiB-like proteins. *Journal of the American Chemical Society*, 133(4), pp.1112–21.

Crack, J.C. et al., 2007. Superoxide-mediated amplification of the oxygen-induced switch from [4Fe-4S] to [2Fe-2S] clusters in the transcriptional regulator FNR. *Proceedings of the National Academy of Sciences of the United States of America*, 104(7), pp.2092–2097.

Crack, J.C. et al., 2006. Detection of Sulfide Release from the Oxygen-sensing [4Fe-4S] Cluster of FNR. *Journal of Biological Chemistry,* 281(28) pp.18909-18913.

Cruz-ramos, H. et al., 2002. NO sensing by FNR : regulation of the *Escherichia coli* NO-detoxifying flavohaemoglobin , Hmp. *The EMBO Journal,* 21(13), pp.3235–3244.

Cupp-Vickery, J.R., Urbina, H. & Vickery, L.E., 2003. Crystal structure of IscS, a cysteine desulfurase from *Escherichia coli*. *Journal of Molecular Biology*, 330(5), pp.1049–1059.

D'Autréaux, B. et al., 2005. A non-haem iron centre in the transcription factor NorR senses nitric oxide. *Nature*, 437(7059), pp.769–772.

Dai, Y. & Outten, F.W., 2012. The *E. coli* SufS-SufE sulfur transfer system is more resistant to oxidative stress than IscS-IscU. *FEBS Letters*, 586(22), pp.4016–4022.

Das, S. et al., 2009. Contribution of lipoproteins and lipoprotein processing to endocarditis virulence in *Streptococcus sanguinis*. *Journal of Bacteriology*, 191(13), pp.4166–4179.

Denham, E.L., Ward, P.N. & Leigh, J. a., 2009. In the absence of Lgt, lipoproteins are shed from *Streptococcus uberis* independently of Lsp. *Microbiology*, 155(1), pp.134–141.

Dev, I.K., Harvey, R.J. & Ray, P.H., 1985. Inhibition of prolipoprotein signal peptidase by globomycin. *Journal of Biological Chemistry*, 260(10), pp.5891–5894.

Doi, Y. & Arakawa, Y., 2007. 16S Ribosomal RNA Methylation: Emerging Resistance Mechanism against Aminoglycosides. *Clinical Infect Dis*, 45(1), pp.88–94.

Dosanjh, M., Newton, G.L. & Davies, J., 2008. Characterization of a mycothiol ligase mutant of *Rhodococcus jostii* RHA1. *Research in Microbiology*, 159(9–10), pp.643–650.

Driessen, A.J.M. & Nouwen, N., 2008. Protein translocation across the bacterial cytoplasmic membrane. *Annual review of biochemistry*, 77, pp.643–667.

Dutta, T., Malhotra, A. & Deutscher, M.P., 2012. Exoribonuclease and endoribonuclease activities of RNase BN/RNase Z both function in vivo. *Journal of Biological Chemistry*, 287(42), pp.35747–35755.

Eccleston, J.F. et al., 2006. The kinetic mechanism of the SufC ATPase: The cleavage step is accelerated by SufB. *Journal of Biological Chemistry*, 281(13), pp.8371–8378.

Edwards, J. et al., 2012. *Neisseria meningitidis* and *Neisseria gonorrhoeae* are differently adapted in the regulation of denitrification: single nucleotide polymorphisms that enable species-specific tuning of the aerobic-anaerobic switch. *The Biochemical journal*, 445(1), pp.69–79.

Elliot, M. A., Buttner, M. J., Nodwell, J. R., 2008. Multicellular Development in Streptomyces. *Myxobacteria, Multicellularity and Differentiation*. pp. 243–267.

Emanuelsson, O. et al., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*, 2(4), pp.953–71.

Erez, E., Fass, D. & Bibi, E., 2009. How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature*, 459(7245), pp.371–378.

Even, S. et al., 2006. Global Control of Cysteine Metabolism by CymR in *Bacillus subtilis*. *Journal of Bacteriology*, 188(6), pp.2184–2197.

Feehily, C. & Karatzas, K.A.G., 2013. Role of glutamate metabolism in bacterial responses towards acid and other stresses. *Journal of Applied Microbiology*, 114(1), pp.11–24.

Fernández-Martínez, L.T. et al., 2011. A transposon insertion single-gene knockout library and new ordered cosmid library for the model organism *Streptomyces coelicolor* A3(2). *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 99(3), pp.515–522.

Fernández-Martínez, L.T. & Bibb, M.J., 2014. Use of the Meganuclease I-SceI of *Saccharomyces cerevisiae* to select for gene deletions in actinomycetes. *Scientific Reports*, 4, p.7100.

Filloux, A., Hachani, A. & Bleves, S., 2008. The bacterial type VI secretion machine: Yet another player for protein transport across membranes. *Microbiology*, 154(6), pp.1570–1583.

Fink, R.C. et al., 2007. FNR is a global regulator of virulence and anaerobic metabolism in *Salmonella enterica* serovar typhimurium (ATCC 14028s). *Journal of Bacteriology*, 189(6), pp.2262–2273.

Finn, R.D. et al., 2016. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, 44(D1), pp.D279–D285.

Flärdh, K., 2003. Essential role of DivIVA in polar growth and morphogenesis in *Streptomyces coelicolor* A3(2). *Molecular microbiology*, 49(6), pp.1523–1536.

Flärdh, K. et al., 2012. Regulation of apical growth and hyphal branching in *Streptomyces*. *Current Opinion in Microbiology*, 15(6), pp.737–743.

Flärdh, K. & Buttner, M.J., 2009. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature reviews. Microbiology*, 7(1), pp.36–49.

Fleischhacker, A.S. et al., 2012. Characterization of the [2Fe-2S] cluster of *Escherichia coli* transcription factor IscR. *Biochemistry*, 51(22), pp.4453–4462.

Fogel, N., 2015. Tuberculosis: A disease without boundaries. *Tuberculosis*, pp.5–9.

Forrester, M.T. & Foster, M.W., 2012. Protection from nitrosative stress: a central role for microbial flavohemoglobin. *Free radical biology & medicine*, 52(9), pp.1620–33.

Fowler-Goldsworthy, K. et al., 2011. The actinobacteria-specific gene *wblA* controls major developmental transitions in *Streptomyces coelicolor* A3(2). *Microbiology*, 157, pp.1312–28.

Freinkman, E., Chng, S.-S. & Kahne, D., 2011. The complex that inserts

lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6), pp.2486–2491.

Fridovich, I. & Fridovich, I., 1997. Superoxide Anion Radical (O_2^-), Superoxide Dismutases, and Related Matters*. *Molecular Biology*, pp.18515–18517.

Friedrich, T., Dekovic, D.K. & Burschel, S., 2016. Assembly of the Escherichia coli NADH:ubiquinone oxidoreductase (respiratory complex I). *Biochimica et Biophysica Acta - Bioenergetics*, 1857(3), pp.214–223.

Fröbel, J., Rose, P. & Müller, M., 2012. Twin-arginine-dependent translocation of folded proteins. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1592), pp.1029–46.

Fukuda, A. et al., 2002. Aminoacylation of the N-terminal cysteine is essential for Lol-dependent release of lipoproteins from membranes but does not depend on lipoprotein sorting signals. *The Journal of biological chemistry*, 277(45), pp.43512–8.

Fuss, J.O. et al., 2015. Emerging critical roles of Fe–S clusters in DNA replication and repair. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853(6), pp.1253–1271.

Füzéry, A.K. et al., 2008. Solution structure of the iron-sulfur cluster cochaperone HscB and its binding surface for the iron-sulfur assembly scaffold protein IscU. *Biochemistry*, 47(36), pp.9394–9404.

Gardner, P.R., 2012. Hemoglobin: a nitric-oxide dioxygenase. *Scientifica*, 2012, p.683729.

Gardner, P.R. et al., 2006. Hemoglobins dioxygenate nitric oxide with high fidelity. *Journal of Inorganic Biochemistry*, 100(4), pp.542–550.

Geiman, D.E. et al., 2006. Differential gene expression in response to exposure to antimycobacterial agents and other stress conditions among seven *Mycobacterium tuberculosis whiB*-like genes. *Antimicrobial agents and chemotherapy*, 50(8), pp.2836–41.

Gerber, N.N. & Lechevalier, H.A., 1965. Geosmin, an earthly-smelling substance isolated from actinomycetes. *Applied microbiology*, 13(6), pp.935–8.

Geukens, N. et al., 2006. The type II signal peptidase of Legionella pneumophila. *Research in microbiology*, 157(9), pp.836–41.

Gey Van Pittius, N.C. et al., 2001. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome biology*, 2(10), p.RESEARCH0044.

Giel, J.L. et al., 2013. Regulation of iron-sulphur cluster homeostasis through transcriptional control of the Isc pathway by [2Fe-2S]-IscR in *Escherichia coli*. *Molecular Microbiology*, 87(3), pp.478–492.

Gill, H.S. & Eisenberg, D., 2001. The crystal structure of phosphinothricin in the active site of glutamine synthetase illuminates the mechanism of enzymatic inhibition. *Biochemistry*, 40(7), pp.1903–1912.

Giménez, M.I., Dilks, K. & Pohlschröder, M., 2007. Haloferax volcanii twin-arginine translocation substates include secreted soluble, C-terminally anchored and lipoproteins. *Molecular Microbiology*, 66(6), pp.1597–1606.

Glazebrook, M. a et al., 1990. Sporulation of *Streptomyces venezuelae* in submerged cultures. *Journal of general microbiology*, 136(3), pp.581–8.

Goosens, V.J., Monteferrante, C.G. & Van Dijl, J.M., 2014. The Tat system of Gram-positive bacteria. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(8), pp.1698–1706.

de Greeff, A. et al., 2003. Lipoprotein signal peptidase of *Streptococcus suis* serotype 2. *Microbiology*, 149(6), pp.1399–1407.

Green, J. et al., 1996. Reconstitution of the [4Fe-4S] cluster in FNR and demonstration of the aerobic-anaerobic transcription switch in vitro. *Biohem J*, 316, pp.887–892.

Gregory, M.A., Till, R. & Smith, M.C.M., 2003. Integration Site for *Streptomyces* Phage φBT1 and Development of Site-Specific Integrating Vectors. *Society*, 185(17), pp.5320–5323.

Grimm, B., 1990. Primary structure of a key enzyme in plant tetrapyrrole synthesis: glutamate 1-semialdehyde aminotransferase. *Proceedings of the National Academy of Sciences of the United States of America*, 87(11), pp.4169–73.

Gullón, S. et al., 2015. Exploring the Feasibility of the Sec Route to Secrete Proteins Using the Tat Route in Streptomyces lividans. *Molecular Biotechnology*, 57(10), pp.931–938.

Gullón, S., Arranz, E.I.G. & Mellado, R.P., 2013. Transcriptional characterisation of the negative effect exerted by a deficiency in type II signal peptidase on

extracellular protein secretion in *Streptomyces lividans*. *Applied microbiology and biotechnology*, 97(23), pp.10069–80.

Guo, M.S. et al., 2014. MicL, a new σ^E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes and Development*, 28(14), pp.1620–1634.

Gupta, S.D. et al., 1993. Characterization of a temperature-sensitive mutant of Salmonella typhimurium defective in apolipoprotein N-acyltransferase. *Journal of Biological Chemistry*, 268(22), pp.16551–16556.

Gupta, S.D., Dowhan, W. & Wu, H.C., 1991. Phosphatidylethanolamine is not essential for the N-acylation of apolipoprotein in *Escherichia coli*. *The Journal of biological chemistry*, 266(15), pp.9983–6.

Gust, B. et al., 2006. Recombineering in *Streptomyces coelicolor*. *FEMS Online Protocols*, pp.1–22.

Gust, B., Kieser, T. & Chater, K., 2002. PCR targeting system in *Streptomyces coelicolor* A3 (2). *John Innes Centre*, 3(2), pp.1–39.

Guyet, A. et al., 2014. Identified members of the *Streptomyces lividans* AdpA regulon involved in differentiation and secondary metabolism. *BMC microbiology*, 14, p.81.

Hahn, J., Oh, S. & Roe, J., 2002. Role of OxyR as a peroxide-sensing positive regulator in *Streptomyces coelicolor* A3(2). *Journal of bacteriology*, 184(19), pp.5214–5222.

Halliwell, B. & Gutteridge, J., 2007. *Free radicals in biology and medicine* xxxvi., New York: Oxford University Press.

Hamilton, A. et al., 2006. Mutation of the maturase lipoprotein attenuates the virulence of *Streptococcus equi* to a greater extent than does loss of general lipoprotein lipidation. *Infection and Immunity*, 74(12), pp.6907–6919.

Härtig, E. & Jahn, D., 2012. Regulation of the Anaerobic Metabolism in *Bacillus subtilis*. *Advances in Microbial Physiology*, 61, pp.195–216.

Hayashi, S. et al., 1985. Modification and processing of internalized signal sequences of prolipoprotein in *Escherichia coli* and in *Bacillus subtilis*. *Journal of Biological Chemistry*, 260(9), pp.5753–5759.

Hedstrom, L., 2009. IMP Dehydrogenase: Structure, Mechanism and Inhibition.

Chemical Reviews, 109(7), pp.2903–2928.

Henares, B. et al., 2014. The ResD response regulator, through functional interaction with NsrR and Fur, plays three distinct roles in *Bacillus subtilis* transcriptional control. *Journal of Bacteriology*, 196(2), pp.493–503.

Henderson, I.R. et al., 2004. Type V Protein Secretion Pathway : the Autotransporter Story Type V Protein Secretion Pathway : the Autotransporter Story. *Microbiology and molecular biology reviews : MMBR*, 68(4), pp.692–744.

den Hengst, C. et al., 2010. Genes essential for morphological development and antibiotic production in *Streptomyces coelicolor* are targets of BldD during vegetative growth. *Mol. Microbiol*, 78, pp.361–379.

Den Hengst, C.D. et al., 2010. Genes essential for morphological development and antibiotic production in *Streptomyces coelicolor* are targets of BldD during vegetative growth. *Molecular Microbiology*, 78(2), pp.361–379.

Henneke, P. et al., 2008. Lipoproteins are critical TLR2 activating toxins in group B *Streptococcal sepsis. Journal of immunology (Baltimore, Md. : 1950)*, 180(9), pp.6149–6158.

Hibbing, M.E. & Fuqua, C., 2011. Antiparallel and interlinked control of cellular iron levels by the Irr and RirA regulators of Agrobacterium tumefaciens. *Journal of Bacteriology*, 193(14), pp.3461–3472.

Hiemstra, H. et al., 1987. Distribution of newly synthesized lipoprotein over the outer membrane and the peptidoglycan sacculus of an *Escherichia coli* lac-lpp strain. *Journal of Bacteriology*, 169(12), pp.5434–5444.

Hiemstra, H. et al., 1986. Induction kinetics and cell surface distribution of *Escherichia coli* lipoprotein under *lac* promoter control. *Journal of Bacteriology*, 168(1), pp.140–151.

Hillmann, F., Argentini, M. & Buddelmeijer, N., 2011. Kinetics and phospholipid specificity of apolipoprotein N-acyltransferase. *Journal of Biological Chemistry*, 286(32), pp.27936–27946.

Hirota, Y. et al., 1977. On the process of cellular division in *Escherichia coli*: a mutant of E. coli lacking a murein-lipoprotein. *Proceedings of the National Academy of Sciences of the United States of America*, 74(4), pp.1417–1420.

Hoff, K.G., Cupp-Vickery, J.R. & Vickery, L.E., 2003. Contributions of the LPPVK motif of the iron-sulfur template protein IscU to interactions with the Hsc66-

Hsc20 chaperone system. *Journal of Biological Chemistry*, 278(39), pp.37582–37589.

Hojati, Z. et al., 2002. Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chemistry and Biology*, 9(11), pp.1175–1187.

Hopkins, A., Buchanan, G. & Palmer, T., 2014. Role of the twin arginine protein transport pathway in the assembly of the *Streptomyces coelicolor* cytochrome bc1 complex. *Journal of Bacteriology*, 196(1), pp.50–59.

Hopwood, D.A., 2007. *Streptomyces in Nature and Medicine*,

Houben, E.N.G., Korotkov, K. V. & Bitter, W., 2014. Take five - Type VII secretion systems of Mycobacteria. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(8), pp.1707–1716.

Hsu, T. et al., 2003. The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proceedings of the National Academy of Sciences of the United States of America*, 100(21), pp.12420–12425.

Hu, J. et al., 2011. The crystal structure of GXGD membrane protease FlaK. *Nature*, 475(7357), pp.528–531.

Huang, Y.T. et al., 2009. In vitro characterization of enzymes involved in the synthesis of nonproteinogenic residue (2S,3S)-β-methylphenylalanine in glycopeptide antibiotic mannopeptimycin. *ChemBioChem*, 10(15), pp.2480–2487.

Hullo, M.F., Martin-Verstraete, I. & Soutourina, O., 2010. Complex phenotypes of a mutant inactivated for CymR, the global regulator of cysteine metabolism in *Bacillus subtilis*. *FEMS Microbiology Letters*, 309(2), pp.201–207.

Hutchings, M.I. et al., 2009. Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em. *Trends in microbiology*, 17(1), pp.13–21.

Imam, S., Noguera, D.R. & Donohue, T.J., 2014. Global Analysis of Photosynthesis Transcriptional Regulatory Networks. *PLoS Genetics*, 10(12), p.e1004837.

Imlay, J. a, 2008. Cellular defenses against superoxide and hydrogen peroxide. *Annu Rev Biochem*, (77), pp.755–776.

Imlay, J. a, 2015. Diagnosing oxidative stress in bacteria: not as easy as you might think. *Current Opinion in Microbiology*, 24, pp.124–131.

Imlay, J. a, 2003. Pathways of oxidative damage. *Annual review of microbiology*, 57, pp.395–418.

Ingolia, N.T. et al., 2009. Genome-wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*, 324(5924), pp.218–223.

Ingolia, N.T., 2014. Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Reviews Genetics*, 15(3), pp.205–213.

Inouye, M., Shaw, J. & Shen, C., 1972. The assembly of a structural lipoprotein in the envelope of Escherichia coli. *Journal of Biological Chemistry*, 247(24), pp.8154–8159.

Inukai M., Nakajima M., Osawa M., H.T. and A.M., 1978. Globomycin, a new peptide antibiotic with spheroplast-formin activity. *The Journal of antibiotic*, XXXI(5), pp.1–5.

Isabella, V.M. et al., 2009. Functional analysis of NsrR, a nitric oxide-sensing Rrf2 repressor in *Neisseria gonorrhoeae*. *Molecular microbiology*, 71(1), pp.227–39.

Isobe, K. et al., 2006. Production of catalase by fungi growing at low pH and high temperature. *Journal of bioscience and bioengineering*, 101(1), pp.73–76.

Jackowski, S. & Rock, C.O., 1986. Transfer of fatty acids from the 1-position of phosphatidylethanolamine to the major outer membrane lipoprotein of *Escherichia coli*. *Journal of Biological Chemistry*, 261(24), pp.11328–11333.

Jakimowicz, D. et al., 2006. Developmental control of a *parAB* promter leads to formation of sporulation associated ParB complexes in *Streptomyces coelicolor*. *J Bacteriol*, 188, pp.1710–1720.

Jeong, Y. et al., 2016. The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). *Nature Communications*, 7(11605), pp.1–11.

Ji, Q. et al., 2012. *Staphylococcus aureus* CymR is a new thiol-based oxidation-sensing regulator of stress resistance and oxidative response. *Journal of Biological Chemistry*, 287(25), pp.21102–21109.

Jiang, J., He, X. & Cane, D., 2007. Biosynthesis of the earthy odorant geosmin by a

bifunctional *Streptomyces coelicolor* enzyme. *Nat Chem Biol*, 3(11), pp.711–715.

Jiang, P., Mayo, A.E. & Ninfa, A.J., 2007. *Escherichia coli* glutamine synthetase adenylyltransferase (ATase, EC 2.7.7.49): Kinetic characterization of regulation by PII, PII-UMP, glutamine, and α-ketoglutarate. *Biochemistry*, 46(13), pp.4133–4146.

Jin, H.K. et al., 2009. Structure and Dynamics of the Iron-Sulfur Cluster Assembly Scaffold Protein IscU and its Interaction with the Cochaperone HscB. *Biochemistry*, 48(26), pp.6062–6071.

Johnson, D.C. et al., 2005. Structure, function, and formation of biological iron-sulfur clusters. *Annual review of biochemistry*, 74, pp.247–281.

Johnson, E.G. et al., 2008. Plant-pathogenic *Streptomyces* species produce nitric oxide synthase-derived nitric oxide in response to host signals. *Chemistry & biology*, 15(1), pp.43–50.

Johnston, A.W.B. et al., 2007. Living without Fur: The subtlety and complexity of iron-responsive gene regulation in the symbiotic bacterium *Rhizobium* and other α-proteobacteria. *BioMetals*, 20(3–4), pp.501–511.

Jongbloed, J.D.H. et al., 2004. Two minimal Tat translocases in *Bacillus*. *Molecular microbiology*, 54(5), pp.1319–25.

Joshi, M. V et al., 2010. The twin arginine protein transport pathway exports multiple virulence proteins in the plant pathogen *Streptomyces scabies*. *Molecular microbiology*, 77(1), pp.252–71.

Kaiser, B.K. & Stoddard, B.L., 2011. DNA recognition and transcriptional regulation by the WhiA sporulation factor. *Scientific reports*, 1, p.156.

Kalb, D., Gressler, J. & Hoffmeister, D., 2016. Active-Site Engineering Expands the Substrate Profile of the Basidiomycete $_L$-Tryptophan Decarboxylase CsTDC. *ChemBioChem*, 17(2), pp.132–136.

Kang, J.Y. et al., 2009. Recognition of Lipopeptide Patterns by Toll-like Receptor 2-Toll-like Receptor 6 Heterodimer. *Immunity*, 31(6), pp.873–884.

Karlinsey, J.E. et al., 2012. The NsrR regulon in nitrosative stress resistance of *Salmonella enterica* serovar Typhimurium. *Molecular microbiology*, 85(6), pp.1179–93.

De Keersmaeker, S. et al., 2007. The Tat pathway in *Streptomyces lividans*: interaction of Tat subunits and their role in translocation. *Microbiology (Reading, England)*, 153(Pt 4), pp.1087–94.

Keon, R.G., Fu, R. & Voordouw, G., 1997. Deletion of two downstream genes alters expression of the *hmc* operon of *Desulfovibrio vulgaris* subsp. *vulgaris* Hildenborough. *Archives of Microbiology*, 167(6), pp.376–383.

van Keulen, G. et al., 2007. The obligate aerobic actinomycete *Streptomyces coelicolor* A3(2) survives extended periods of anaerobic stress. *Environmental microbiology*, 9(12), pp.3143–9.

Khoroshilova, N. et al., 1997. Iron-sulfur cluster disassembly in the FNR protein of *Escherichia coli* by O2: [4Fe-4S] to [2Fe-2S] conversion with loss of biological activity. *Proceedings of the National Academy of Sciences of the United States of America*, 94(12), pp.6087–6092.

Kiley, P.J. & Beinert, H., 1998. Oxygen sensing by the global regulator, FNR: the role of the iron- sulfur cluster. *FEMS Microbiol Rev*, 22(5), p.341–52.

Kim, H.M. et al., 2014. Inverse regulation of Fe- and Ni-containing SOD genes by a fur family regulator Nur through small RNA processed from 3'UTR of the *sodF* mRNA. *Nucleic Acids Research*, 42(3), pp.2003–2014.

Kim, J.H. et al., 2012. Specialized Hsp70 chaperone (HscA) binds preferentially to the disordered form, whereas J-protein (HscB) binds preferentially to the structured form of the iron-sulfur cluster scaffold protein (IscU). *Journal of Biological Chemistry*, 287(37), pp.31406–31413.

Kimura, S. & Suzuki, T., 2014. Iron–sulfur proteins responsible for RNA modifications. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853, pp.1272–1283.

Kinshita, S., 2005. *A Short History of the Birth of the Amino Acid Industry in Japan. In: Handbook of Corynebacterium glutamicum*

Kirby, R. et al., 2012. Draft genome sequence of the human pathogen *Streptomyces somaliensis*, a significant cause of actinomycetoma. *Journal of Bacteriology*, 194(13), pp.3544–3545.

Kodani, S. et al., 2004. The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene *ramS* in *Streptomyces coelicolor*. *Proceedings of the National Academy of Sciences of the United States of*

America, 101(31), pp.11448–11453.

Kol, S., Nouwen, N. & Driessen, A.J.M., 2008. Mechanisms of YidC-mediated insertion and assembly of multimeric membrane protein complexes. *The Journal of biological chemistry*, 283(46), pp.31269–73.

Komar, J. et al., 2015. ACEMBLing a Multiprotein Transmembrane Complex: The Functional SecYEQ-SecDF-YajC-YidC Holotranslocon Protein Secretase/Insertase. *Methods in Enzymology*, 556, pp.23-49.

Kommineni, S. et al., 2012. Global transcriptional control by NsrR in *Bacillus subtilis*. *Journal of bacteriology*, 194(7), pp.1679–88.

Kommineni, S. et al., 2010. Nitric oxide sensitive and insensitive interaction of *Bacillus subtilis* NsrR with a ResDE-controlled promoter. *Mol Microbial*, 78(5), pp.1280–1293.

Kommineni, S. et al., 2013. The direct synergistic roles of the NO-sensitive NsrR repressor and ResD response regulator in the transcriptional control in *Bacillus subtilis*. *Molecular Microbiology*.

Kouidmi, I., Levesque, R.C. & Paradis-Bleau, C., 2014. The biology of Mur ligases as an antibacterial target. *Molecular Microbiology*, 94(2), pp.242–253.

Kudva, R. et al., 2013. Protein translocation across the inner membrane of Gram-negative bacteria: The Sec and Tat dependent protein transport pathways. *Research in Microbiology*, 164(6), pp.505–534.

Kurokawa, K. et al., 2012. Environment-mediated accumulation of diacyl lipoproteins over their triacyl counterparts in *Staphylococcus aureus*. *Journal of bacteriology*, 194(13), pp.3299–306.

Kurokawa, K. et al., 2009. The Triacylated ATP Binding Cluster Transporter Substrate-binding Lipoprotein of *Staphylococcus aureus* Functions as a Native Ligand for Toll-like Receptor 2. *The Journal of biological chemistry*, 284(13), pp.8406–11.

Laemmli, U., 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227, pp.680–685.

Lai, J.S., Philbrick, W.M. & Wu, H.C., 1980. Acyl moieties in phospholipids are the precursors for the fatty acids in murein lipoprotein of *Escherichia coli*. *Journal of Biological Chemistry*, 255(11), pp.5384–5387.

Lambden, P. & Guest, J., 1976. Mutants of *Escherichia coli* K12 unable to use fumarate as an anaerobic electron acceptor. *J Gen Microbial*, 97(2), pp.145–60.

Lange, H. et al., 2000. A mitochondrial ferredoxin is essential for biogenesis of cellular iron-sulfur proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 97(3), pp.1050–1055.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, pp.357–359.

Layer, G. et al., 2007. SufE transfers sulfur from SufS to SufB for iron-sulfur cluster assembly. *Journal of Biological Chemistry*, 282(18), pp.13342–13350.

Lee, C. et al., 2004. Redox regulation of OxyR requires specific disulfide bond formation involving a rapid kinetic reaction path. *Nature structural & molecular biology*, 11(12), pp.1179–1185.

Lee, J.-W., Soonsanga, S. & Helmann, J.D., 2007. A complex thiolate switch regulates the *Bacillus subtilis* organic peroxide sensor OhrR. *Proceedings of the National Academy of Sciences of the United States of America*, 104(21), pp.8743–8748.

Lee, K.-L. et al., 2015. Factors affecting redox potential and differential sensitivity of SoxR to redox-active compounds. *Molecular Microbiology*.

Leskelä, S. et al., 1999. Lipid modification of prelipoproteins is dispensable for growth but essential for efficient protein secretion in *Bacillus subtilis*: characterization of the Lgt gene. *Molecular microbiology*, 31(4), pp.1075–85.

Lewis, J., Jergic, S. & Dixon, N., 2016. The E. coli DNA Replication Fork. *Enzymes*, 39, pp.31–88.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, pp.2078–2079.

de Lima Procópio, R.E. et al., 2012. Antibiotics produced by *Streptomyces*. *Brazilian Journal of Infectious Diseases*, 16(5), pp.466–471.

Liu, Y.B. et al., 2013. Physiological roles of mycothiol in detoxification and tolerance to multiple poisonous chemicals in *Corynebacterium glutamicum*. *Archives of Microbiology*, 195(6), pp.419–429.

Machata, S. et al., 2008. Lipoproteins of *Listeria monocytogenes* are critical for

virulence and TLR2-mediated immune activation. *Journal of immunology (Baltimore, Md. : 1950)*, 181(3), pp.2028–2035.

Madigan, M. & Martinko, J., 2006. *Person-To-Person Microbial Diseases In: Brock, Biology of Microorganisms.*,

Magnet, S. et al., 2008. Identification of the L,D-transpeptidases for peptidoglycan cross-linking in *Escherichia coli*. *Journal of Bacteriology*, 190(13), pp.4782–4785.

Manjunath, K., Jeyakanthan, J. & Sekar, K., 2015. Catalytic pathway, substrate binding and stability in SAICAR synthetase: A structure and molecular dynamics study. *Journal of Structural Biology*, 191(1), pp.22–31.

Marteyn, B.S., Gazi, A.D. & Sansonetti, P.J., 2012. *Shigella*: A model of virulence regulation in vivo. *Gut Microbes*, 3(2), pp.104–120.

Mashalidis, E.H. et al., 2011. Rv2607 from *Mycobacterium tuberculosis* is a pyridoxine 5'-phosphate oxidase with unusual substrate specificity. *PLoS ONE*, 6(11).

Masi, M. & Wandersman, C., 2010. Multiple signals direct the assembly and function of a type 1 secretion system. *Journal of Bacteriology*, 192(15), pp.3861–3869.

Matos, C.F.R.O., Di Cola, A. & Robinson, C., 2009. TatD is a central component of a Tat translocon-initiated quality control system for exported FeS proteins in *Escherichia coli*. *EMBO reports*, 10(5), pp.474–9.

Maurer, C. et al., 2010. TatB functions as an oligomeric binding site for folded Tat precursor proteins. *Molecular biology of the cell*, 21(23), pp.4151–4161.

McCormick, J. & Flardh, K., 2012. Signals and Regulators That Govern *Streptomyces* Development. , 36(1), pp.206–231.

McCormick, J.R., 2009. Cell division is dispensable but not irrelevant in *Streptomyces*. *Current Opinion in Microbiology*, 12(6), pp.689–698.

McCormick, J.R. & Losick, R., 1996. Cell division gene *ftsQ* is required for efficient sporulation but not growth and viability in Cell Division Gene *ftsQ* Is Required for Efficient Sporulation but Not Growth and Viability in *Streptomyces coelicolor* A3 (2). , 3(2), pp.5295–5301.

McLaughlin, K.J. et al., 2010. Structural Basis for NADH/NAD+ Redox Sensing by a

Rex Family Repressor. *Molecular Cell*, 38(4), pp.563–575.

McLean, S., Bowman, L. a H. & Poole, R.K., 2010. Peroxynitrite stress is exacerbated by flavohaemoglobin-derived oxidative stress in *Salmonella typhimurium* and is relieved by nitric oxide. *Microbiology (Reading, England)*, 156(Pt 12), pp.3556–65.

Membrillo-Hernández, J. et al., 1998. A novel mechanism for upregulation of the *Escherichia coil* K-12 hmp (flavohaemoglobin) gene by the "NO releaser", S-nitrosoglutathione: Nitrosation of homocysteine and modulation of MetR binding to the *glyA-hmp* intergenic region. *Molecular Microbiology*, 29(4), pp.1101–1112.

Messner, K.R. & Imlay, J. a., 2002. Mechanism of superoxide and hydrogen peroxide formation by fumarate reductase, succinate dehydrogenase, and aspartate oxidase. *Journal of Biological Chemistry*, 277(45), pp.42563–42571.

Mettert, E.L. & Kiley, P.J., 2014. Fe–S proteins that regulate gene expression. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853(6), pp.1284–1293.

Meyer, J., 2008. Iron-sulfur protein folds, iron-sulfur chemistry, and evolution. *Journal of Biological Inorganic Chemistry*, 13(2), pp.157–170.

Midorikawa, T. et al., 2009. An Rrf2-type transcriptional regulator is required for expression of *psaAB* genes in the *Cyanobacterium synechocystis* sp. PCC 6803. *Plant physiology*, 151(2), pp.882–892.

Midorikawa, T., Narikawa, R. & Ikeuchi, M., 2012. A deletion mutation in the spacing within the *psaA* core promoter enhances transcription in a *Cyanobacterium synechocystis* sp. PCC 6803. *Plant and Cell Physiology*, 53(1), pp.164–172.

Miggiano, R. et al., 2016. Crystal structure of *Mycobacterium tuberculosis* O6-methylguanine-DNA methyltransferase protein clusters assembled on to damaged DNA. *Biochemical Journal*, 473(2), pp.123–133.

Miller, A.F., 2004. Superoxide dismutases: Active sites that save, but a protein that kills. *Current Opinion in Chemical Biology*, 8(2), pp.162–168.

Miller, A.F., 2012. Superoxide dismutases: Ancient enzymes and new insights. *FEBS Letters*, 586(5), pp.585–595.

Miller, H.K. & Auerbuch, V., 2015. *Bacterial iron-sulfur cluster sensors in mammalian pathogens.*, Royal Society of Chemistry.

Minghetti, K. & Gennis, R., 1988. The two terminal oxidases of the aerobic respiratory chain of *Escherichia coli* each yield water and not peroxide as a final product. *Biochem Biophys Res Commun*, 155(1), pp.243–8.

Mishra, S. & Imlay, J., 2012. Why do bacteria use so many enzymes to scavenge hydrogen peroxide? *Archives of Biochemistry and Biophysics*, 525(2), pp.145–160.

Mizuno, T. & Kageyama, M., 1979. Isolation and characterization of major outer membrane proteins of *Pseudomonas aeruginosa* strain PAO with special reference to peptidoglycan-associated protein. *Journal of biochemistry*, 86(4), pp.979–989.

Mohiman, N. et al., 2012. The ppm Operon Is Essential for Acylation and Glycosylation of Lipoproteins in *Corynebacterium glutamicum*. *PLoS ONE*, 7(9).

Morris, R.P. et al., 2005. Ancestral antibiotic resistance in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, 102(34), pp.12200–5.

Mortenson, L., Valentine, R. & Carnahan, J., 1962. An electron transport factor from *Clostridium pasteurianum*. *Biochem Biophys Res Commun*, 4(7), pp.448–52.

Mühlig, A. et al., 2014. Stress Response of *Salmonella enterica* Serovar *typhimurium* to Acidified Nitrite. *Applied and environmental microbiology*, 80(20), pp.6373–82.

Münch, R. et al., 2005. Virtual Footprint and PRODORIC: An integrative framework for regulon prediction in prokaryotes. *Bioinformatics*, 21(22), pp.4187–4189.

Muñiz, M., Morsomme, P. & Riezman, H., 2001. Protein sorting upon exit from the endoplasmic reticulum. *Cell*, 104(2), pp.313–20.

Munnoch, J.T., et al., 2016. Characterization of a putative NsrR homologue in *Streptomyces venezuelae* reveals a new member of the Rrf2 superfamily. *Nature Publishing Group*, (July), pp.1–14.

Munnoch, J.T., et al., 2016. Cosmid based mutagenesis causes genetic instability in *Streptomyces coelicolor*, as shown by targeting of the lipoprotein signal peptidase gene. *Scientific Reports*, 6(29495), pp.1–10.

Muñoa, F.J. et al., 1991. Membrane topology of *Escherichia coli* prolipoprotein signal peptidase (signal peptidase II). *Journal of Biological Chemistry*, 266(26),

pp.17667–17672.

Nakamura, A. et al., 2007. Crystal structure of TTHA1657 (AT-rich DNA-binding protein; p25) from Thermus ther- mophilus HB8 at 2.16 A resolution. *Proteins*, 66, pp.755–759.

Narita, S.I. & Tokuda, H., 2011. Overexpression of LolCDE allows deletion of the *Escherichia coli* gene encoding apolipoprotein N-acyltransferase. *Journal of Bacteriology*, 193(18), pp.4832–4840.

Naseer, N., Shapiro, J. a. & Chander, M., 2014. RNA-Seq Analysis Reveals a Six-Gene SoxR Regulon in *Streptomyces coelicolor*. *PLoS ONE*, 9(8), p.e106181.

Natale, P., Brüser, T. & Driessen, A.J.M., 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms. *Biochimica et biophysica acta*, 1778(9), pp.1735–56.

Nesbit, a. D. et al., 2009. Sequence-Specific Binding to a Subset of IscR-Regulated Promoters Does Not Require IscR Fe-S Cluster Ligation. *Journal of Molecular Biology*, 387(1), pp.28–41.

Newton, G.L. & Fahey, R.C., 2008. Regulation of mycothiol metabolism by sigma(R) and the thiol redox sensor anti-sigma factor RsrA. *Molecular microbiology*, 68(4), pp.805–9.

Nicol, J.W. et al., 2009. The Integrated Genome Browser: Free software for distribution and exploration of genome-scale datasets. *Bioinformatics*, 25, pp.2730–2731.

Nodwell, J.R. & Losick, R., 1998. Purification of an extracellular signaling molecule involved in production of aerial mycelium by *Streptomyces coelicolor*. *Journal of Bacteriology*, 180(5), pp.1334–1337.

Nodwell, J.R., McGovern, K. & Losick, R., 1996. An oligopeptide permease responsible for the import of an extracellular signal governing aerial mycelium formation in *Streptomyces coelicolor*. *Molecular microbiology*, 22(5), pp.881–893.

Novichkov, P.S. et al., 2009. RegPrecise: A database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Research*, 38(SUPPL.1), pp.111–118.

Okuda, S. & Tokuda, H., 2011. Lipoprotein sorting in bacteria. *Annual review of microbiology*, 65, pp.239–59.

Okuda, S. & Tokuda, H., 2009. Model of mouth-to-mouth transfer of bacterial lipoproteins through inner membrane LolC, periplasmic LolA, and outer membrane LolB. *Proceedings of the National Academy of Sciences of the United States of America*, 106(14), pp.5877–5882.

Okugawa, S. et al., 2012. Lipoprotein biosynthesis by prolipoprotein diacylglyceryl transferase is required for efficient spore germination and full virulence of Bacillus anthracis. *Molecular Microbiology*, 83(1), pp.96–109.

Osborne, A.R., Rapoport, T. a & van den Berg, B., 2005. Protein translocation by the Sec61/SecY channel. *Annual review of cell and developmental biology*, 21, pp.529–550.

Pacher, P., Beckman, J. & Liaudet, L., 2007. Nitric Oxide and Peroxynitrite: in Health and disease. *Physiological reviews*, 87(1), pp.315–424.

Paetzel, M., 2014. Structure and mechanism of *Escherichia coli* type I signal peptidase. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(8), pp.1497–1508.

Pagels, M. et al., 2010. Redox sensing by a Rex-family repressor is involved in the regulation of anaerobic gene expression in *Staphylococcus aureus*. *Molecular Microbiology*, 76(5), pp.1142–1161.

Pailler, J. et al., 2012. Phosphatidylglycerol::Prolipoprotein diacylglyceryl transferase (Lgt) of *Escherichia coli* has seven transmembrane segments, and its essential residues are embedded in the membrane. *Journal of Bacteriology*, 194(9), pp.2142–2151.

Paitan, Y. et al., 1999. A nonessential signal peptidase II (Lsp) of *Myxococcus xanthus* might be involved in biosynthesis of the polyketide antibiotic TA. *Journal of bacteriology*, 181(18), pp.5644–51.

Palmer, T. & Berks, B.C., 2012. The twin-arginine translocation (Tat) protein export pathway. *Nature Reviews Microbiology*, 10(7), pp.483–496.

Paradis-bleau, C. et al., 2011. Lipoprotein cofactors located in the outer membrane activate bacterial cell wall polymerases. *Cell*, 143(7), pp.1110–1120.

Partridge, J.D. et al., 2009. NsrR targets in the *Escherichia coli* genome: New insights into DNA sequence requirements for binding and a role for NsrR in the regulation of motility. *Molecular Microbiology*, 73(4), pp.680–694.

Perry, J.J.P. et al., 2010. The structural biochemistry of the superoxide dismutases.

Biochimica et Biophysica Acta - Proteins and Proteomics, 1804(2), pp.245–262.

Petit, C.M. et al., 2001. Lipid modification of prelipoproteins is dispensable for growth in vitro but essential for virulence in *Streptococcus pneumoniae. FEMS Microbiology Letters*, 200(2), pp.229–233.

Petrovic, A. et al., 2008. Hydrodynamic characterization of the SufBC and SufCD complexes and their interaction with fluorescent adenosine nucleotides. *Protein science : a publication of the Protein Society*, 17(7), pp.1264–1274.

du Plessis, D.J.F., Nouwen, N. & Driessen, A.J.M., 2011. The Sec translocase. *Biochimica et biophysica acta*, 1808(3), pp.851–65.

Pogliano, J. a & Beckwith, J., 1994. SecD and SecF facilitate protein export in *Escherichia coli. The EMBO journal*, 13(3), pp.554–61.

Poole, R.K. & Hughes, M.N., 2000. New functions for the ancient globin family: Bacterial responses to nitric oxide and nitrosative stress. *Molecular Microbiology*, 36(4), pp.775–783.

Prágai, Z. et al., 1997. The signal peptidase II (Isp) gene of *Bacillus subtilis*. *Microbiology*, 143 (Pt 4(143), pp.1327–1333.

Prischi, F. et al., 2010. Structural bases for the interaction of frataxin with the central components of iron-sulphur cluster assembly. *Nature communications*, 1(7), p.95.

Pullan, S.T. et al., 2011. Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. *BMC genomics*, 12(1), p.175.

Pullan, S.T. et al., 2007. Nitric oxide in chemostat-cultured *Escherichia coli* is sensed by Fnr and other global regulators: Unaltered methionine biosynthesis indicates lack of S nitrosation. *Journal of Bacteriology*, 189(5), pp.1845–1855.

Py, B. & Barras, F., 2015. Genetic approaches of the Fe–S cluster biogenesis process in bacteria: Historical account, methodological aspects and future challenges. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853(6), pp.1429–1435.

Pym, A.S. et al., 2003. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nature medicine*, 9(5), pp.533–539.

Qi, H.Y. et al., 1995. Structure-function relationship of bacterial prolipoprotein diacylglyceryl transferase: functionally significant conserved regions. *Journal of bacteriology*, 177(23), pp.6820–4.

Quiblier, C. et al., 2013. Secretome Analysis Defines the Major Role of SecDF in *Staphylococcus aureus* Virulence. *PLoS ONE*, 8(5), pp.1–12.

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

Rampini, S.K. et al., 2008. LspA inactivation in Mycob*acterium tuberculosis* results in attenuation without affecting phagosome maturation arrest. *Microbiology*, 154(10), pp.2991–3001.

Randall, L.L. et al., 1998. Calorimetric analyses of the interaction between SecB and its ligands. *Protein science : a publication of the Protein Society*, 7(5), pp.1195–1200.

Raulfs, E.C. et al., 2008. In vivo iron-sulfur cluster formation. *Proceedings of the National Academy of Sciences of the United States of America*, 105(25), pp.8591–8596.

Ravcheev, D. a. et al., 2012. Transcriptional regulation of central carbon and energy metabolism in bacteria by redox-responsive repressor rex. *Journal of Bacteriology*, 194(5), pp.1145–1157.

Rawat, M. et al., 2002. Mycothiol-deficient *Mycobacterium smegmatis* mutants are hypersensitive to alkylating agents, free radicals, and antibiotics. *Antimicrobial Agents and Chemotherapy*, 46(11), pp.3348–3355.

Redenbach, M. et al., 1996. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Molecular microbiology*, 21(1), pp.77–96.

Remans, K. et al., 2010. Hydrophobic Surface Patches on LolA of *Pseudomonas aeruginosa* are Essential for Lipoprotein Binding. *Journal of Molecular Biology*, 401(5), pp.921–930.

Rico, S. et al., 2014. Deciphering the regulon of *Streptomyces coelicolor* AbrC3, a positive response regulator of antibiotic production. *Applied and Environmental Microbiology*, 80(8), pp.2417–2428.

Robichon, C., Vidal-Ingigliardi, D. & Pugsley, A.P., 2005. Depletion of apolipoprotein N-acyltransferase causes mislocalization of outer membrane lipoproteins in

Escherichia coli. *Journal of Biological Chemistry*, 280(2), pp.974–983.

Robinson, C. & Bolhuis, A., 2004. Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochimica et biophysica acta*, 1694(1–3), pp.135–47.

Roche, B. et al., 2013. Iron/sulfur proteins biogenesis in prokaryotes: Formation, regulation and diversity. *Biochimica et Biophysica Acta - Bioenergetics*, 1827(8–9), pp.923–937.

Rodionov, D. a. et al., 2005. Dissimilatory metabolism of nitrogen oxides in bacteria: Comparative reconstruction of transcriptional networks. *PLoS Computational Biology*, 1(5), pp.0415–0431.

Rogers, S.D. et al., 1991. Cloning and characterization of *cutE*, a gene involved in copper transport in *Escherichia coli*. *Journal of bacteriology*, 173(21), pp.6742–6748.

Rollenhagen, C. & Bumann, D., 2006. *Salmonella enterica* Highly Expressed Genes Are Disease Specific. *American Society for Microbiology*, 74(3), pp.1649–1660.

Ronald, A., 2003. The etiology of urinary tract infection: traditional and emerging pathogens. *Dis Mon*, 49(2), pp.71–82.

Ruiz, N., Kahne, D. & Silhavy, T.J., 2009. Transport of lipopolysaccharide across the cell envelope: the long road of discovery. *Nature reviews. Microbiology*, 7(9), pp.677–683.

Rybniker, J. et al., 2010. Insights into the function of the WhiB-like protein of mycobacteriophage TM4 - A transcriptional inhibitor of WhiB2. *Molecular Microbiology*, 77(June), pp.642–657.

Ryding, N.J. et al., 1998. A developmentally regulated gene encoding a repressor-like protein is essential for sporulation in *Streptomyces coelicolor* A3(2). *Molecular microbiology*, 29(1), pp.343–357.

Sahoo, R. et al., 2009. A novel role of catalase in detoxification of peroxynitrite in *S. cerevisiae*. *Biochemical and Biophysical Research Communications*, 385(4), pp.507–511.

Saini, A. et al., 2010. SufD and SufC ATPase activity are required for iron acquisition during in vivo Fe-S cluster formation on SufB. *Biochemistry*, 49(43), pp.9402–9412.

Saini, V., Farhana, A. & Steyn, A.J.C., 2012. WhiB3: A Novel Iron–Sulfur Cluster

Protein That Regulates Redox Homeostasis and Virulence. *Antioxidants & Redox Signaling*, 16(7), pp.687–697.

Salerno, P. et al., 2009. One of the two genes encoding nucleoid-associated HU proteins in *Streptomyces coelicolor* is developmentally regulated and specifically involved in spore maturation. *Journal of bacteriology*, 191(21), pp.6489–500.

Sampson, T.R. et al., 2013. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, 497(7448), pp.254–7.

Sander, P. et al., 2004. Lipoprotein processing is required for virulence of *Mycobacterium tuberculosis*. *Molecular Microbiology*, 52(6), pp.1543–1552.

Sankaran, K. et al., 1997. Roles of histidine-103 and tyrosine-235 in the function of the prolipoprotein diacylglyceryl transferase of *Escherichia coli*. *Journal of bacteriology*, 179(9), pp.2944–8.

Sankaran, K. & Wu, H.C., 1994. Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *Journal of Biological Chemistry*, 269(31), pp.19701–19706.

Santos, J. a., Pereira, P.J.B. & Macedo-Ribeiro, S., 2015. What a difference a cluster makes: The multifaceted roles of IscR in gene regulation and DNA recognition. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, pp.1–12.

Sassetti, C.M., Boyd, D.H. & Rubin, E.J., 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology*, 48(1), pp.77–84.

Schäberle, T.F., Orland, A. & König, G.M., 2014. Enhanced production of undecylprodigiosin in *Streptomyces coelicolor* by co-cultivation with the corallopyronin A-producing myxobacterium, *Corallococcus coralloides*. *Biotechnology Letters*, 36(3), pp.641–648.

Schaerlaekens, K. et al., 2004. The importance of the Tat-dependent protein secretion pathway in *Streptomyces* as revealed by phenotypic changes in tat deletion mutants and genome analysis. *Microbiology*, 150(1), pp.21–31.

Schreur, P.J.W. et al., 2011. Lgt processing is an essential step in *Streptococcus suis* lipoprotein mediated innate immune activation. *PLoS ONE*, 6(7).

Schroeder, G.N. & Hilbi, H., 2008. Molecular pathogenesis of *Shigella* spp.:

Controlling host cell signaling, invasion, and death by type III secretion. *Clinical Microbiology Reviews*, 21(1), pp.134–156.

Schwartz, C.J. et al., 2001. IscR, an Fe-S cluster-containing transcription factor, represses expression of *Escherichia coli* genes encoding Fe-S cluster assembly proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 98(26), pp.14895–14900.

Selbach, B.P., Pradhan, P.K. & Dos Santos, P.C., 2013. Protected sulfur transfer reactions by the *Escherichia coli* Suf system. *Biochemistry*, 52(23), pp.4089–4096.

Sevcikova, B. et al., 2010. The anti-anti-sigma factor BldG is involved in activation of the stress response sigma factor σ^H in *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*, 192(21), pp.5674–5681.

Sharma, C.M. et al., 2010. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature*, 464(7286), pp.250–255.

Shen, G. et al., 2007. SufR coordinates two [4Fe-4S]2+,1+ clusters and functions as a transcriptional repressor of the *sufBCDS* operon and an autoregulator of *sufR* in *Cyanobacteria*. *Journal of Biological Chemistry*, 282(44), pp.31909–31919.

Shepard, W. et al., 2011. Insights into the Rrf2 repressor family the structure of CymR, the global cysteine regulator of *Bacillus subtilis*. *The FEBS journal*, 278(15), pp.2689–701.

Shi, R. et al., 2010. Structural basis for Fe-S cluster assembly and tRNA thiolation mediated by IscS protein-protein interactions. *PLoS Biology*, 8(4).

Shin, J.H. et al., 2011. Activation of the SoxR regulon in *Streptomyces coelicolor* by the extracellular form of the pigmented antibiotic actinorhodin. *Journal of Bacteriology*, 193(1), pp.75–81.

Sickmier, E.A. et al., 2005. X-ray structure of a Rex-family repressor/NADH complex insights into the mechanism of redox sensing. *Structure*, 13(1), pp.43–54.

Sievers, F. et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), p.539.

Singh, A. et al., 2009. *Mycobacterium tuberculosis* WhiB3 Maintains redox homeostasis by regulating virulence lipid anabolism to modulate macrophage

response. *PLoS Pathogens*, 5(8).

Singh, A. et al., 2007. *Mycobacterium tuberculosis* WhiB3 responds to O2 and nitric oxide via its [4Fe-4S] cluster and is essential for nutrient starvation survival. *Proceedings of the National Academy of Sciences of the United States of America*, 104(28), pp.11562–11567.

Slauch, J., 2011. How does the oxidative burst of macrophages kill bacteria? Still an open question. *Mol Microbiol*, 80(3), pp.580–583.

Smith, A.D. et al., 2005. Role of conserved cysteines in mediating sulfur transfer from IscS to IscU. *FEBS Letters*, 579(23), pp.5236–5240.

Smith, a. D. et al., 2001. Sulfur Transfer from IscS to IscU: The First Step in Iron-Sulfur Cluster Biosynthesis. *Journal of the American Chemical Society*, 123(44), pp.11103–11104.

Smith, L.J. et al., 2010. *Mycobacterium tuberculosis* WhiB1 is an essential DNA-binding protein with a nitric oxide-sensitive iron-sulfur cluster. *The Biochemical journal*, 432(3), pp.417–427.

Smith, P.K. et al., 1985. Measurement of protein using bicinchoninic acid. *Analytical biochemistry*, 150, pp.76–85.

Soliveri, J. et al., 2000. Multiple paralogous genes related to the *Streptomyces coelicolor* developmental regulatory gene whiB are present in *Streptomyces* and other actinomycetes. *Microbiology*, 146(Pt 2), pp.333–43.

Soutourina, O. et al., 2009. CymR, the master regulator of cysteine metabolism in *Staphylococcus aureus*, controls host sulphur source utilization and plays a role in biofilm formation. *Molecular Microbiology*, 73(2), pp.194–211.

Soutourina, O. et al., 2010. The pleiotropic cymr regulator of *Staphylococcus aureus* plays an important role in virulence and stress response. *PLoS Pathogens*, 6(5), pp.1–13.

Stanley, S.A. et al., 2003. Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22), pp.13001–13006.

Stapleton, M.R. et al., 2012. *Mycobacterium tuberculosis* WhiB1 represses transcription of the essential chaperonin GroEL2. *Tuberculosis*, 92(4), pp.328–332.

Stevanin, T.M., Read, R.C. & Poole, R.K., 2007. The hmp gene encoding the NO-inducible flavohaemoglobin in *Escherichia coli* confers a protective advantage in resisting killing within macrophages, but not in vitro: Links with swarming motility. *Gene*, 398(1–2 SPEC. ISS.), pp.62–68.

Stoll, H. et al., 2005. *Staphylococcus aureus* deficient in lipidation of prelipoproteins is attenuated in growth and immune activation. *Infection and Immunity*, 73(4), pp.2411–2423.

Streicher, S., Gurney, E. & Valentine, R.C., 1971. Transduction of the nitrogen-fixation genes in *Klebsiella pneumoniae*. *Proceedings of the National Academy of Sciences of the United States of America*, 68(6), pp.1174–1177.

Takano, E. et al., 2003. A rare leucine codon in *adpA* is implicated in the morphological defect of *bldA* mutants of *Streptomyces coelicolor*. *Molecular Microbiology*, 50(2), pp.475–486.

Takeda, K. et al., 2003. Crystal structures of bacterial lipoprotein localization factors, LolA and LolB. *EMBO Journal*, 22(13), pp.3199–3209.

Taniguchi, N., Matsuyama, S.I. & Tokuda, H., 2005. Mechanisms Underlying Energy-independent Transfer of Lipoproteins from LolA to LolB, Which Have Similar Unclosed β-Barrel Structures. *Journal of Biological Chemistry*, 280(41), pp.34481–34488.

Tanous, C. et al., 2008. The CymR regulator in complex with the enzyme CysK controls cysteine metabolism in *Bacillus subtilis*. *Journal of Biological Chemistry*, 283(51), pp.35551–35560.

Tesson, A.R. et al., 2003. Revisiting the steady state kinetic mechanism of glutamine-dependent asparagine synthetase from *Escherichia coli*. *Archives of Biochemistry and Biophysics*, 413(1), pp.23–31.

Thompson, B.J. et al., 2010. Investigating lipoprotein biogenesis and function in the model Gram-positive bacterium *Streptomyces coelicolor*. *Molecular microbiology*, 77(June), pp.943–957.

Thompson, B.J., 2010. INVESTIGATING THE LIPOPROTEIN BIOSYNTHETIC PATHWAY IN STREPTOMYCES SPECIES.

Tjalsma, H., Zanen, G., et al., 1999. The potential active site of the lipoprotein-specific (type II) signal peptidase of *Bacillus subtilis*. *Journal of Biological Chemistry*, 274(40), pp.28191–28197.

Tjalsma, H., Kontinen, V.P., et al., 1999. The role of lipoprotein processing by signal peptidase II in the Gram-positive eubacterium *Bacillus subtilis*. Signal peptidase II is required for the efficient secretion of alpha-amylase, a non-lipoprotein. *The Journal of biological chemistry*, 274(3), pp.1698–1707.

Tokuda, H., Sander, P. & Lee, B., 2014. Bacterial lipoproteins: biogenesis, virulence/pathogenicity and trafficking. In et al. Remaut H, Fronzes R, ed. *Bacterial Membranes: Structural and Molecular Biology*. UK: Caister Academic Press.

Tokumoto, U. & Takahashi, Y., 2001. Genetic Analysis of the isc Operon in Escherichia coli Involved in the Biogenesis of Cellular Iron-Sulfur Proteins. *The Journal of Biochemistry*, 130(1), pp.63–71.

Tokunaga, M., Loranger, J.M. & Wu, H.C., 1984. Prolipoprotein modification and processing enzymes in Escherichia coli. *Journal of Biological Chemistry*, 259(6), pp.3825–3830.

Tokunaga, M., Tokunaga, H. & Wu, H.C., 1982. Post-translational modification and processing of *Escherichia coli* prolipoprotein in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 79(7), pp.2255–2259.

Tong, Y. et al., 2015. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synthetic Biology*, p.150325161425009.

Tran, N.T. et al., 2011. Identification and characterization of CdgB, a diguanylate cyclase involved in developmental processes in *Streptomyces coelicolor*. *Journal of Bacteriology*, 193(12), pp.3100–3108.

Trapnell, C. et al., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7(3), pp.562–78.

Tschowri, N. et al., 2014. Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces development*. *Cell*, in press(5), pp.1136–1147.

Tschumi, A. et al., 2012. Functional analyses of mycobacterial lipoprotein diacylglyceryl transferase and comparative secretome analysis of a mycobacterial *lgt* mutant. *Journal of bacteriology*, 194(15), pp.3938–49.

Tschumi, A. et al., 2009. Identification of apolipoprotein N-acyltransferase (Lnt) in

mycobacteria. *Journal of Biological Chemistry*, 284(40), pp.27146–27156.

Tucker, N.P. et al., 2008. The transcriptional repressor protein NsrR senses nitric oxide directly via a [2Fe-2S] cluster. *PloS one*, 3(11), p.e3623.

Tucker, N.P. et al., 2010. There's NO stopping NsrR, a global regulator of the bacterial NO stress response. *Trends in microbiology*, 18(4), pp.149–56.

Typas, A. et al., 2010. Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell*, 143(7), pp.1097–1109.

Udaka, S., 2008. *The Discovery of Corynebacterium glutamicum and Birth of Amino Acid Fermentation Industry in Japan*,

Ueda, K. et al., 2005. Dual Transcriptional Control of *amfTSBA*, Which Regulates the Onset of Cellular Differentiation in *Streptomyces griseus*. *Society*, 187(1), pp.135–142.

Urbina, H.D. et al., 2001. Transfer of Sulfur from IscS to IscU during Fe/S Cluster Assembly. *Journal of Biological Chemistry*, 276(48), pp.44521–44526.

Ventura, M. et al., 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and molecular biology reviews : MMBR*, 71(3), pp.495–548.

Vidal-Ingigliardi, D., Lewenza, S. & Buddelmeijer, N., 2007. Identification of essential residues in apolipoprotein N-acyl transferase, a member of the CN hydrolase family. *Journal of bacteriology*, 189(12), pp.4456–64.

Vine, C.E., Purewal, S.K. & Cole, J. a, 2011. NsrR-dependent method for detecting nitric oxide accumulation in the *Escherichia coli* cytoplasm and enzymes involved in NO production. *FEMS microbiology letters*, 325(2), pp.108–14.

Vinella, D. et al., 2013. In vivo [Fe-S] cluster acquisition by IscR and NsrR, two stress regulators in *Escherichia coli*. *Molecular microbiology*, 87(3), pp.493–508.

Vogt, R.N. et al., 2003. The metabolism of nitrosothiols in the Mycobacteria: identification and characterization of S-nitrosomycothiol reductase. *The Biochemical journal*, 374(Pt 3), pp.657–666.

Wang, E. et al., 2011. Small-angle X-ray scattering study of a rex family repressor: Conformational response to NADH and NAD+ binding in solution. *Journal of Molecular Biology*, 408(4), pp.670–683.

Wang, P. et al., 2008. *Escherichia coli* signal peptide peptidase A is a serine-lysine protease with a lysine recruited to the nonconserved amino-terminal domain in the S49 protease family. *Biochemistry*, 47(24), pp.6361–6369.

Wang, T. et al., 2004. The *sufR* Gene (*sll0088* in *Synechocystis* sp. Strain PCC 6803) Functions as a Repressor of the *sufBCDS* Operon in Iron-Sulfur Cluster Biogenesis in Cyanobacteria. *Journal of Bacteriology*, 186(4), pp.956–967.

Wayne Outten, F., 2014. Recent advances in the Suf Fe–S cluster biogenesis pathway: Beyond the Proteobacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853(6), pp.1464–1469.

Weidinger, A. & Kozlov, A., 2015. Biological Activities of Reactive Oxygen and Nitrogen Species: Oxidative Stress versus Signal Transduction. *Biomolecules*, 5(2), pp.472–484.

Weiss, B., 2006. Evidence for mutagenesis by nitric oxide during nitrate metabolism in *Eschenchia coli*. *Journal of Bacteriology*, 188(3), pp.829–833.

Weiss, J.B., Ray, P.H. & Bassford, P.J., 1988. Purified *secB* protein of *Escherichia coli* retards folding and promotes membrane translocation of the maltose-binding protein in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 85(23), pp.8978–8982.

Wexler, M. et al., 2000. TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in sec-independent protein export. *The Journal of biological chemistry*, 275(22), pp.16717–22.

Widdick, D. a et al., 2011. Dissecting the complete lipoprotein biogenesis pathway in *Streptomyces scabies*. *Molecular microbiology*, 80(5), pp.1395–412.

Widdick, D. a et al., 2006. The twin-arginine translocation pathway is a major route of protein export in *Streptomyces coelicolor*. *Proceedings of the National Academy of Sciences of the United States of America*, 103(47), pp.17927–32.

Wild, K. et al., 2004. SRP meets the ribosome. *Nature structural & molecular biology*, 11(11), pp.1049–53.

Willemse, J., Mommaas, A. & van Wezel, G., 2012. Constitutive expression of *ftsZ* overrides the whi developmental genes to initiate sporulation of *Streptomyces coelicolor*. *Antonie van Leeuwenhoek*, 101(3), pp.619–32.

Willey, J. et al., 1991. Extracellular complementation of a developmental mutation implicates a small sporulation protein in aerial mycelium formation by *S.*

coelicolor. *Cell*, 65(4), pp.641–650.

Williams, M.G. et al., 1989. Identification and genetic mapping of the structural gene for an essential *Escherichia coli* membrane protein. *Journal of Bacteriology*, 171(1), pp.565–568.

Xiao, Y. et al., 2012. Myxobacterium-produced antibiotic TA (myxovirescin) inhibits type II signal peptidase. *Antimicrobial Agents and Chemotherapy*, 56(4), pp.2014–2021.

Xiao, Y. & Wall, D., 2014. Genetic redundancy, proximity, and functionality of *lspA*, the target of antibiotic TA, in the *Myxococcus xanthus* producer strain. *Journal of Bacteriology*, 196(6), pp.1174–1183.

Xu, W. et al., 2010. Regulation of morphological differentiation in *S. coelicolor* by RNase III (AbsB) cleavage of mRNA encoding the AdpA transcription factor. *Molecular Microbiology*, 75(3), pp.781–791.

Xu, Y. et al., 2012. A two-step mechanism for the activation of actinorhodin export and resistance in *Streptomyces coelicolor*. *mBio*, 3(5), pp.1–11.

Yamagata, H., Daishima, K. & Mizushima, S., 1983. Cloning and expression of a gene coding for the prolipoprotein signal peptidase of *Escherichia coli*. *FEBS letters*, 158(2), pp.301–304.

Yamazaki, H., Ohnishi, Y. & Horinouchi, S., 2000. An A-factor-dependent extracytoplasmic function sigma factor (σ(AdsA)) that is essential for morphological development in *Streptomyces griseus*. *Journal of Bacteriology*, 182(16), pp.4596–4605.

Yamazaki, K., 2014. Glutamine–Fructose-6-Phosphate Transaminase 1,2 (GFPT1,2). In *Handbook of Glycosyltransferases and Related Genes*. pp. 1465–1479.

Yang, J. et al., 2015. Deletion of the proposed iron chaperones IscA/SufA results in accumulation of a red intermediate cysteine desulfurase IscS in *Escherichia coli*. *Journal of Biological Chemistry*, 290(22), p.jbc.M115.654269.

Yang, Y.H. et al., 2008. Finding new pathway-specific regulators by clustering method using threshold standard deviation based on DNA chip data of *Streptomyces coelicolor*. *Applied Microbiology and Biotechnology*, 80(4), pp.709–717.

Yem, D.W. & Wu, H.C., 1978. Physiological characterization of an *Escherichia coli*

mutant altered in the structure of murein lipoprotein. *Journal of Bacteriology*, 133(3), pp.1419–1426.

Yen, M.-R. et al., 2002. Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Archives of microbiology*, 177(6), pp.441–50.

Yeo, W.S. et al., 2006. IscR acts as an activator in response to oxidative stress for the *suf* operon encoding Fe-S assembly proteins. *Molecular Microbiology*, 61(1), pp.206–218.

Yoshioka, H., Bouteau, F. & Kawano, T., 2008. Discovery of oxidative burst in the field of plant immunity: Looking back at the early pioneering works and towards the future development. *Plant signaling & behavior*, 3(3), pp.153–155.

Youn, H.D. et al., 1996. A novel nickel-containing superoxide dismutase from *Streptomyces* spp. *The Biochemical journal*, 318 (Pt 3, pp.889–896.

Yukl, E. et al., 2008. Transcription Factor NsrR from *Bacillus subtilis* Senses Nitric Oxide with a 4Fe− 4S Cluster†. *Biochemistry*, 47(49), pp.13084–13092.

Yuwen, L. et al., 2013. The role of aromatic L-amino acid decarboxylase in bacillamide C biosynthesis by *Bacillus atrophaeus* C89. *Scientific reports*, 3, p.1753.

Zhou, Z. et al., 2014. Function and Evolution of Two Forms of SecDF Homologs in *Streptomyces coelicolor*. *PloS one*, 9(8), p.e105237.

Zückert, W.R., 2014. Secretion of Bacterial Lipoproteins: Through the Cytoplasmic Membrane, the Periplasm and Beyond. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(8), pp.1509–1516.

7 Appendix and Supplementary data

File 1

Published paper 1 (NsrR from *Streptomyces coelicolor* is a Nitric Oxide-Sensing [4Fe-4S] Cluster Protein with a Specialized Regulatory Function)

File 2

Published paper 2 (Differentiated, Promoter-specific Response of [4Fe-4S] NsrR DNA-binding to Reaction with Nitric Oxide)

File 3

Published paper 3 (Cosmid based mutagenesis causes genetic instability in *Streptomyces coelicolor*, as shown by targeting of the lipoprotein signal peptidase gene)

File 4

Published paper 4 (RsrR: a novel redox sensitive Rrf2 family transcription factor in Streptomyces venezuelae)

File 5

S3.1 - Predicted NsrR binding sites 2008 Rodionov.docx

File 6

S3.2 - Virtual footprinting.xlsx

File 7

S4.1 - *rrf2* Reciprical BLAST results strepDB.xlsx

File 8

S4.2 - Streptomyces Rrf2 protein Aligenment.docx

File 9

S4.3 - pMS82 Sven nsrR-3xflag.ape

File 10

S4.4 – Combined NGS (by tab - Genome annotation, Combined ChIP and dRNA-seq, ChIP-Seq CLC workbench 8, TSS WT +ve strand (TSSAR), TSS WT -ve

strand (TSSAR), TSS RsrR +ve strand (TSSAR), TSS RsrR -ve strand (TSSAR), dRNA-seq (Tuxedo), dRNA-seq >1 and <-1 (Tuxedo), dRNA-seq >2 and <-2 (Tuxedo) ChIP dRNA-seq overlap)

File 11

S4.5 - BLAST Heat map.tiff

File 12

S5.1-S5.4 - Lsp chapter supplementary data.docx (S5.1. Sequence of *scr6809,* S5.2. Sequence of M145 *sco6811-08* loci, S5.3. Sequence of BJT1004 *sco6811-08* loci, S5.4. Sequence results of JTM018.01, combined sequences from JM0150 and JM0151).

⌘ *Author's Choice*

NsrR from *Streptomyces coelicolor* Is a Nitric Oxide-sensing [4Fe-4S] Cluster Protein with a Specialized Regulatory Function*

Jason C. Crack[‡], John Munnoch[§], Erin L. Dodd[‡], Felicity Knowles[§], Mahmoud M. Al Bassam[§], Saeed Kamali[¶], Ashley A. Holland[‖], Stephen P. Cramer[¶], Chris J. Hamilton[**], Michael K. Johnson[‖], Andrew J. Thomson[‡], Matthew I. Hutchings[§1], and Nick E. Le Brun[‡2]

From the [‡]*Centre for Molecular and Structural Biochemistry, School of Chemistry*, the [§]*School of Biological Sciences*, and the [**]*School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom*, the [¶]*Department of Chemistry, University of California, Davis, California 95616*, and the [‖]*Department of Chemistry and Center for Metalloenzyme Studies, University of Georgia, Athens, Georgia 30602*

Background: NsrR family proteins are [2Fe-2S] or [4Fe-4S] cluster-containing global regulators.
Results: *Streptomyces coelicolor* NsrR regulates only three genes, and it is the [4Fe-4S] form of the protein that binds tightly to NsrR-regulated promoters.
Conclusion: [4Fe-4S] NsrR has a specialized function associated only with nitric oxide stress response.
Significance: Members of the NsrR family are most likely all [4Fe-4S] proteins.

The Rrf2 family transcription factor NsrR controls expression of genes in a wide range of bacteria in response to nitric oxide (NO). The precise form of the NO-sensing module of NsrR is the subject of controversy because NsrR proteins containing either [2Fe-2S] or [4Fe-4S] clusters have been observed previously. Optical, Mössbauer, resonance Raman spectroscopies and native mass spectrometry demonstrate that *Streptomyces coelicolor* NsrR (ScNsrR), previously reported to contain a [2Fe-2S] cluster, can be isolated containing a [4Fe-4S] cluster. ChIP-seq experiments indicated that the ScNsrR regulon is small, consisting of only *hmpA1*, *hmpA2*, and *nsrR* itself. The *hmpA* genes encode NO-detoxifying flavohemoglobins, indicating that ScNsrR has a specialized regulatory function focused on NO detoxification and is not a global regulator like some NsrR orthologues. EMSAs and DNase I footprinting showed that the [4Fe-4S] form of ScNsrR binds specifically and tightly to an 11-bp inverted repeat sequence in the promoter regions of the identified target genes and that DNA binding is abolished following reaction with NO. Resonance Raman data were consistent with cluster coordination by three Cys residues and one oxygen-containing residue, and analysis of ScNsrR variants suggested that highly conserved Glu-85 may be the fourth ligand. Finally, we demonstrate that some low molecular weight thiols, but importantly not physiologically relevant thiols, such as cysteine and an analogue of mycothiol, bind weakly to the [4Fe-4S] cluster, and exposure of this bound form to O_2 results in cluster conversion to the [2Fe-2S] form, which does not bind to DNA. These data help to account for the observation of [2Fe-2S] forms of NsrR.

Nitric oxide (NO) is a reactive, lipophilic radical that can freely diffuse into cells. At low (nanomolar) concentrations, NO functions principally as a signaling molecule (*e.g.* via the reversible coordination of NO to the heme group in soluble guanylate cyclase to facilitate vasodilation in higher eukaryotes) and more widely through the process of thiol *S*-nitrosation, a regulatory process well characterized in eukaryotes (1) but also now recognized in bacteria (2). At higher concentrations (micromolar), NO is cytotoxic due to its reactivity with a wide range of targets resulting in nitrosation of amino acids (*e.g.* tryptophan) (3), nitrosative DNA damage (4), and nitrosylation of protein metallocofactors, particular those containing iron-sulfur (Fe-S) clusters (5). This property is exploited by mammalian macrophages in response to infection by pathogenic bacteria (6). Non-pathogenic bacteria also encounter significant concentrations of NO, through the activity of denitrifying species but also through the internal generation of NO resulting from the reduction of nitrite by nitrate reductases (7) and by the bacterial NO synthase enzymes encoded by some Gram-positive soil bacteria (8).

In order to survive, bacteria need to be able to counter the deleterious effects of NO. As a result, many bacteria have evolved a suite of specific iron-containing proteins to sense NO. Although the bacterial regulators SoxR and FNR are involved in coordinating the cell's response to NO (9–11), the primary functions of these regulators lie in sensing superoxide/redox stress and O_2, respectively. However, two recently discovered regulatory proteins in *Escherichia coli* appear to be dedicated to sensing NO. NorR senses NO directly through a non-heme iron center and responds by switching on expression of the flavorubredoxin NorVW to detoxify NO (12). NsrR has also been shown to sense NO in *E. coli* and to switch on a regulon of at

[1] To whom correspondence may be addressed. E-mail m.hutchings@uea.ac.uk.
[2] To whom correspondence may be addressed. E-mail: n.le-brun@uea.ac.uk.

NsrR Is a [4Fe-4S] Regulator of NO Stress Response

least 60 genes (13), including *hmp*, which encodes an NO-detoxifying flavohemoglobin (14) that converts NO to nitrate (or nitrous oxide under anaerobic conditions). This suggested that NsrR is a global regulator of NO-induced stress, whereas NorR has a more specific role in NO detoxification and a small regulon of only three genes, *norR-VW* (15).

NsrR belongs to the Rrf2 superfamily of regulators that includes the Fe-S cluster biosynthesis regulator IscR (16). Sequence alignment of NsrR proteins from a range of organisms revealed three conserved cysteine residues (Cys-93, Cys-99, and Cys-105 in *Streptomyces coelicolor* NsrR) in the C terminus region that probably act as cluster ligands (17). Consistent with this, Cys to Ala substitutions in *Neisseria gonorrhoeae* NsrR relieved repression of a target promoter and reduced DNA binding activity *in vitro* (18). Purified NsrR from *S. coelicolor* (19), *N. gonorrhoeae* (18), and *Bacillus subtilis* (20) have all been shown to be Fe-S cluster-binding proteins. However, the nature of the cluster and the mechanism by which the protein functions to coordinate the response to NO stress are not clear. Our studies are focused on NsrR from *S. coelicolor*, a model organism for the genus *Streptomyces*, which are widespread saprophytic soil bacteria that produce more than half of all known antibiotics and belong to the high GC Gram-positive phylum Actinobacteria. *S. coelicolor* is an obligate aerobe and encodes two homologues of flavohemoglobin, HmpA1 (SCO7428) and HmpA2 (SCO7094). The gene encoding one of these homologues (HmpA1) is adjacent to the gene encoding NsrR (SCO7427). Initial aerobic purification of *S. coelicolor* NsrR (produced in *E. coli*) resulted in a [2Fe-2S] cluster form (19) that was found to bind specifically to the *S. coelicolor hmpA1* and *hmpA2* promoter regions. This was consistent with data for *N. gonorrhoeae* NsrR, which suggested that it also contains a [2Fe-2S] cluster (18). However, anaerobically purified *B. subtilis* NsrR was found to contain a [4Fe-4S] cluster (20) and was recently shown to bind the *B. subtilis nasD* (nitrite reductase) promoter in an NO-sensitive manner (21). Therefore, the current literature on NsrR does not provide a consistent view of the nature of the Fe-S cluster.

Here we report ChIP-seq analysis to define the *S. coelicolor* NsrR regulon and DNase I footprinting and EMSA studies that confirm the target promoters and binding site. Spectroscopic and native mass spectrometry studies of anaerobically purified NsrR are also described, which, together with DNA binding studies, establish the physiologically relevant form of NsrR. These also reveal conditions under which facile cluster conversion occurs, accounting for the observation of different cluster types in purified NsrR proteins.

EXPERIMENTAL PROCEDURES

Strains, Plasmids, Cosmids, Primers, and Growth Conditions—The strains, plasmids, cosmids, and primers used in this study are listed in Table 1. *E. coli* was routinely grown on Luria-Bertani (LB) broth or agar or modified LB lacking NaCl to select for hygromycin resistance. *S. coelicolor* strains were grown on mannitol soya flour agar (20 g of mannitol, 20 g of soya flour, 20 g of agar in 1 liter of tap water), Difco nutrient agar (BD Biosciences). Liquid cultures were grown in Difco

TABLE 1
Strains, plasmids, and primers used in this study

Strains/plasmids	Description	Reference
Strains		
S. coelicolor		
M145	SCP1⁻ SCP2⁻ *S. coelicolor* wild-type strain	Ref. 22
JTM001	M145 Δ*nsrR::apr*	This work
JTM002	M145 Δ*nsrR* (unmarked)	This work
JTM003	JTM002 containing pJM001	This work
E. coli		
DH5α + BT340	Flp recombinase *E. coli* expression strain	Ref. 60
BW25113 (pIJ790)	*E. coli* BW25113 containing λ RED recombination plasmid pIJ790	Ref. 60
ET12567 (pUZ8002)	*E. coli* Δ*dam dcm* strain containing helper plasmid pUZ8002	Ref. 60
Plasmid/cosmids		
St5C11	*S. coelicolor* cosmid containing genes SCO7423–SCO7460	Ref. 61
St3A4.2.A04	*S. coelicolor* cosmid containing genes SCO7067–SCO7103 and a hygromycin-marked transposon-disrupting SCO7094	Ref. 26
pSET152	Integrative *Streptomyces* vector	Ref. 22
pGS21a	*E. coli* expression vector	Ref. 62
pNsrR	Expression construct for untagged native ScNsrR	Ref. 19
pJM001	pSET152 encoding NsrR with a C-terminal 3 × FLAG tag sequence	This work
pJM002	pGS21a encoding NsrR with a C-terminal His₆ tag sequence	This work
pJM003	pJM002 containing a E85A mutation	This work
pJM004	pJM002 containing a D96A mutation	This work
pJM005	pJM002 containing a D113A mutation	This work
pJM006	pJM002 containing a E116A mutation	This work
pJM007	pJM002 containing a D123A mutation	This work
pJM008	pJM002 containing a D129A mutation	This work

nutrient broth (BD Biosciences) or a 50:50 mix of Tryptone soy broth and yeast extract/malt extract (22).

ChIP-seq—Experiments were performed using a Δ*nsrR* mutant strain expressing a C-terminal 3×FLAG-tagged NsrR protein with the parent Δ*nsrR* strain as a control. The coding sequence for NsrR-3×FLAG was synthesized by Genscript with the native *nsrR* promoter and introduced into *S. coelicolor* Δ*nsrR* on the integrative vector pMS82 (23). 1×10^8 spores of each strain were inoculated onto cellophane disks on mannitol soya flour agar plates (20 plates/strain) that were then grown for 48 h at 30 °C. The disks were removed and flipped so that the mycelium was submerged in 10 ml of a 1% (v/v) formaldehyde solution, within the Petri dish lids, for 20 min at room temperature to cross-link proteins to DNA. The disks were incubated in 10 ml of 0.5 M glycine for 5 min, and the mycelium was harvested, washed twice with 25 ml ice-cold PBS (pH 7.4), and incubated in 1 ml of lysis buffer (10 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mg/ml lysozyme, 1× protease inhibitor (Roche Applied Science, complete mini EDTA-free tablets) at 25 °C for 25 min. Samples were then placed on ice, and 1 ml of IP[3] buffer (100 mM Tris-HCl, pH 8.0, 250 mM NaCl, 0.5% (v/v) Triton X-100, 0.1% (w/v) SDS, 1× protease inhibitor) was added for 2 min prior to sonicating seven times at 50 Hz for 15 s each time. Material was centrifuged at 16,200 × g for 10 min at 4 °C, and

---

[3] The abbreviations used are: IP, immunoprecipitation; ScNsrR, *S. coelicolor* NsrR; Bistris propane, 1,3-bis[tris(hydroxymethyl)methylamino]propane; ESI, electrospray ionization; dMSH, 2-(*N*-acetylcysteinyl)amido-2-deoxy-D-glucopyranoside.

supernatants were recentrifuged as above. A 1-ml sample was used for IP, 25 μl was used to prepare total DNA, and the excess was stored at -20 °C.

The 1-ml IP sample was precleared using 100 μl of equilibrated 50% (v/v) protein A-Sepharose beads and incubated at 4 °C for 1 h on a rotating wheel. Samples were then centrifuged at 16,200 × g for 15 min at 4 °C, and 100 μl of a 1 mg/ml solution of α-FLAG antibody (Sigma-Aldrich) was added, and the solution was incubated overnight at 4 °C on a rotating wheel. 100 μl of equilibrated protein A-Sepharose beads was added and incubated for 4 h at 4 °C on a rotating wheel. Samples were centrifuged at 1200 × g for 5 min and washed twice with 1 ml of 0.5× IP buffer for 15 min with gentle agitation and then twice with 1 ml of 1× IP buffer for 15 min with gentle agitation. Each sample was split into two 0.5-ml aliquots and centrifuged at 1200 × g for 5 min to remove all supernatant before 150 μl of elution buffer (50 mM Tris-HCl, pH 7.6, 10 mM EDTA, 1% (w/v) SDS) was added (or 10 μl for the total DNA samples) and incubated at 65 °C overnight. Tubes were inverted seven times and centrifuged at 16,200 × g for 5 min. Supernatants were retained, and the beads were washed with a further 50 μl of Tris (10 mM) and EDTA (1 mM) (pH 7.8) at 65 °C for 5 min before centrifuging at 16,200 × g for 5 min. Supernatants were pooled and centrifuged again at 16,200 × g for 1 min, and proteinase K (2 μl of a 10 mg/ml stock) was added to the supernatant and incubated at 55 °C for 90 min. Then 200 μl of phenol/chloroform was added, and samples were vortex-mixed for 3 min and then centrifuged for 3 min at 16,200 × g. The upper phase was stored, and the organic phase was re-extracted with 100 μl of Tris (10 mM) and EDTA (1 mM) (pH 7.8). Samples were then purified using a QIAquick kit (Qiagen), eluted with 50 μl of ultrapure water (Sigma), and re-eluted with the eluate. DNA was quantified using a nanodrop ND2000c spectrophotometer (Thermo Fisher), and libraries were constructed and sequenced by the Genome Analysis Centre (Norwich, UK).

Electrophoretic Mobility Shift Assays (EMSAs)—DNA fragments carrying the *hmpA1* (SCO7428), *hmpA2* (SCO7094), or *nsrR* (SCO7427) promoters were PCR-amplified using *S. coelicolor* genomic DNA with 5′ 6-carboxyfluorescein-modified primers (see Table 1). The PCR products were extracted and purified using a QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. Probes were quantitated using a nanodrop ND2000c. The molecular weights of the double-stranded 6-carboxyfluorescein-labeled probes were calculated using OligoCalc (24). Band shift reactions (20 μl) were carried out in 10 mM Tris, 54 mM KCl, 0.3% (v/v) glycerol, 1.32 mM glutathione, pH 7.5. Briefly, 1 μl of DNA was titrated with aliquots of NsrR (20 μl final volume), typically to a 20-fold molar excess, and incubated on ice for ~10 min. Loading dye (2 μl, containing 0.3% (w/v) bromphenol blue), was added and the reaction mixtures were immediately separated at 30 mA for 30 min on a 5% (w/v) polyacrylamide gel in 1× TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA), using a Mini Protean III system (Bio-Rad). Gels were visualized (excitation, 488 nm; emission, 530 nm) on a molecular imager FX Pro (Bio-Rad). Polyacrylamide gels were prerun at 30 mA for 2 min prior to use.

DNase I Footprinting—Footprinting was carried out as described previously (25) with the following modifications.

DNA fragments carrying the *hmpA1* (SCO7428), *hmpA2* (SCO7094), or *nsrR* (SCO7427) promoters were PCR-amplified using the *S. coelicolor* cosmids 5C11 (*hmpA1* and *nsrR*) and 3A4 2.A04 (*hmpA2*) as templates (26, 27). In each case, one primer was end-labeled with ^{32}P (PerkinElmer Life Sciences) using T4 polynucleotide kinase (New England Biolabs) in a 20-μl labeling reaction (2.5 μl of primer (10 pmol/μl), 11.5 μl of water, 2 μl of 10× T4 polynucleotide kinase buffer, 1 μl of T4 polynucleotide kinase, and 3 μl of γ-^{32}P) incubated at 37 °C for 2 h and then 65 °C for 20 min. To this labeling reaction 30 μl of PCR mix was added (2.5 μl of second primer (10 pmol/μl), 1 μl of template (100 ng/μl), 1 μl of dNTP mix, 10 μl of 5× Q5 buffer, 10 μl of 5× GC enhancer, 5 μl of water, 0.5 μl of Q5 (supplied by New England Biolabs)), and thermal cycling conditions previously optimized using non-radiolabeled reagents were used. The subsequent PCR products were purified using QIAquick columns (Qiagen) according to the manufacturer's instructions. Binding reactions between DNA (~100,000 cpm) and NsrR (0–2 μM) were carried out for 30 min at room temperature in 40 μl of reaction buffer (10 mM Tris, 54 mM KCl, 0.3% (v/v) glycerol, pH 7.5) before treatment with 10 units of DNase I (Promega) and 1 μl of 100 mM CaCl$_2$ for 10–150 s. To terminate the reactions, 140 μl of stop solution (192 mM sodium acetate, 32 mM EDTA, 0.14% (w/v) SDS, 70 μg/ml yeast tRNA) was added and mixed by vortexing. Samples were extracted with 190 μl of phenol/chloroform, and the DNA-containing aqueous phase was ethanol-precipitated with 540 μl of 96% (v/v) ethanol. Pellets were dried and resuspended in 4 μl of loading dye (80% (v/v) formamide, 10 mM NaOH, 1 mM EDTA, 0.1% (w/v) xylene cyanol, 0.1% (w/v) bromphenol blue). A 6% (w/v) polyacrylamide sequencing gel with 8 M urea (Severn Biotech) was loaded with each sample in 1× TBE running buffer. The gel was maintained at 50 °C running at 1200 V to ensure uniform DNA denaturation and separation. Gels were transferred from glass plates to Whatman paper and dried for 30 min under vacuum. Labeled DNA was visualized using a PhosphorImager plate exposed for 16–24 h and scanned at 635 nm on a Typhoon FLA 9500 (GE Healthcare).

G+A ladders were produced based on the Sure track footprinting method. Labeled DNA (~150,000 cpm) was incubated with 1 μg of poly(dI-dC) and 1 μl of 4% (v/v) formic acid for 25 min at 37 °C. Tubes were placed on ice, and 150 μl of fresh 1 M piperidine was added and incubated for 30 min at 90 °C. Reactions were cooled on ice for 5 min, and 1 ml of butanol was added to the mixture and vortexed vigorously. Samples were then centrifuged for 2 min, the supernatant was removed, and 150 μl of 1% (w/v) SDS and 1 ml of butanol were added and vortexed vigorously at room temperature. Reactions were then centrifuged for 2 min at room temperature, and pellets were washed two times with 0.5 ml of butanol (stored at -20 °C), centrifuging between washes at 4 °C. The supernatant was removed, and the pellet was checked using a Geiger-Müller counter. Pellets were dried for 5–10 min in a vacuum concentrator and then dissolved in 2–5 μl of loading dye (80% (v/v) formamide, 10 mM NaOH, 1 mM EDTA, 0.1% (w/v) xylene cyanol, and 0.1% (w/v) bromphenol blue).

Purification of S. coelicolor NsrR—Wild type NsrR was overproduced in aerobically grown *E. coli* strain BL21λDE3 cultures

harboring pNsrR, as described previously (19). Cell pellets were washed with lysis buffer (50 mM Tris-HCl, 50 mM NaCl, 5% (v/v) glycerol, pH 7.1), transferred to the anaerobic cabinet, and stored at $-10\ ^{\circ}C$ in an anaerobic freezer (Belle Technology) until required. For Mössbauer studies, ^{57}Fe (Goss Scientific)-labeled ScNsrR was produced *in vivo* as described previously (28). Unless otherwise stated, all subsequent purification steps were performed under anaerobic conditions inside an anaerobic cabinet ($O_2 < 4$ ppm). Cell pellets were resuspended in lysis buffer with the addition of lysozyme (0.4 mg/ml), DNase I (1.3 μg/ml), 2 mM PMSF, and 1.3% (v/v) ethanol. The cell suspension was thoroughly homogenized by syringe, removed from the anaerobic cabinet, sonicated twice while on ice, and returned to the anaerobic cabinet. The cell suspension was transferred to O-ring sealed centrifuge tubes (Nalgene) and centrifuged outside of the cabinet at $40,000 \times g$ for 45 min at $1\ ^{\circ}C$. The supernatant was passed through a HiTrap DEAE column (2×5 ml; GE Healthcare), and the eluate was immediately loaded onto a HiTrap heparin column (3×5 ml; GE Healthcare) and washed with lysis buffer until $A_{280\ nm}$ was ≤ 0.1. The heparin column was then washed with buffer A (50 mM Tris-HCl, 50 mM NaCl, 5% (v/v) glycerol, pH 8.0), and bound proteins were eluted (1 ml/min) using a linear gradient (20 ml) from 10 to 100% (v/v) buffer B (50 mM Tris, 2 M NaCl, 5% (v/v) glycerol, pH 8.0). Fractions (1 ml) containing NsrR were pooled, diluted 10-fold with lysis buffer, transferred to O-ring sealed centrifuge tubes (Nalgene), and centrifuged outside of the cabinet at $40,000 \times g$ for 30 min at $1\ ^{\circ}C$. The supernatant was passed through a HiTrap DEAE column (5 ml) and immediately loaded onto a HiTrap heparin column (3×1 ml). The heparin column was then washed with buffer A containing 3% (v/v) buffer B and eluted using a linear gradient (2 ml) from 3 to 100% (v/v) buffer B. Fractions (1 ml) containing NsrR were pooled and stored in an anaerobic freezer until needed. Where necessary, gel filtration was carried out under anaerobic conditions using a Sephacryl S-100HR 16/50 column (GE Healthcare), equilibrated in buffer C (50 mM Tris, 100 mM NaCl, 5% (v/v) glycerol, pH 8) with a flow rate of 1 ml/min.

Protein concentrations were determined using the method of Smith (Pierce) (29) with bovine serum albumin as the standard. The iron and sulfide content of proteins were determined as described previously (30). This gave an extinction coefficient of $\epsilon_{406\ nm} = 13.30 \pm 0.19\ mM^{-1}\ cm^{-1}$, which was subsequently used to determine the $[4Fe-4S]^{2+}$ cluster concentration.

C-terminal His-tagged NsrR proteins (wild type and variants D85A, E96A, E113A, D116A, E123A, and E129A) were overproduced from pJM plasmids containing the SCO7427 sequence codon-optimized for *E. coli* (Genscript (Piscataway, NJ); see Table 1) in aerobically grown *E. coli* strain BL21λDE3, as described previously (28), except that 10 μM isopropyl 1-thio-β-D-galactopyranoside was used to induce protein expression. Cells were lysed in buffer C, as described above. The cleared cell lysate was loaded onto a HiTrap Ni^{2+} chelating column (2×5 ml), previously equilibrated with buffer C, and washed with 5% (v/v) buffer D (50 mM Tris, 100 mM NaCl, 200 mM L-histidine, 5% glycerol, pH 8.0). Bound proteins were eluted using a linear gradient (30 ml) from 5 to 50% (v/v) buffer D. Fractions (1 ml) containing NsrR were pooled, immediately

loaded onto a HiTrap heparin column, and eluted with buffer B, as described above.

Preparation of [2Fe2S]-NsrR—An aliquot of [4Fe-4S] NsrR was diluted to a final concentration of $\sim 70\ \mu$M cluster with 20 mM Tris, 20 mM Mes, 20 mM Bistris propane, 100 mM NaCl, 5% (v/v) glycerol, 5 mM DTT, pH 8.7, containing dissolved atmospheric oxygen, and gently agitated for ~ 50 min. The sample was immediately returned to the anaerobic chamber and buffer-exchanged (PD10 column, GE Healthcare) into phosphate buffer (50 mM potassium phosphate, 200 mM NaCl, pH 7.5). The sample was incubated at an ambient temperature for ~ 5 min and then centrifuged at $14,100 \times g$ for 2 min. The red pellet, containing [2Fe-2S] NsrR, was briefly washed with a minimal amount of phosphate buffer before being redissolved in buffer A containing 25 mM DTT. The supernatant, containing DTT-modified [4Fe-4S] NsrR, was discarded.

Preparation of Apo-NsrR—Native apo-NsrR was prepared from holoprotein using EDTA and potassium ferricyanide, as described previously (31), except that it was dialyzed against buffer A containing 5 mM DTT, and a HiTrap heparin column (5×1 ml) was used to isolate and concentrate the protein following dialysis. Briefly, the column was equilibrated with buffer A, and bound proteins were washed with 10 ml of buffer A containing 5.6 mM tris(2-carboxyethyl)phosphine and eluted using a linear gradient (20 ml) from 0% to 100% (v/v) buffer B.

Spectroscopy and Mass Spectrometry—UV-visible absorbance measurements were made with a Jasco V500 spectrometer, and CD spectra were measured with a Jasco J810 spectropolarimeter. Dissociation constants for the binding of low molecular weight thiols to [4Fe-4S] NsrR were determined by fitting plots of $\Delta CD_{374\ nm}$ *versus* thiol concentration to a single site binding equation using Origin software (version 8; OriginLab, Northampton, MA).

2-(*N*-Acetylcysteinyl)amido-2-deoxy-D-glucopyranoside (dMSH) was prepared as described previously (32). A 13.87 mM dMSH stock solution was prepared and was determined to be ~ 60% reduced (9.09 mM free thiol form) using a 5,5'-dithiobis-(nitrobenzoic acid) assay ($\epsilon_{412\ nm} \sim 14,150\ M^{-1}\ cm^{-1}$ (33)). To investigate the stability of the iron-sulfur cluster toward O_2, aliquots of protein ($\sim 10-45\ \mu$M cluster final concentration) and assay buffer (20 mM Tris, 20 mM MES, 20 mM Bistris propane, 100 mM NaCl, 5% (v/v) glycerol, pH 8.0) containing dissolved atmospheric O_2 ($234 \pm 3\ \mu$M) were combined and mixed by inversion in a sealed cuvette outside of the anaerobic cabinet in the presence or absence of dithiothreitol. Loss of the iron-sulfur cluster was monitored at 406 nm as a function of time.

Resonance Raman spectra were recorded at 21 K using a scanning Ramanor U1000 spectrometer (Instruments SA, Edison, NJ) and an Innova 10-watt argon ion laser (Coherent, Santa Clara, CA), with 15-μl frozen droplets of sample mounted on the cold finger of a Displex model CSA-202E closed cycle refrigerator (Air Products, Allentown, PA). Laser power at the sample was 30 milliwatts, and the spectrum reported was the sum of 90 scans, with each scan involving photon counting for 1 s every $0.5\ cm^{-1}$ and a spectral bandwidth of $7\ cm^{-1}$. Mössbauer measurements were performed using an MS4 spectrometer operating in the constant acceleration mode in transmission geometry. The measurements were performed at 10 K using a Janis

SVT-400 cryostat. 100 mCi of ^{57}Co in rhodium held at room temperature was used as the source. Centroid shifts, δ, are given with respect to metallic α-iron at room temperature. The spectra were least square fitted using Recoil software (34).

For native MS analysis, His-tagged NsrR was exchanged into 250 mM ammonium acetate, pH 7.1, using Zeba spin desalting columns (Thermo Scientific), diluted to ~6 μM cluster (6 pmol/μl), and infused directly (0.3 ml/h) into the ESI source of a Bruker micrOTOF-QIII mass spectrometer (Bruker Daltonics, Coventry, UK) operating in the positive ion mode. To study the effect of O_2 and low molecular weight thiols, His-tagged NsrR was exchanged into ammonium acetate under anaerobic conditions. The resulting sample was diluted to ~7 μM cluster with ammonium acetate buffer containing dissolved atmospheric oxygen (~240 μM) and 5 mM DTT or 1.1 M β-mercaptoethanol. Full mass spectra (*m/z* 50–3500) were recorded for 5 min. Spectra were combined, processed using the ESI Compass version 1.3 Maximum Entropy deconvolution routine in Bruker Compass Data analysis version 4.1 (Bruker Daltonik, Bremen, Germany). The mass spectrometer was calibrated with ESI-L low concentration tuning mix in the positive ion mode (Agilent Technologies, San Diego, CA).

RESULTS

Identification of ScNsrR Binding Sites in Vivo—To determine where ScNsrR binds on the *S. coelicolor* chromosome, ChIP-seq analysis was carried out on 48-h, mannitol soya flour agar-grown cultures of the Δ*nsrR* strain with and without an NsrR-3×FLAG expression construct, integrated in single copy. ChIP was performed using monoclonal anti-FLAG antibodies, and immunoprecipitated DNA was sequenced using Illumina Hi-Seq.

The most significantly enriched DNA sequences in the NsrR-3×FLAG strain (compared with the control strain) mapped to the promoter regions of *hmpA1* and *nsrR* (Fig. 1*A*). This was surprising, because *hmpA1* is a weak match to the previously predicted NsrR binding site (35), and *nsrR* does not match at all. Furthermore, the *hmpA2* promoter, which shows a strong match to the predicted binding site, showed relatively low (<2-fold) enrichment in the ChIP-seq data (Fig. 1*A*), although it was previously shown to be bound by purified ScNsrR *in vitro* (19). Alignment of the *nsrR*, *hmpA1*, and *hmpA2* promoters using MEME identified a conserved sequence at all three promoters, and alignment of these sequences generated a 23-base pair consensus ScNsrR binding site, which consists of two 11-base pair inverted repeats separated by a single base pair (Fig. 1*B*). This binding site contains the DNA sequence previously shown to be bound by ScNsrR at the *hmpA1* and *hmpA2* promoters using AUC (19) but is significantly different in sequence to both the experimentally verified *E. coli* and *B. subtilis* NsrR binding sites (13, 20) and the predicted binding site for *Streptomyces* and Bacillales NsrR (35).

Anaerobic Purification of ScNsrR Results in a [4Fe-4S] Cluster-bound Dimer—In order to validate the ChIP-seq data and analyze the ScNsrR binding sites at the three target promoters *in vitro*, it was necessary to purify the ScNsrR protein. Previous aerobic purification of ScNsrR in the presence of DTT, following overproduction in *E. coli*, resulted in a [2Fe-2S] form at a

level of ~30% cluster incorporation (19). A new strategy was devised to purify ScNsrR under anaerobic conditions and in the absence of any low molecular weight thiols (see "Experimental Procedures"). This resulted in a dark brown solution indicative of the presence of an Fe-S cluster. The UV-visible absorbance spectrum (Fig. 2*A*) revealed a broad absorbance band with a maximum at 406 nm (ε = 13302 ± 196 M^{-1} cm^{-1}) and a pronounced shoulder feature at 320 nm. Broad weaker bands were observed in the 550–750 nm region. The spectrum is very similar in form to a number of [4Fe-4S] cluster-containing proteins (30, 31) and is quite distinct from that previously published for NsrR, which was characteristic of the redder color of a [2Fe-2S] cluster (19).

Because the electronic transitions of iron-sulfur clusters become optically active as a result of the fold of the protein in which they are bound, CD spectra reflect the cluster environment (36). The near UV-visible CD spectrum of NsrR (Fig. 2*B*) contained two positive features at 330 and 530 nm and a major negative feature at 400 nm, with smaller features at 570 and 640 nm. Although the sign of the band at 330 nm is reversed, the spectrum is otherwise similar to that of *S. coelicolor* WhiD, which contains a [4Fe-4S] cluster (31), and is again quite distinct from the previously published CD spectrum of NsrR (19).

Mössbauer spectroscopy provides definitive and quantitative determination of the type of iron-sulfur clusters present in a sample (37), so the spectrum of as-isolated, ^{57}Fe-enriched NsrR was measured (Fig. 2*C*). The data fit best to two quadrupole doublets with similar isomer shifts (δ) and quadrupole splitting (ΔE$_Q$), one having δ = 0.442 mm/s and ΔE$_Q$ = 1.031 mm/s and the other having δ = 0.481 mm/s and ΔE$_Q$ = 1.309 mm/s. Each doublet arises from a valence-delocalized [2Fe-2S]$^+$ pair that couple together to form an S = 0 [4Fe-4S]$^{2+}$ cluster (38). The isomer shifts and quadrupole splittings of both doublets are characteristic of [4Fe-4S]$^{2+}$ clusters and are very similar to those reported for MiaB and lipoyl synthase, which both contain [4Fe-4S]$^{2+}$ clusters that are coordinated by three Cys residues (39, 40). Furthermore, the Mössbauer parameters are markedly different from those of [2Fe-2S]$^{2+}$ clusters, including that of IscR (41).

The low temperature (21 K) resonance Raman spectrum of NsrR (488-nm excitation) in the iron-sulfur stretching region (250–450 cm^{-1}) is shown in Fig. 2*D*. The Fe-S stretching frequencies and relative resonance enhancements are characteristic of a [4Fe-4S]$^{2+}$ cluster (42, 43) and are similar to those reported for *B. subtilis* [4Fe-4S] NsrR at room temperature (20). The bands are readily assigned by analogy with isotopically labeled model complexes and simple [4Fe-4S] ferredoxins under idealized T_d or D_{2d} symmetry (42), with mainly terminal Fe-S stretching modes at ~389 and 363 cm^{-1} and mainly bridging Fe-S stretching modes at ~389, 343, 300, 281, 266, and 253 cm^{-1} (both terminal and bridging Fe-S stretching modes are likely to contribute to the broad band at 389 cm^{-1}). Previous studies of proteins have identified the frequency of the intense symmetric Fe-S stretching mode of the [4Fe-4S]$^{2+}$ core as an indicator of ligation of a unique iron site by an oxygenic ligand, with all cysteinyl-ligated [4Fe-4S]$^{2+}$ exhibiting frequencies spanning 333–339 cm^{-1} and those with one Asp or Ser ligand exhibiting frequencies spanning 340–343 cm^{-1} at low temper-
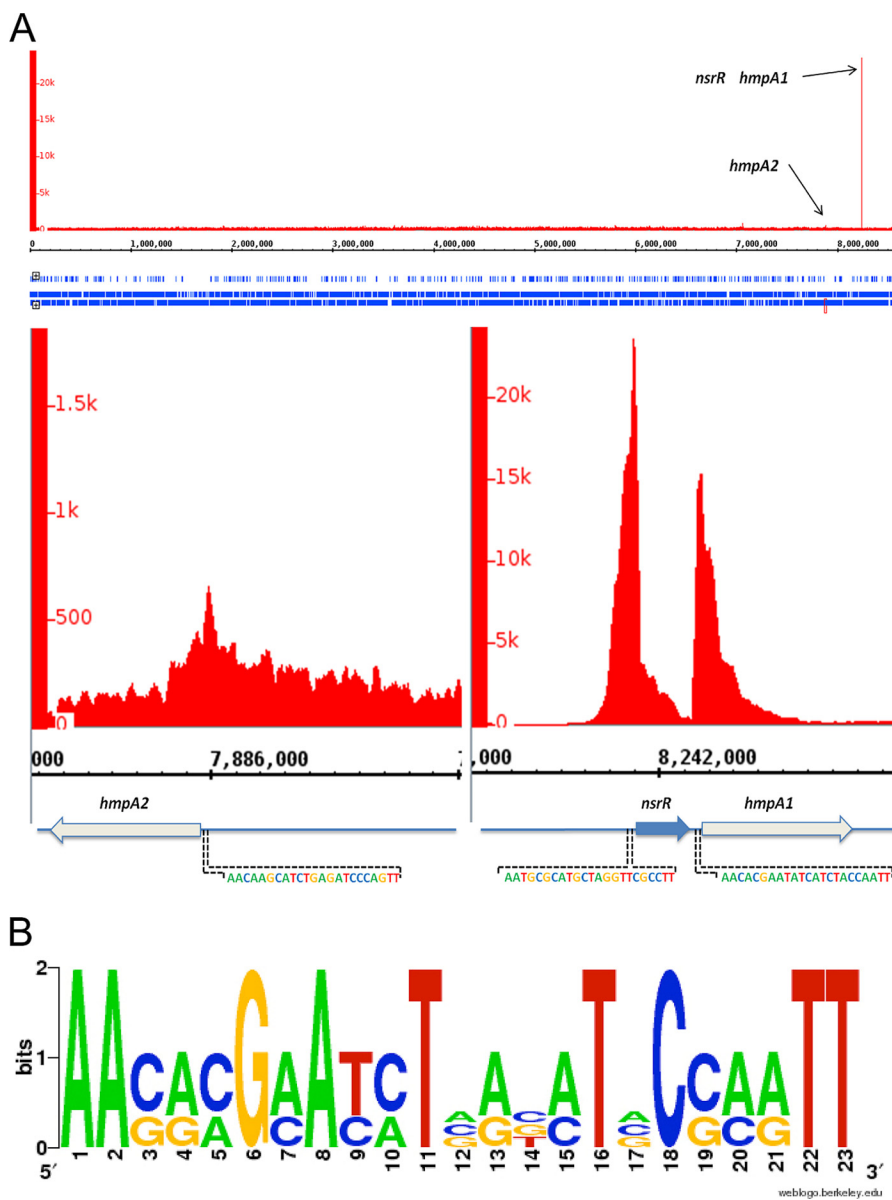
NsrR Is a [4Fe-4S] Regulator of NO Stress Response

FIGURE 1. **Genes regulated by NsrR in _S. coelicolor_.** _A, top,_ whole genome view of the ChIP-seq data for strain JM1002 (Δ_nsrR_ expressing NsrR-3×FLAG), visualized using the Integrated Genome Browser (available on the BioViz Web site), showing only three enriched peaks when compared with the control strain JM1001 (Δ_nsrR_) that map to the _nsrR, hmpA1,_ and _hmpA2_ promoters. _Bottom,_ the same JM1002 ChIP-seq data but _zoomed in_ to view the _hmpA2, nsrR,_ and _hmpA1_ genes and the enrichment peaks at their respective promoters. The MEME-predicted ScNsrR binding site at each promoter is also shown. _B,_ NsrR WebLogo generated by alignment of the three MEME-predicted NsrR sites at the _nsrR, hmpA1,_ and _hmpA2_ promoters.

atures (≥77 K) (43, 44). Consequently, the high frequency of the symmetric bridging Fe-S stretching mode of the $[4Fe-4S]^{2+}$ in NsrR (343 cm^{-1}) is highly indicative of oxygenic ligation at a unique site of the $[4Fe-4S]^{2+}$ cluster.

Native mass spectrometry was used to provide high resolution mass data of cluster-bound NsrR (see Fig. 2E). Here, a C-terminal His-tagged form of the protein was ionized in a volatile aqueous buffered solution that enabled it to remain folded with its bound cluster intact. The deconvoluted mass spectrum contained several peaks. The apoprotein was observed at 17,474 Da (predicted mass 17,474 Da), and there were adduct peaks at +23 and +64 Da due to Na$^+$ (commonly observed in native mass spectra) and most likely two additional sulfurs (Cys residues of iron-sulfur cluster proteins appear to readily pick up additional sulfurs as persulfides (45)), respec-

tively. The peak at 17,823 Da corresponds to the protein containing a [4Fe-4S] cluster with three deprotonated coordinating Cys residues (predicted mass = 17,474 − 3 + 352 = 17,823 Da). As for the apoprotein, peaks corresponding to Na$^+$ and sulfur adducts of the [4Fe-4S] species were observed.

Previous studies of _S. coelicolor_ NsrR revealed that the protein was a dimer in both [2Fe-2S] and apo-forms (19). The native mass spectrum of [4Fe-4S] NsrR did not reveal a dimeric form of NsrR. This may be because the dimeric form is not able to survive the ionization/vaporization process or because the protein is monomeric. To investigate this, anaerobic gel filtration of as-isolated NsrR (containing 60% holoprotein) was carried out. This gave a single elution band corresponding to a molecular mass of ~37 kDa (see Fig. 2F). Removal of the cluster to generate a homogeneous apoprotein sample also gave rise to
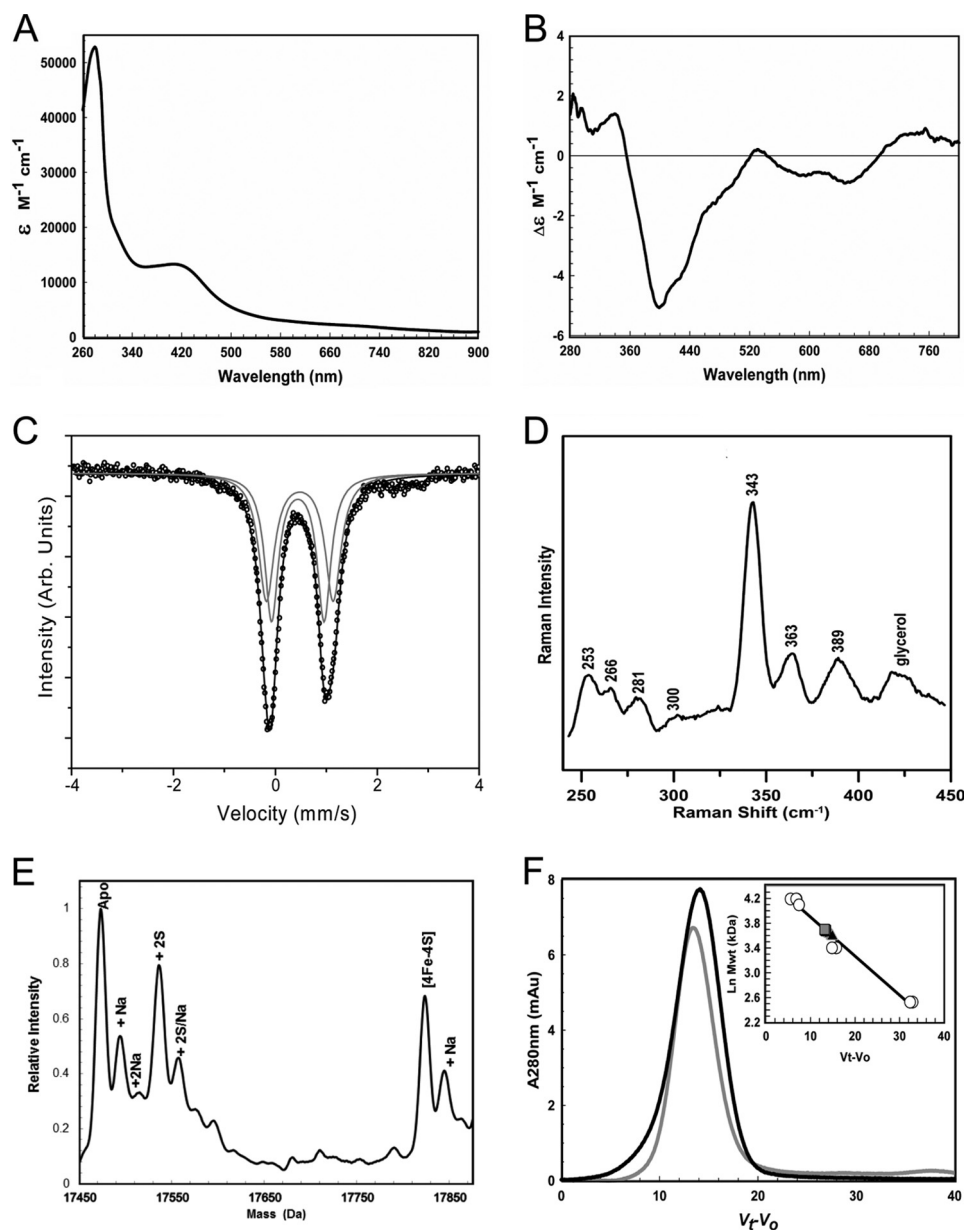
FIGURE 2. **Spectroscopic characterization of NsrR.** *A*, UV-visible absorption spectrum of 665 μM [4Fe-4S] NsrR, as isolated ($\sim$60% cluster-loaded). *B*, CD spectrum of an identical sample. Extinction coefficients relate to the [4Fe-4S] cluster. The buffer was 50 mM Tris, 800 mM NaCl, 5% (v/v) glycerol, pH 8.0. *C*, Mössbauer spectrum of $\sim$0.75 mM [4Fe-4S] NsrR enriched with ^{57}Fe. *D*, resonance Raman spectrum of $\sim$1.60 mM [4Fe-4S] NsrR. Excitation was at 488 nm, and temperature was 21 K. We note that the higher frequencies compared with those reported for *B. subtilis* NsrR are at least in part due to temperature difference (room temperature for *B. subtilis* NsrR) (20). The buffer was 50 mM Tris, 2 M NaCl, 5% (v/v) glycerol, pH 8.0, for *B* and *C*, respectively. *E*, positive ion mode ESI-TOF native mass spectrum of $\sim$7.5 μM [4Fe-4S] NsrR in 250 mM ammonium acetate pH 8.0. *m/z* spectra were deconvoluted with Bruker Compass Data analysis with the Maximum Entropy plugin. *F*, gel filtration analysis of NsrR association state. [4Fe-4S] (*black line*) and apo-form (*gray line*) samples of varying concentration (4–32 μM protein) were loaded in the presence or absence of DTT. *Inset*, calibration curve for the Sephacryl 100HR column. Standard proteins (*open circles*) were BSA (66 kDa), apo-FNR (30 kDa), and cytochrome *c* (13 kDa). [4Fe-4S]-NsrR and apo-NsrR are shown as a *black triangle* and *gray square*, respectively. The buffer was 50 mM Tris, 50 mM NaCl, 5% glycerol, $\pm$2.5 mM DTT, pH 8.0. *mAu*, milliabsorbance units.

a single elution band at a mass of $\sim$40 kDa, consistent with the previous report (19).

The data presented here clearly indicate that under anaerobic conditions, the protein is isolated containing a [4Fe-4S]$^{2+}$ cluster and is a homodimer, irrespective of the presence of a cluster.

[4Fe-4S] ScNsrR Binds Tightly to NsrR-regulated Promoters— It was previously concluded that [2Fe-2S] ScNsrR binds to the *hmpA1* and *hmpA2* promoters (19) and the ChIP-seq data show that both *hmpA* promoters are bound by ScNsrR *in vivo* (Fig.

1A). Thus, it was of interest to investigate the binding properties of the [4Fe-4S] form with the same promoters and with the *nsrR* promoter, which we identified as an additional ScNsrR target using ChIP-seq (Fig. 1A). EMSA experiments were conducted with fluorescently (6-carboxyfluorescein) labeled PCR fragments carrying each of the three promoters, as described under "Experimental Procedures," and the data for binding to *hmpA1* are shown in Fig. 3A. Increasing the concentration of as-isolated ScNsrR resulted in a clear shift in the mobility of the promoter DNA, and although the significance of the double
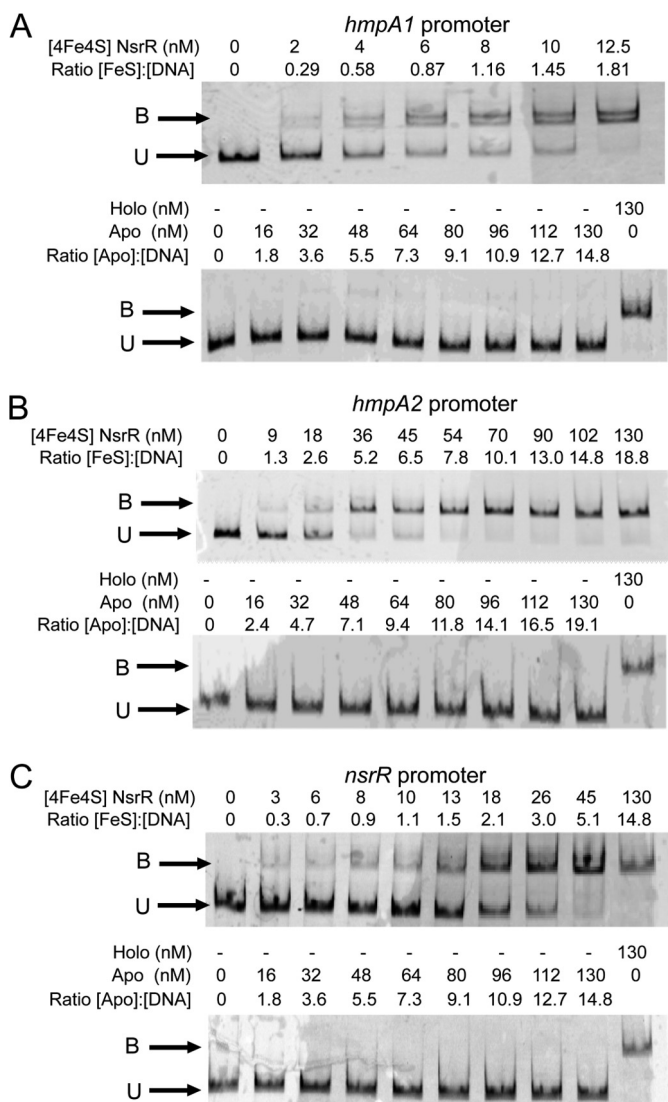
FIGURE 3. **Cluster-dependent DNA binding by [4Fe-4S] NsrR.** EMSAs using [4Fe-4S] or apo-NsrR (as indicated) and the *hmpA1* (*A*), *hmpA2* (*B*), and *nsrR* (*C*) promoters. Ratios of [4Fe-4S] NsrR to DNA are indicated. DNA concentration were 6.9 and 8.8 nM (*hmpA1* and *hmpA2*) or 8.8 and 8.8 nM (*nsrR*) for the [4Fe-4S] and apo-NsrR experiments, respectively. The binding buffer contained 10 mM Tris, 54 mM KCl, 0.3% (v/v) glycerol, 1.32 mM GSH, pH 7.5.

band observed at low levels of ScNsrR is not known, the data demonstrate tight binding. The nature of the binding species is not completely clear because the ScNsrR sample contains both [4Fe-4S] and apo-forms, so a sample consisting entirely of apo-NsrR was also investigated. No evidence of binding was observed (Fig. 3*A*), demonstrating that the [4Fe-4S] cluster form of ScNsrR is the DNA-binding form of the protein. We conclude that the binding interaction between ScNsrR and the *hmpA1* promoter is tight, with full binding observed at a level of ~2 [4Fe-4S] NsrR monomers per DNA. This is significantly tighter than previously reported for [2Fe-2S] NsrR, for which full binding of *hmpA1* DNA was not observed even with a several hundred-fold excess of protein (19). Data for the *hmpA2* and *nsrR* promoters are shown in Fig. 3, *B* and *C*, respectively. Binding of the [4Fe-4S] form was again observed, whereas the apo-form did not bind. For the *hmpA2* and *nsrR* promoters, full binding was observed at an excess of [4Fe-4S] NsrR over DNA

of 8 and 5, respectively, indicating that ScNsrR binds the *hmpA1* promoter most tightly, consistent with the enrichment seen in the ChIP-seq experiment.

ScNsrR Binds to an 11-bp Inverted Repeat Sequence—MEME analysis revealed that all three ScNsrR target promoters contain a DNA sequence that resembles the 11-bp inverted repeat structure of the known NsrR binding sites in *E. coli* and *B. subtilis* (13, 46). To confirm that ScNsrR binds specifically to these sites at the *hmpA1*, *hmpA2*, and *nsrR* promoters, DNase I footprinting experiments were performed using ^{32}P-labeled DNA fragments carrying each promoter. When the *nsrR* promoter fragment was incubated with ScNsrR and subjected to different DNase I digestion times the footprint was clearly visible (Fig. 4*A*), and this was confirmed in a separate experiment in which all three promoter fragments were incubated with increasing concentrations of ScNsrR before the addition of DNase I (Fig. 4*B*). The results clearly show a protected region covering the predicted binding site at each of the three promoters.

To probe important features of the binding site, additional EMSAs were performed. Deletion of the conserved AA and TT from either end of the *hmpA1* binding site, to make a truncated 19-bp site, abolished binding by ScNsrR, indicating that these conserved features are essential for recognition by ScNsrR (Fig. 4*C*). Similarly, substitution of all of the conserved A:T base pairs within the 23 bp site by C:G also abolished binding, suggesting that the unusual AT-rich features of the binding site are essential for recognition by ScNsrR (Fig. 4*C*). Probes carrying the experimentally verified NsrR binding sites from the *E. coli* and *B. subtilis hmpA* promoters were only very weakly bound by ScNsrR, indicating that the differences in DNA sequence are crucial for tight and specific binding of the NsrR proteins from these distantly related species (Fig. 4*C*).

[4Fe-4S] ScNsrR Binding to DNA Is Abolished by Reaction with NO—Exposure of [4Fe-4S] NsrR to a ~20-fold excess of NO resulted in loss of binding to all three NsrR-regulated gene promoter regions (see Fig. 5). Thus, the high affinity DNA binding exhibited by [4Fe-4S] NsrR is sensitive to NO, consistent with its role as an NO sensor. Further details of the [4Fe-4S] cluster nitrosylation reaction will be described elsewhere.

Identification of the Non-Cys Ligand in [4Fe-4S] ScNsrR—The resonance Raman spectrum of [4Fe-4S] ScNsrR indicated that an oxygenic ligand coordinates the cluster in addition to the three conserved Cys residues. Alignment of the characterized NsrR proteins from *E. coli*, *B. subtilis*, *Neisseria*, and *S. coelicolor* show that possible oxygenic ligands include Glu-85, Asp-123, and Asp-129, which are absolutely conserved, and Asp-96, Asp-113, and Glu-116, which are not conserved (17). A series of site-directed variants of NsrR was generated, in which carboxylate residues in these regions (which lie close to the three conserved Cys residues) were substituted by non-coordinating Ala. His-tagged variants E85A, D96A, D113A, E116A, D123A, and D129A ScNsrR were purified, and UV-visible spectra were recorded along with that of His-tagged wild-type ScNsrR (see Fig. 6*A*). Each protein was able to bind a cluster *in vivo* although at variable levels of incorporation and with somewhat variable absorbance properties. In particular, spectra due to E85A, D113A, and D123A NsrR were unusual in that absorp-
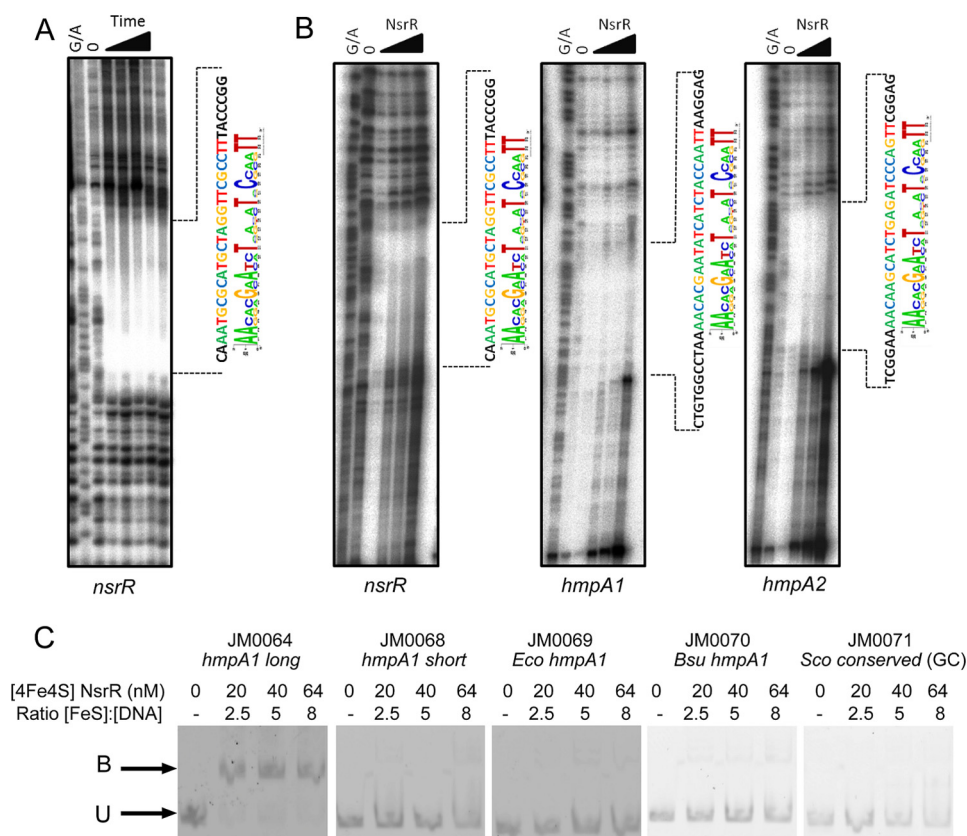
FIGURE 4. **DNase I footprinting and EMSA analysis of [4Fe-4S] NsrR binding to the *nsrR*, *hmpA1*, and *hmpA2* promoters.** *A*, footprint of NsrR bound to its own promoter. NsrR [4Fe-4S] at 2 µM was incubated with radiolabeled DNA for 0, 1, 2.5, 5, 7.5, and 10 min, respectively. *B*, footprints of increasing concentrations of NsrR bound to *nsrR*, *hmpA1*, and *hmpA2* promoters. The NsrR protein concentrations used were 0, 100, 250, 1000, and 2000 nM. *G/A*, Maxam and Gilbert sequence ladder. The regions protected by NsrR binding are indicated by *dotted lines*, and the sequence of the predicted binding site is shown *beside* the MEME-predicted consensus. *C*, EMSAs showing DNA probes bound (*B*) and unbound (*U*) by [4Fe-4S] NsrR. Probes used were JM0086, which has the conserved AA and TT removed from the ends of the binding site (*hmpA1 short*); JM0069, which contains the *E. coli hmpA1* binding sequence; JM0070, which contains the *B. subtilis hmpA1* binding sequence; and JM0071, in which the most conserved A/T base pairs have been changed to C/G. JM0064, containing the identified binding site (*hmpA1 long*) was included as a control.
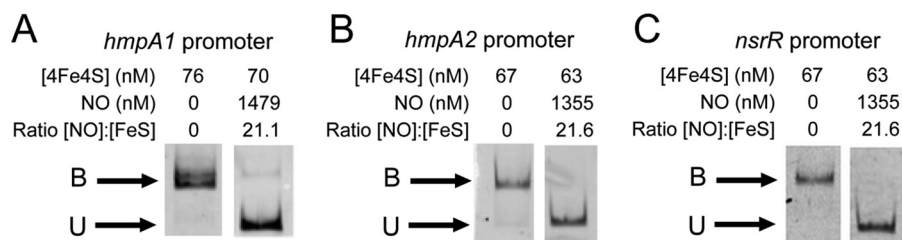
FIGURE 5. **Nitrosylation of NsrR [4Fe-4S] cluster abolishes DNA binding.** EMSAs using [4Fe-4S] before and after the addition of excess NO (as indicated) and the *hmpA1* (*A*), *hmpA2* (*B*), and *nsrR* (*C*) promoters. DNA concentrations were 10.6 nM (*hmpA1*), 5.9 nM (*hmpA2*), and 4.6 nM (*nsrR*). The binding buffer contained 10 mM Tris, 54 mM KCl, 0.3% (v/v) glycerol, 1.32 mM GSH, pH 7.5.

tion due to the cluster was shifted to higher energy (Fig. 6, *A* and *B*). The ability of the variant proteins to bind the *hmpA1* promoter region *in vitro* was investigated using EMSAs. The wild-type His-tagged protein, which fully bound DNA at a ratio of ~6 NsrR/DNA, exhibited somewhat weaker binding than was observed for the non-tagged wild-type ScNsrR protein (Figs. 3*A* and 6*C*). The DNA-binding behavior of (His-tagged) D96A, E116A, D123A, and D129A NsrR proteins was similar to that of the tagged wild-type protein, with only minor variation in apparent affinities. Only the E85A and D113A ScNsrR proteins showed behavior different from that of wild-type ScNsrR. For D113A ScNsrR, specific DNA binding was observed, but as protein concentration increased, additional binding (as evidenced

by a supershifted band) occurred, and, at the highest concentrations used here, aggregation occurred, with the protein and DNA remaining in the wells. Thus, although substitution of Asp-113 caused perturbations of the cluster environment, leading to aggregation at higher concentration, this variant was still able to bind specifically to DNA. This, alongside the fact that it is not conserved in other NsrR proteins, suggests that it is not a cluster ligand. In the case of E85A ScNsrR, however, there was no evidence of significant DNA binding, even at a ~17-fold excess of protein, suggesting a significant loss of DNA binding activity (Fig. 6*C*). This loss of activity, combined with the fact that Glu-85 is well conserved, suggests that it may be the fourth ligand for the Fe-S cluster in NsrR.

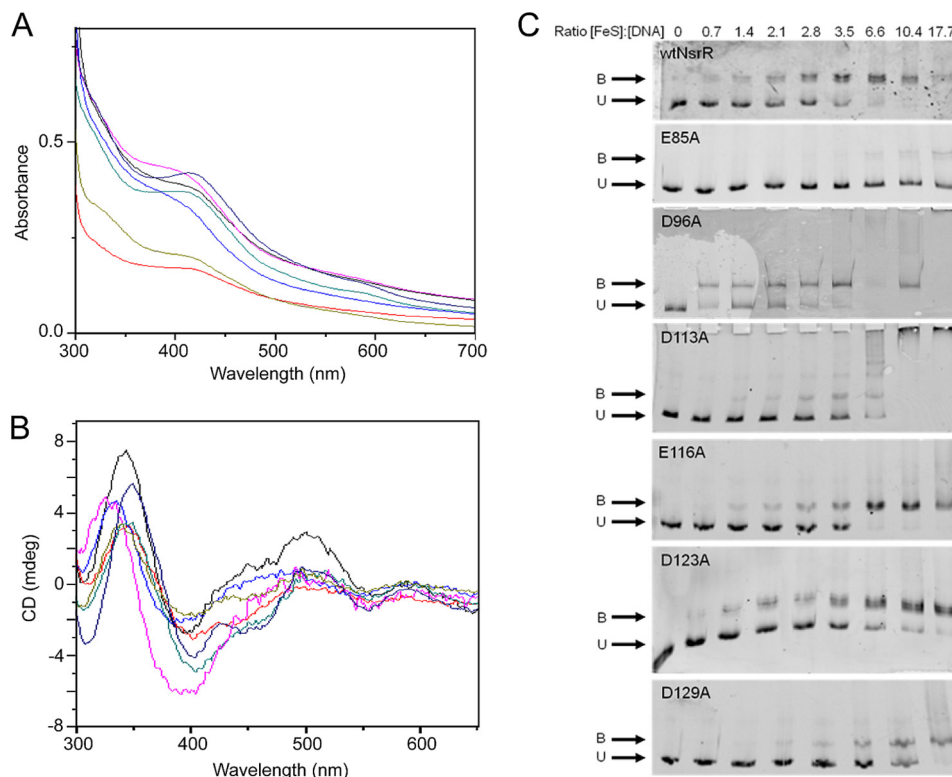FIGURE 6. **Spectroscopic and DNA binding properties of NsrR site-directed variants.** Shown are overlaid UV-visible absorbance (*A*) and circular dichroism (*B*) spectra of NsrR variant proteins: E85A (*black*), D96A (*red*), D113A (*royal blue*), E116A (*cyan*), D123A (*magenta*), D129A (*olive green*). The spectrum of wild-type NsrR (*navy blue*) was included for comparison. Spectra were not corrected for concentration (path length, 1 cm). The absorbance spectrum of E85A was magnified ×5 to enable comparison; the CD spectra of D96A and D123A were magnified ×3 and ×2, respectively. The buffer was 50 mM Tris, 800 mM NaCl, 5% (v/v) glycerol, pH 8.0. *C*, EMSAs for site-directed variants of NsrR, as indicated, with wild-type NsrR shown for comparison. The *hmpA1* promoter was used as the DNA probe at concentrations of 15.1 nM (E85A and D123A), 14.5 nM (D96A), and 7.6 nM (wild type NsrR, D113A, E116A, and D129A). Ratios of [FeS] NsrR to DNA are as shown except for D96A wells 7 (empty) and 9 ([FeS]/[DNA] 20.8).

Selective Interaction of Low Molecular Weight Thiols with [4Fe-4S] NsrR—It was recently reported that *B. subtilis* NsrR interacts with dithiothreitol (20), leading to the suggestion that low molecular weight thiols might be able to displace the non-Cys native ligand, resulting in all-thiolate coordination. To investigate whether ScNsrR also interacts with thiols, [4Fe-4S] ScNsrR was titrated with a range of low molecular weight thiols, including dithiothreitol, glutathione, and the more physiologically relevant mycothiol analogue *des-myo*-inositol mycothiol (dMSH), and visible CD spectra were recorded after each addition. Fig. 7*A* shows that dithiothreitol had a significant effect, with the major negative feature shifting to 374 nm with an isodichroic point at ~382 nm. A plot of $\Delta CD_{374\ nm}$ represents a binding isotherm, and fitting to a simple binding equation gave a K_d of 9.9 mM (Fig. 7*B*), indicating a relatively weak interaction. Stronger binding was observed for β-mercaptoethanol (K_d ~3.8 mM) and thioethane (K_d ~1.9 mM) (see Fig. 7*B*). Glutathione, cysteine, and thiosulfate had no effect on the CD spectrum, indicating that they do not bind to [4Fe-4S] ScNsrR. Mycothiol (1-D-*myo*-inosityl-2-(*N*-acetylcysteinyl)amido-2-deoxy-α-D-glucopyranoside), an abundant low molecular weight thiol found at millimolar concentrations in most actinomycetes (47), serves as the major thiol redox buffer for *S. coelicolor*. dMSH is an analogue of mycothiol, only lacking the inositol group. The addition of dMSH caused only very minor changes in the spectrum (Fig. 7*C*), which were different in form from those above. A plot of $\Delta CD_{374\ nm}$ over a physiologically relevant range

(0–2.5 mM) (Fig. 7*D*) shows no evidence of dMSH binding. The observed changes suggest that dMSH may increase the [4Fe-4S] cluster content of ScNsrR, perhaps through promoting the repair of minor components of damaged cluster.

Thiol-mediated Conversion of the NsrR [4Fe-4S] Cluster to a [2Fe-2S] Form—As isolated, [4Fe-4S] ScNsrR is unreactive toward O_2 (Fig. 7*E*, *inset*), with no loss of cluster observed up to 43 min after the addition of O_2 and only 8% cluster loss observed after 120 min (not shown). Very different behavior was observed in the presence of 5 mM dithiothreitol, however. The addition of O_2 resulted in a reddening of the color of the sample, and the UV-visible absorption and CD spectra of the resulting ScNsrR sample (Fig. 7, *E* and *F*) are very similar to those previously reported for [2Fe-2S] ScNsrR (19), consistent with the O_2- and thiolate-mediated conversion of the [4Fe-4S] to a [2Fe-2S] form. The time course of the reaction ($\Delta A_{474\ nm}$ *versus* time) (Fig. 7*E*, *inset*) shows that the conversion reaction was complete within 1 h. Similar experiments were conducted with β-mercaptoethanol and dMSH. In the presence of β-mercaptoethanol, [4Fe-4S] ScNsrR underwent a change similar to that observed with dithiothreitol, whereas dMSH had no effect on the O_2 stability of the [4Fe-4S] cluster (not shown).

To investigate this further, native MS was employed. The addition of dithiothreitol to [4Fe-4S] ScNsrR resulted in the series of mass spectra shown in Fig. 8*A*. Over a period of 30 min, the peak at 17,823 Da due to [4Fe-4S] ScNsrR decreased, whereas a new peak at 17,647 Da was observed to increase in
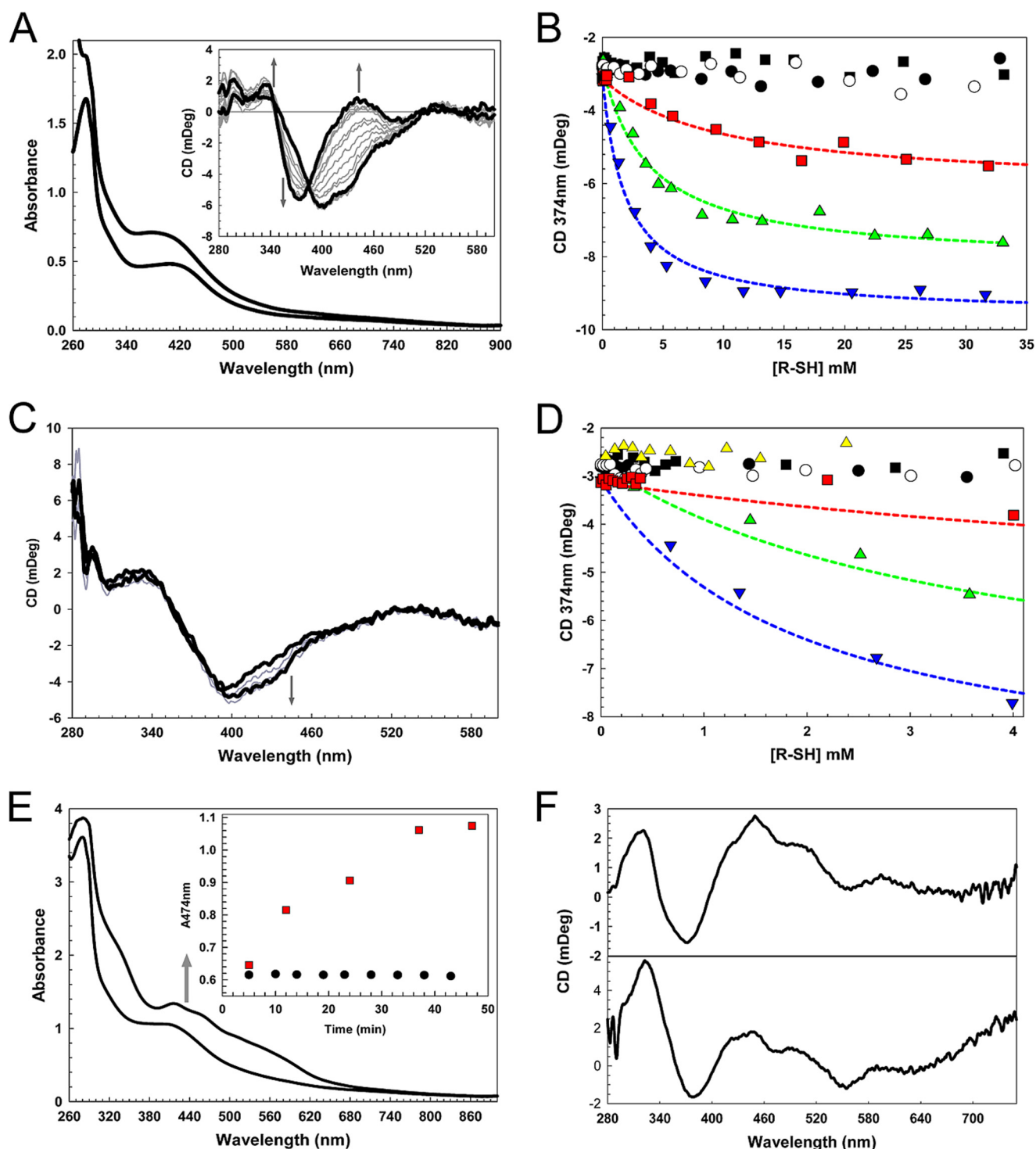
FIGURE 7. **Investigation of low molecular weight thiol binding to [4Fe-4S] NsrR and O_2-mediated cluster conversion.** *A*, UV-visible absorbance spectrum of 36.3 μM [4Fe-4S] NsrR in the presence (*black line*) and absence (*gray line*) of 35 mM DTT. *Inset*, CD spectra resulting from a titration of an identical sample of [4Fe-4S] NsrR with DTT up to 35 mM. *Arrows* indicate the direction of spectra changes. *B*, changes in the CD spectrum ($CD_{374\,nm}$) in response to glutathione (*open circles*), L-cysteine (*black circles*), thiosulfate (*black squares*), DTT (*red squares*; K_d = 9.9 mM), β-mercaptoethanol (*green triangles*; K_d = 3.8 mM), and thioethane (*blue triangles*; K_d = 1.9 mM). Fits to a simple binding equation (*dashed lines*) provide an estimate of the K_d for each thiol. *C*, CD spectra of [4Fe-4S] NsrR titrated with *N*-acetylcysteine-glucose amine (dMSH). Minor changes between 400 and 460 nm suggest that dMSH may repair damaged FeS clusters in NsrR. *D*, changes in the CD spectrum ($CD_{374\,nm}$) in response to dMSH (*yellow triangles*) over a physiologically relevant concentration range (the responses due to other thiols shown in *B* are also plotted for comparison. The buffer was 20 mM Tris, 20 mM MES, 20 mM Bistris propane, 100 mM NaCl, 5% (v/v) glycerol, pH 8.0. *E*, absorption spectrum of NsrR exposed to O_2 in the presence of 5 mM DTT for 5 and 47 min. The absorption spectrum is similar to that reported previously [2Fe-2S] NsrR (19). *Inset*, O_2-induced absorbance changes at 474 nm in the presence (*red squares*) and absence (*black circles*) of 5 mM DTT. The buffer was 20 mM Tris, 20 mM MES, 20 mM Bistris propane, 100 mM NaCl, 5% (v/v) glycerol, pH 8.0. *F*, CD spectra of O_2-modified NsrR (*top*) and [2Fe-2S] NsrR prepared as reported previously (19). The buffer was 50 mM Tris, 50 mM NaCl, 5 mM DTT, 5% (v/v) glycerol, pH 8.0.

⬱ASBMB

FIGURE 8. **Native MS analysis of O$_2$- and low molecular weight thiol-induced cluster conversion of [4Fe-4S] NsrR.** Shown are ESI-TOF MS spectra of [4Fe-4S] NsrR (7 μM) in the presence of O$_2$ ($\sim$220 μM) 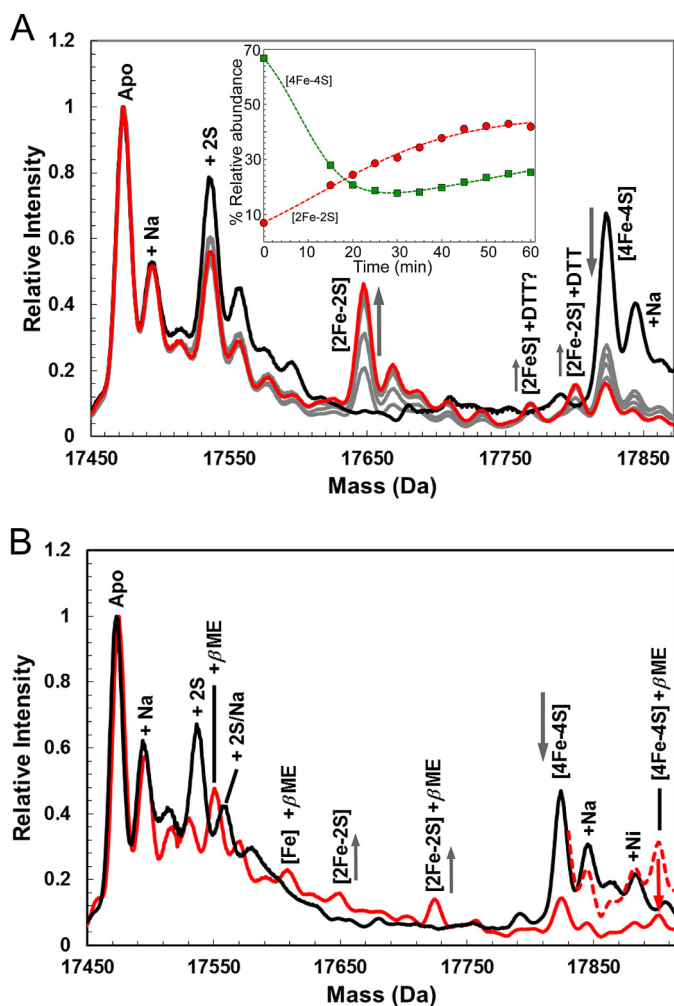and 5 mM DTT (*A*) and 1.1 M β-mercaptoethanol (*B*). Prior to the addition of thiol/O$_2$ (*black line*), no [2Fe-2S] clusters were observed. In *A*, mass spectra were recorded at 0 min (*black line*); 15, 30, 45, and 55 min (*gray lines*); and 65 min (*red line*) postexposure. Plots of relative intensity of [4Fe-4S] and [2Fe-2S] NsrR as a function of time are shown in the *inset*. *Trend lines* are drawn in. In *B*, mass spectra were recorded at 0 min (*black line*) and 15 min (*red line*) postexposure. Prior to the addition of β-mercaptoethanol/O$_2$ (*black line*), no [2Fe-2S] clusters were observed. After 15 min (*red line*, *dashed red line* multiplied $\times$3.5), β-mercaptoethanol adducts of [2Fe-2S] and [4Fe-4S] NsrR were observed. *m/z* spectra, recorded in the positive ion mode, were deconvoluted using Bruker Compass Data analysis software with the Maximum Entropy plugin. The buffer was 250 mM ammonium acetate, pH 8.0.

intensity. The new peak corresponds to ScNsrR with a [2Fe-2S] cluster bound to three deprotonated Cys residues (predicted mass = 17,474 − 3 + 176 = 17,647 Da). The small peak at 17,801 Da could be due to the [2Fe-2S] form bound by dithiothreitol (predicted 17,647 + 154 = 17,801 Da). Changes in relative intensity for the [4Fe-4S] and [2Fe-2S] forms are plotted as a function of time (Fig. 8*A*, *inset*). Similar experiments were conducted with β-mercaptoethanol, and similar effects were observed, with the [4Fe-4S] NsrR peak losing intensity and a [2Fe-2S] peak appearing (Fig. 8*B*). In this case, however, adducts of β-mercaptoethanol are more abundant, such that a β-mercaptoethanol-bound form of [2Fe-2S] ScNsrR, at 17,725 Da (predicted 17,647 + 78 = 17,725 Da) is more abundant than the [2Fe-2S] form. A β-mercaptoethanol-bound form of the
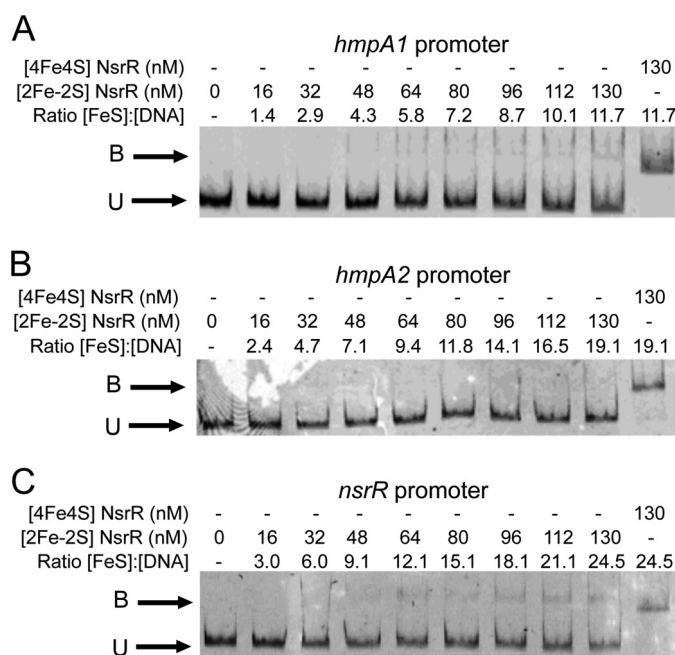
FIGURE 9. **DNA binding of [2Fe-2S] NsrR.** Shown are EMSAs using [2Fe-2S] NsrR (as indicated) and the *hmpA1* (11.1 nM) (*A*), *hmpA2* (6.8 nM) (*B*), and *nsrR* (5.3 nM) (*C*) promoters. Ratios of [FeS] NsrR to DNA are indicated. The binding buffer contained 10 mM Tris, 54 mM KCl, 0.3% (v/v) glycerol, 1.32 mM GSH, pH 7.5.

[4Fe-4S] species at 17,901 Da (predicted 17,823 + 78 = 17,901 Da) was also observed. It is not absolutely clear that these are due to β-mercaptoethanol bound to the cluster, however, because a β-mercaptoethanol adduct is also detected for the apoprotein. However, together with the dithiothreitol experiment, the data are consistent with thiol binding of the cluster. To determine whether [2Fe-2S] ScNsrR binds to the promoter regions of *hmpA1*, *hmpA2*, or *nsrR*, EMSA experiments were repeated using a [2Fe-2S] ScNsrR sample treated to remove all traces of residual (non-converted) [4Fe-4S] NsrR (see "Experimental Procedures." Fig. 9 shows that very little binding to *hmpA1* was observed, even at an excess of >10 [2Fe-2S] ScNsrR/DNA molecule. No evidence for binding to *hmpA2* was obtained even when [2Fe-2S] ScNsrR was present in 15-fold excess.

Overall, these data demonstrate that low molecular weight thiols that are able to bind to the cluster promote its reaction with O$_2$, resulting in conversion to a [2Fe-2S] form in which the thiol may remain bound. This form of ScNsrR does not bind significantly to the *hmpA1* and *hmpA2* promoters.

DISCUSSION

Anaerobic purification of *S. coelicolor* NsrR resulted in a cluster-bound form of the protein that is different from that reported previously (19). Here we have demonstrated that this form of the protein contains a [4Fe-4S]$^{2+}$ cluster and is a homodimer whether the cluster is present or not. The [4Fe-4S] form is stable to O$_2$, consistent with the fact that *S. coelicolor* is an obligate aerobe, and binds tightly in a cluster-dependent manner to an 11-bp inverted repeat sequence in the promoter regions of *hmpA1*, *hmpA2*, and *nsrR*.

The relationship between [4Fe-4S] ScNsrR and the previously reported [2Fe-2S] form (19) was initially unclear. We

noted that the resonance Raman spectrum of *B. subtilis* [4Fe-4S] NsrR was somewhat affected by the presence of dithiothreitol (with a decrease in the frequency of the symmetric bridging Fe-S stretching mode from 338 to 335 cm^{-1}), and its reaction with O_2 was markedly affected by the presence of dithiothreitol, resulting in a stabilization of the cluster and a mixture of [4Fe-4S] and [2Fe-2S] clusters (20). We have found that a number of low molecular weight thiols bind with low affinity to [4Fe-4S] NsrR, altering the spectroscopic properties of the cluster. All of the thiols that were found to bind (dithiothreitol, β-mercaptoethanol, and thioethane) are simple organic molecules with one or more thiol groups and no net charge. A number of thiols tested were found not to bind, and these were either more complex molecules with large substituents in addition to the thiol group (glutathione and dMSH) or were charged (thiosulfate). Thus, electrostatic and/or steric effects appear to be important for access of the thiol to the cluster site. Where binding was observed, the thiol was most likely able to compete for one of the iron sites, and it is also likely that this is the one that is not already coordinated by a thiol Cys. In those cases, the binding affinities reported here are better described as competition exchange constants rather than absolute binding constants, reflecting the competition between the natural ligand and the low molecular weight thiol.

Binding of an exogenous thiol to [4Fe-4S] NsrR was found to drastically reduce the O_2 stability of the cluster, leading to rapid and stoichiometric conversion to a [2Fe-2S] form. These data explain why NsrR was previously characterized as a [2Fe-2S] cluster protein; in the original report, the protein was purified in the presence of dithiothreitol under aerobic conditions (19). Here, little or no binding of the [2Fe-2S] form to *hmpA1* and *hmpA2* promoters was detected. These observations appear to be inconsistent with the previous report of DNA binding by [2Fe-2S] NsrR. However, in those experiments, a several hundred-fold excess of [2Fe-2S] NsrR was present; in the current experiments, stoichiometric or near stoichiometric binding was observed for [4Fe-4S] NsrR binding to *hmpA1*, *hmpA2*, and *nsrR* promoters.

Thus, NsrR can accommodate either [4Fe-4S] or [2Fe-2S] clusters, and the O_2-mediated conversion from the [4Fe-4S] to the [2Fe-2S] form is dependent on the presence of a coordinating low molecular weight thiol. Evidence from absorbance spectroscopy (where there was an increase in absorbance observed upon cluster conversion consistent with increased iron-thiolate coordination) and native MS (where [2Fe-2S] NsrR-thiol adducts were directly observed) suggests that the thiol remains bound to the [2Fe-2S] cluster and probably stabilizes it against further O_2-mediated breakdown. Importantly, the physiologically relevant thiols L-cysteine and thiosulfate and the mycothiol analogue dMSH (48, 49) did not promote [4Fe-4S] to [2Fe-2S] cluster conversion. This suggests that cluster conversion is a result of *in vitro* protein handling, and we conclude that the [4Fe-4S] form of NsrR is the active form of the protein in the cytoplasm of aerobically growing *S. coelicolor* cells. However, given the facile nature of the cluster conversion reaction, albeit under specific conditions, we cannot rule out that this could have physiological significance in *Streptomyces* or other organisms. In the case of *S. coelicolor*, this would

involve regulation of genes different from those identified here because we found no evidence of DNA binding for the [2Fe-2S] form.

Reaction of [4Fe-4S] NsrR with NO led to the loss of DNA binding, consistent with NsrR acting as an NO sensor. The data indicate that NsrR functions as a repressor under normal conditions. In the presence of NO, a conformational change must occur that disrupts DNA binding, resulting in derepression of genes encoding NO-detoxifying enzymes. The nature of the reaction with NO is currently under investigation, but it may be similar to the nitrosylation reactions of other [4Fe-4S] regulatory proteins, involving a rapid and complex reaction with up to eight NO molecules per cluster (10, 50).

NsrR proteins contain three conserved Cys residues that coordinate the cluster. Resonance Raman spectroscopy indicated that the fourth cluster ligand is oxygenic, and studies of site-directed variants highlighted two proteins with unusual properties (E85A and D113A), and of these only E85A showed no evidence of DNA binding. Furthermore, Glu-85 is totally conserved in experimentally verified NsrR proteins (17). Although we are not aware of an unambiguous example of cluster coordination by three Cys and one Glu residue, several instances of [4Fe-4S] clusters coordinated by three Cys residues and one Asp are known (*e.g. Pyrococcus furiosus* ferredoxin (43, 51), *Desulfovibrio africanus* ferredoxin III (52), and *B. subtilis* FNR (53)). Therefore, Glu-85 is a reasonable candidate for the fourth ligand. Some caution is required, however, because substitutions of non-coordinating residues could *indirectly* affect the cluster environment and/or DNA binding properties of the protein. We note that the yield of variant E85A was much lower than for the other variants, and this could be a result of impaired stability. Therefore, although our data point to Glu-85 being the fourth cluster ligand, further confirmation is needed before a definitive assignment can be made, and this may require a high resolution structure.

Another well characterized member of the Rrf2 family of regulators, IscR, has been shown to bind a [2Fe-2S] cluster (41, 54). Although approximately 30% identity exists between IscR and NsrR, and the three Cys residues that coordinate the cluster are conserved, the spacing between them is not, and the fourth ligand to the cluster is different. For IscR, this was recently shown to be His-107 (41), a residue that is not conserved in NsrR proteins. The equivalent residue of Glu-85 in IscR is Asp-84, but substitution of this residue had no effect on IscR activity (41).

As clearly demonstrated here, NsrR appears to have inherent flexibility in its cluster-binding site, and IscR might share this flexibility. The variations in the nature and precise arrangements of coordinating ligands are likely to be important in determining the balance of stabilities between the different cluster forms.

Although IscR is arguably the best characterized Rrf2 protein, NsrR is the most widely conserved in the bacterial kingdom and has been characterized not just in Gram-negative gammaproteobacteria like *E. coli* K12, *E. coli* O157:H7, and *Salmonella* (13, 55, 56) but also in the Gram-negative betaproteobacteria *N. gonorrhoeae* and *Neisseria meningitidis* (18, 57), in the low GC Gram-positive Firmicute *B. subtilis*, and in the high

A NsrR auto-binding sites

```
S. coelicolor    ACCCCGGGTAAAGGCGAACCTAGCATGCGCATTTGATAGCGTCCTGGTATG
S. scabies       ACTCCCTCAAAATAAGTATCTGAGATGCAATATTGCTAGGCTGCTGCCATG
S. clavuligerus  GCGACTTTTAAGTGGTATCTAAGATTCGCATTAACCTAAGGTGGCGTTATG
S. albus S4      CGAGAGCCTTAAGTTGCATCGCATGTGCAAGTTTCTAGGCTGCGGGCCATG
S. avermitilis   TTCGCCCGATAATTTGAACCGTGGATGCGCATTATTTACGCTGGCGCCATG
S. griseus       CGGACCACTTAAGTTGCATCAAGGATGCAACTTATCTACGGTGGACCCATG
```

Aligned:

```
S. coelicolor    AAGGCGAACCTAGCATGCGCATT
S. scabies       AATAAGTATCTGAGATGCAATAT
S. clavuligerus  AAGTGGTATCTAAGATTCGCATT
S. albus S4      AAGTTGCATCGCATGTGCAAGTT
S. avermitilis   AATTTGAACCGTGGATGCGCATT
S. griseus       AAGTTGCATCAAGGATGCAACTT
```
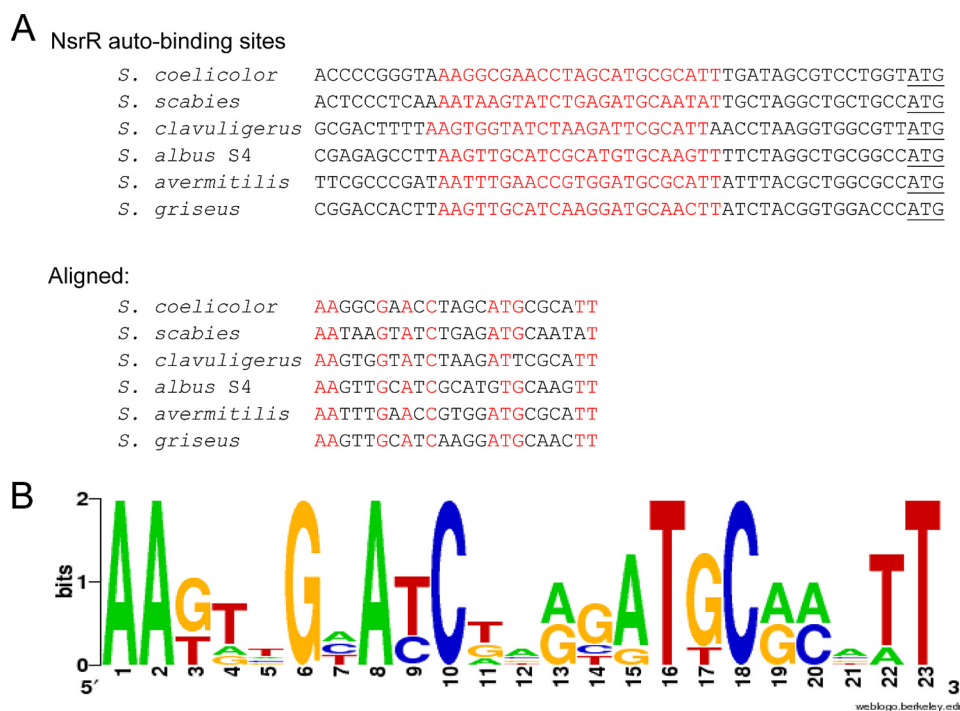
B

FIGURE 10. **Alignment of *nsrR* promoters.** *A*, alignment of *nsrR* promoters from six *Streptomyces* species (taken from StrepDB), revealing a conserved NsrR binding site at each. *B*, WebLogo generated by aligning these putative NsrR binding sites.

GC Gram-positive actinomycete *S. coelicolor* (19). In all four branches of the bacteria, NsrR senses NO via an Fe-S cluster, and its primary function is to detoxify NO. *Neisseria* NsrR acts solely as a repressor and has a relatively small regulon of five genes, including *nsrR* (18, 57). It is somewhat unusual because it controls NO metabolism not via HmpA but by coordinating expression of the nitrite reductase (*aniA*) and NO reductase (*norB*) genes such that nitrite can be converted to nitrous oxide without a toxic build up of the intermediate NO. It has been argued that *Neisseria* strains are undergoing host adaptation by losing the ability to denitrify, through loss of *aniA*, and evolving into NO-tolerant aerobes (58). *Neisseria* NsrR also controls expression of *mobB*, which encodes an enzyme involved in molybdenum metabolism, and *dnrN*, which encodes a protein involved in repairing Fe-S cluster proteins damaged by nitrosative or oxidative stress (57). In *E. coli* and *B. subtilis*, NsrR regulates NO detoxification by controlling the production of HmpA. However, *E. coli* K12 NsrR regulates >60 target genes, and *B. subtilis* NsrR has a regulon of ~35 target genes, many of which do not have an obvious role in NO metabolism (13, 46). *E. coli* and *B. subtilis* NsrR proteins regulate many of these target genes by binding to half-sites, but we could not detect any binding of ScNsrR to EMSA probes carrying artificial half-sites (not shown). This is consistent with the ChIP-seq analysis in which all three experimentally validated targets have full 11-bp inverted repeat binding sites. Intriguingly, in *E. coli* O157:H7, NsrR binds to full inverted repeat sequences at the promoters of the locus of enterocyte effacement LEE1 and LEE4 genes, and a half-site at the LEE5 promoter, all of which are on a chromosomal pathogenicity island. Bound NsrR activates these promoters by recruiting RNA polymerase, and activation is abolished by the addition of the NO releaser Nor-4 to cultures (56).

To our knowledge, this is the only example of NsrR directly activating gene expression because all other reports describe it as a transcriptional repressor.

S. coelicolor NsrR has the smallest regulon reported to date and appears to be unique (thus far) in that its function is specialized to NO detoxification. Maintenance of an Fe-S-containing regulator to control its own expression plus that of two *hmpA* genes suggests that NO is a significant threat to *S. coelicolor* in its natural habitat. Of the complete genomes in the *Streptomyces* genome database, StrepDB, NsrR is conserved in all except for *Streptomyces venezuelae*, which also lacks an HmpA homologue but encodes a bacterial NO synthase enzyme. *S. venezuelae* may have lost NsrR-HmpA because it interferes with endogenous NO production via bacterial NO synthase. *Streptomyces scabies* also encodes bacterial NO synthase and has NsrR-HmpA, but the production of NO in *S. scabies* is tightly coupled to the biosynthesis of the phytotoxin thaxtomin (59). In the streptomycetes that encode NsrR, all of the *nsrR* genes contain a full NsrR binding site upstream of the translational start codon (Fig. 10), suggesting that autoregulation is a conserved feature.

REFERENCES

1. Hess, D. T., and Stamler, J. S. (2012) Regulation by *S*-nitrosylation of protein post-translational modification. *J. Biol. Chem.* **287,** 4411–4418
2. Seth, D., Hausladen, A., Wang, Y. J., and Stamler, J. S. (2012) Endogenous protein *S*-nitrosylation in *E. coli*: regulation by OxyR. *Science* **336,**

470–473

3. Kwon, Y. M., and Weiss, B. (2009) Production of 3-nitrosoindole derivatives by *Escherichia coli* during anaerobic growth. *J. Bacteriol.* **191,** 5369–5376

4. Weiss, B. (2006) Evidence for mutagenesis by nitric oxide during nitrate metabolism in *Escherichia coli. J. Bacteriol.* **188,** 829–833

5. Crack, J. C., Green, J., Thomson, A. J., and Le Brun, N. E. (2012) Iron-sulfur cluster sensor-regulators. *Curr. Opin. Chem. Biol.* **16,** 35–44

6. Bruckdorfer, R. (2005) The basics about nitric oxide. *Mol. Aspects Med.* **26,** 3–31

7. Vine, C. E., and Cole, J. A. (2011) Unresolved sources, sinks, and pathways for the recovery of enteric bacteria from nitrosative stress. *FEMS Microbiol. Lett.* **325,** 99–107

8. Sudhamsu, J., and Crane, B. R. (2009) Bacterial nitric oxide synthases: what are they good for? *Trends Microbiol.* **17,** 212–218

9. Demple, B. (2002) Signal transduction by nitric oxide in cellular stress responses. *Mol. Cell Biochem.* **234,** 11–18

10. Crack, J. C., Stapleton, M. R., Green, J., Thomson, A. J., and Le Brun, N. E. (2013) Mechanism of [4Fe-4S](Cys)$_4$ cluster nitrosylation is conserved among NO-responsive regulators. *J. Biol. Chem.* **288,** 11492–11502

11. Hutchings, M. I., Crack, J. C., Shearer, N., Thompson, B. J., Thomson, A. J., and Spiro, S. (2002) Transcription factor FnrP from *Paracoccus denitrificans* contains an iron-sulfur cluster and is activated by anoxia: identification of essential cysteine residues. *J. Bacteriol.* **184,** 503–508

12. D'Autréaux, B., Tucker, N. P., Dixon, R., and Spiro, S. (2005) A non-haem iron centre in the transcription factor NorR senses nitric oxide. *Nature* **437,** 769–772

13. Partridge, J. D., Bodenmiller, D. M., Humphrys, M. S., and Spiro, S. (2009) NsrR targets in the *Escherichia coli* genome: new insights into DNA sequence requirements for binding and a role for NsrR in the regulation of motility. *Mol. Microbiol.* **73,** 680–694

14. Stevanin, T. M., Read, R. C., and Poole, R. K. (2007) The *hmp* gene encoding the NO-inducible flavohaemoglobin in *Escherichia coli* confers a protective advantage in resisting killing within macrophages, but not *in vitro*: links with swarming motility. *Gene* **398,** 62–68

15. Hutchings, M. I., Mandhana, N., and Spiro, S. (2002) The NorR protein of *Escherichia coli* activates expression of the flavorubredoxin gene *norV* in response to reactive nitrogen species. *J. Bacteriol.* **184,** 4640–4643

16. Rajagopalan, S., Teter, S. J., Zwart, P. H., Brennan, R. G., Phillips, K. J., and Kiley, P. J. (2013) Studies of IscR reveal a unique mechanism for metal-dependent regulation of DNA binding specificity. *Nat. Struct. Mol. Biol.* **20,** 740–747

17. Tucker, N. P., Le Brun, N. E., Dixon, R., and Hutchings, M. I. (2010) There's NO stopping NsrR, a global regulator of the bacterial NO stress response. *Trends Microbiol.* **18,** 149–156

18. Isabella, V. M., Lapek, J. D., Jr., Kennedy, E. M., and Clark, V. L. (2009) Functional analysis of NsrR, a nitric oxide-sensing Rrf2 repressor in *Neisseria gonorrhoeae. Mol. Microbiol.* **71,** 227–239

19. Tucker, N. P., Hicks, M. G., Clarke, T. A., Crack, J. C., Chandra, G., Le Brun, N. E., Dixon, R., and Hutchings, M. I. (2008) The transcriptional repressor protein NsrR senses nitric oxide directly via a [2Fe-2S] cluster. *PLoS One* **3,** e3623

20. Yukl, E. T., Elbaz, M. A., Nakano, M. M., and Moënne-Loccoz, P. (2008) Transcription factor NsrR from *Bacillus subtilis* senses nitric oxide with a 4Fe-4S cluster. *Biochemistry* **47,** 13084–13092

21. Kommineni, S., Yukl, E., Hayashi, T., Delepine, J., Geng, H., Moënne-Loccoz, P., and Nakano, M. M. (2010) Nitric oxide-sensitive and -insensitive interaction of *Bacillus subtilis* NsrR with a ResDE-controlled promoter. *Mol. Microbiol.* **78,** 1280–1293

22. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000) *Practical Streptomyces Genetics*, p. 412, John Innes Foundation, Norwich

23. Gregory, M. A., Till, R., and Smith, M. C. (2003) Integration site for Streptomyces phage phiBT1 and development of site-specific integrating vectors. *J. Bacteriol.* **185,** 5320–5323

24. Kibbe, W. A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35,** W43–W46

25. Al-Bassam, M. M., Bibb, M. J., Bush, M. J., Chandra, G., and Buttner, M. J.

26. (2014) Response regulator heterodimer formation controls a key stage in *Streptomyces* development. *PLoS Genet.* **10,** e1004554

26. Fernández-Martínez, L. T., Del Sol, R., Evans, M. C., Fielding, S., Herron, P. R., Chandra, G., and Dyson, P. J. (2011) A transposon insertion single-gene knockout library and new ordered cosmid library for the model organism *Streptomyces coelicolor* A3(2). *Antonie Van Leeuwenhoek* **99,** 515–522

27. Redenbach, M., Kieser, H. M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H., and Hopwood, D. A. (1996) A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol. Microbiol.* **21,** 77–96

28. Crack, J. C., Green, J., Thomson, A. J., and Le Brun, N. E. (2014) Techniques for the production, isolation, and analysis of iron-sulfur proteins. *Methods Mol. Biol.* **1122,** 33–48

29. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150,** 76–85

30. Crack, J. C., Gaskell, A. A., Green, J., Cheesman, M. R., Le Brun, N. E., and Thomson, A. J. (2008) Influence of the environment on the [4Fe-4S]$^{2+}$ to [2Fe-2S]$^{2+}$ cluster switch in the transcriptional regulator FNR. *J. Am. Chem. Soc.* **130,** 1749–1758

31. Crack, J. C., den Hengst, C. D., Jakimowicz, P., Subramanian, S., Johnson, M. K., Buttner, M. J., Thomson, A. J., and Le Brun, N. E. (2009) Characterization of [4Fe-4S]-containing and cluster-free forms of *Streptomyces* WhiD. *Biochemistry* **48,** 12252–12264

32. Stewart, M. J., Jothivasan, V. K., Rowan, A. S., Wagg, J., and Hamilton, C. J. (2008) Mycothiol disulfide reductase: solid phase synthesis and evaluation of alternative substrate analogues. *Org. Biomol. Chem.* **6,** 385–390

33. Riddles, P. W., Blakeley, R. L., and Zerner, B. (1983) Reassessment of Ellman's reagent. *Methods Enzymol.* **91,** 49–60

34. Lagarec, K., and Rancourt, D. C. (1998) *Recoil: Mössbauer spectral analysis software for Windows*, version 1.0, University of Ottawa, Canada

35. Rodionov, D. A., Dubchak, I. L., Arkin, A. P., Alm, E. J., and Gelfand, M. S. (2005) Dissimilatory metabolism of nitrogen oxides in bacteria: comparative reconstruction of transcriptional networks. *PLoS Comput. Biol.* **1,** e55

36. Stephens, P. J., Thomson, A. J., Dunn, J. B., Keiderling, T. A., Rawlings, J., Rao, K. K., and Hall, D. O. (1978) Circular dichroism and magnetic circular dichroism of iron-sulfur proteins. *Biochemistry* **17,** 4770–4778

37. Beinert, H., Holm, R. H., and Münck, E. (1997) Iron-sulfur clusters: Nature's modular, multipurpose structures. *Science* **277,** 653–659

38. Dickson, D. P. E., Johnson, C. E., Thompson, C. L., Cammack, R., Evans, M. C. W., Hall, D. O., Rao, K. K., and Weser, U. (1974) Mössbauer effect studies on the four iron centres of two iron-sulphur proteins. *J. Phys. Colloques* 10.1051/jphyscol:1974659

39. Cicchillo, R. M., Lee, K. H., Baleanu-Gogonea, C., Nesbitt, N. M., Krebs, C., and Booker, S. J. (2004) *Escherichia coli* lipoyl synthase binds two distinct [4Fe-4S] clusters per polypeptide. *Biochemistry* **43,** 11770–11781

40. Hernández, H. L., Pierrel, F., Elleingand, E., García-Serres, R., Huynh, B. H., Johnson, M. K., Fontecave, M., and Atta, M. (2007) MiaB, a bifunctional radical-*S*-adenosylmethionine enzyme involved in the thiolation and methylation of tRNA, contains two essential [4Fe-4S] clusters. *Biochemistry* **46,** 5140–5147

41. Fleischhacker, A. S., Stubna, A., Hsueh, K. L., Guo, Y., Teter, S. J., Rose, J. C., Brunold, T. C., Markley, J. L., Münck, E., and Kiley, P. J. (2012) Characterization of the [2Fe-2S] cluster of *Escherichia coli* transcription factor IscR. *Biochemistry* **51,** 4453–4462

42. Czernuszewicz, R. S., Macor, K. A., Johnson, M. K., Gewirth, A., and Spiro, T. G. (1987) Vibrational-mode structure and symmetry in proteins and analogs containing Fe$_4$S$_4$ clusters: resonance Raman evidence for different degrees of distortion in HiPIP and Ferredoxin. *J. Am. Chem. Soc.* **109,** 7178–7187

43. Conover, R. C., Kowal, A. T., Fu, W. G., Park, J. B., Aono, S., Adams, M. W., and Johnson, M. K. (1990) Spectroscopic characterization of the novel iron-sulfur cluster in *Pyrococcus furiosus* ferredoxin. *J. Biol. Chem.* **265,** 8533–8541

44. Brereton, P. S., Duderstadt, R. E., Staples, C. R., Johnson, M. K., and Ad-

ams, M. W. (1999) Effect of serinate ligation at each of the iron sites of the [Fe$_4$S$_4$] cluster of *Pyrococcus furiosus* ferredoxin on the redox, spectroscopic, and biological properties. *Biochemistry* **38,** 10594–10605

45. Zhang, B., Crack, J. C., Subramanian, S., Green, J., Thomson, A. J., Le Brun, N. E., and Johnson, M. K. (2012) Reversible cycling between cysteine persulfide-ligated [2Fe-2S] and cysteine-ligated [4Fe-4S] clusters in the FNR regulatory protein. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 15734–15739

46. Nakano, M. M., Geng, H., Nakano, S., and Kobayashi, K. (2006) The nitric oxide-responsive regulator NsrR controls ResDE-dependent gene expression. *J. Bacteriol.* **188,** 5878–5887

47. Jothivasan, V. K., and Hamilton, C. J. (2008) Mycothiol: synthesis, biosynthesis and biological functions of the major low molecular weight thiol in actinomycetes. *Nat. Prod. Rep.* **25,** 1091–1117

48. Newton, G. L., Bewley, C. A., Dwyer, T. J., Horn, R., Aharonowitz, Y., Cohen, G., Davies, J., Faulkner, D. J., and Fahey, R. C. (1995) The structure of U17 isolated from *Streptomyces clavuligerus* and its properties as an antioxidant thiol. *Eur. J. Biochem.* **230,** 821–825

49. Newton, G. L., Fahey, R. C., Cohen, G., and Aharonowitz, Y. (1993) Low molecular weight thiols in streptomycetes and their potential role as antioxidants. *J. Bacteriol.* **175,** 2734–2742

50. Crack, J. C., Smith, L. J., Stapleton, M. R., Peck, J., Watmough, N. J., Buttner, M. J., Buxton, R. S., Green, J., Oganesyan, V. S., Thomson, A. J., and Le Brun, N. E. (2011) Mechanistic insight into the nitrosylation of the [4Fe-4S] cluster of WhiB-like proteins, *J. Am. Chem. Soc.* **133,** 1112–1121

51. Calzolai, L., Gorst, C. M., Zhao, Z. H., Teng, Q., Adams, M. W., and La Mar, G. N. (1995) ^{1}H NMR investigation of the electronic and molecular structure of the four-iron cluster ferredoxin from the hyperthermophile *Pyrococcus furiosus*: identification of Asp14 as a cluster ligand in each of the four redox states. *Biochemistry* **34,** 11373–11384

52. George, S. J., Armstrong, F. A., Hatchikian, E. C., and Thomson, A. J. (1989) Electrochemical and spectroscopic characterization of the conversion of the 7Fe into the 8Fe form of ferredoxin III from *Desulfovibrio africanus*: Identification of a [4Fe-4S] cluster with one non-cysteine ligand. *Biochem. J.* **264,** 275–284

53. Gruner, I., Frädrich, C., Böttger, L. H., Trautwein, A. X., Jahn, D., and Härtig, E. (2011) Aspartate 141 is the fourth ligand of the oxygen-sensing [4Fe-4S]$^{2+}$ cluster of *Bacillus subtilis* transcriptional regulator Fnr. *J. Biol. Chem.* **286,** 2017–2021

54. Schwartz, C. J., Giel, J. L., Patschkowski, T., Luther, C., Ruzicka, F. J., Beinert, H., and Kiley, P. J. (2001) IscR, an Fe-S cluster-containing transcription factor, represses expression of *Escherichia coli* genes encoding Fe-S cluster assembly proteins. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 14895–14900

55. Karlinsey, J. E., Bang, I. S., Becker, L. A., Frawley, E. R., Porwollik, S., Robbins, H. F., Thomas, V. C., Urbano, R., McClelland, M., and Fang, F. C. (2012) The NsrR regulon in nitrosative stress resistance of *Salmonella enterica* serovar *Typhimurium*. *Mol. Microbiol.* **85,** 1179–1193

56. Branchu, P., Matrat, S., Vareille, M., Garrivier, A., Durand, A., Crépin, S., Harel, J., Jubelin, G., and Gobert, A. P. (2014) NsrR, GadE, and GadX interplay in repressing expression of the *Escherichia coli* O157:H7 LEE pathogenicity island in response to nitric oxide. *PLoS Pathog.* **10,** e1003874

57. Heurlier, K., Thomson, M. J., Aziz, N., and Moir, J. W. (2008) The nitric oxide (NO)-sensing repressor NsrR of *Neisseria meningitidis* has a compact regulon of genes involved in NO synthesis and detoxification. *J. Bacteriol.* **190,** 2488–2495

58. Moir, J. W. (2011) A snapshot of a pathogenic bacterium mid-evolution: *Neisseria meningitidis* is becoming a nitric oxide-tolerant aerobe. *Biochem. Soc. Trans.* **39,** 1890–1894

59. Kers, J. A., Wach, M. J., Krasnoff, S. B., Widom, J., Cameron, K. D., Bukhalid, R. A., Gibson, D. M., Crane, B. R., and Loria, R. (2004) Nitration of a peptide phytotoxin by bacterial nitric oxide synthase. *Nature* **429,** 79–82

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

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

62. Cervantes, S., Bunnik, E. M., Saraf, A., Conner, C. M., Escalante, A., Sardiu, M. E., Ponts, N., Prudhomme, J., Florens, L., and Le Roch, K. G. (2014) The multifunctional autophagy pathway in the human malaria parasite, *Plasmodium falciparum*. *Autophagy* **10,** 80–92

NsrR from *Streptomyces coelicolor* Is a Nitric Oxide-sensing [4Fe-4S] Cluster Protein with a Specialized Regulatory Function

Jason C. Crack, John Munnoch, Erin L. Dodd, Felicity Knowles, Mahmoud M. Al Bassam, Saeed Kamali, Ashley A. Holland, Stephen P. Cramer, Chris J. Hamilton, Michael K. Johnson, Andrew J. Thomson, Matthew I. Hutchings and Nick E. Le Brun

Access the most updated version of this article at doi: 10.1074/jbc.M115.643072

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 60 references, 21 of which can be accessed free at
http://www.jbc.org/content/290/20/12689.full.html#ref-list-1

⌘ *Author's Choice*

Differentiated, Promoter-specific Response of [4Fe-4S] NsrR DNA Binding to Reaction with Nitric Oxide*

Jason C. Crack[‡]**, Dimitri A. Svistunenko**[§]**, John Munnoch**[¶]**, Andrew J. Thomson**[‡]**, Matthew I. Hutchings**[¶]**, and Nick E. Le Brun**[‡1]

From the [‡]*Centre for Molecular and Structural Biochemistry, School of Chemistry, and the* [¶]*School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ and the* [§]*School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom*

NsrR is an iron-sulfur cluster protein that regulates the nitric oxide (NO) stress response of many bacteria. NsrR from *Streptomyces coelicolor* regulates its own expression and that of only two other genes, *hmpA1* and *hmpA2*, which encode HmpA enzymes predicted to detoxify NO. NsrR binds promoter DNA with high affinity only when coordinating a [4Fe-4S] cluster. Here we show that reaction of [4Fe-4S] NsrR with NO affects DNA binding differently depending on the gene promoter. Binding to the *hmpA2* promoter was abolished at ~2 NO per cluster, although for the *hmpA1* and *nsrR* promoters, ~4 and ~8 NO molecules, respectively, were required to abolish DNA binding. Spectroscopic and kinetic studies of the NO reaction revealed a rapid, multi-phase, non-concerted process involving up to 8–10 NO molecules per cluster, leading to the formation of several iron-nitrosyl species. A distinct intermediate was observed at ~2 NO per cluster, along with two further intermediates at ~4 and ~6 NO. The NsrR nitrosylation reaction was not significantly affected by DNA binding. These results show that NsrR regulates different promoters in response to different concentrations of NO. Spectroscopic evidence indicates that this is achieved by different NO-FeS complexes.

The gaseous, lipophilic molecule nitric oxide (NO) is an important signaling molecule in animals and there is growing evidence that it also has a signaling role in bacteria (1). At higher concentrations (micromolar) NO is a cytotoxin, a property exploited by the innate immune response of eukaryotes to infection by pathogenic organisms. The toxicity of NO is conferred by its reactivity toward DNA (nitrosative DNA damage (2)) and proteins (*e.g. S*-nitrosation (3) and *N*-nitrosation (4)) and protein metal cofactors, such as iron-sulfur (FeS) clusters (5), which are important for many cellular functions (6). The generation of NO in the presence of superoxide can also lead to the formation of peroxynitrite, leading to toxic effects (7).

The toxicity of NO is exploited by mammalian macrophages in their response to infection by pathogenic bacteria (8). The ability to sense and respond to high concentrations of NO is therefore a key component of stress response mechanisms of pathogenic organisms (9). Detoxification of NO is also important in many non-pathogenic organisms (10). For example, NO can be generated endogenously at significant concentrations in bacterial cells that are respiring anaerobically using nitrate/nitrite as terminal electron acceptors (11, 12) and NO is generated via the activity of NO synthases in some Gram-positive soil bacteria (13).

NsrR has been identified as a regulator of the NO stress response in a number of bacteria, including *Escherichia coli* (14) *Bacillus subtilis* (15) and pathogens such as *Neisseria gonorrhoeae* (16). In most of the organisms investigated to date, NsrR is a global regulator, controlling a complex network of genes, only some of which are directly related to NO detoxification. In the soil bacterium *Streptomyces coelicolor*, however, NsrR has a more specialized function, regulating only the *nsrR* gene itself and two *hmp* genes (*hmpA1* and *hmpA2*) (24). These genes encode NO detoxifying flavohemeoglobins (17) that convert NO to nitrate (or nitrous oxide under anaerobic conditions). Therefore, in this organism, NsrR appears to regulate only the detection and detoxification of NO.

NsrR is a member of the Rrf2 family of transcriptional regulators, which includes IscR that regulates FeS cluster biosynthesis (18, 19). Like IscR, NsrR contains three conserved cysteine residues in the C terminus region that act as ligands to an iron-sulfur cluster (20–22). Recently it was shown that NsrR from *S. coelicolor* (*Sc*NsrR),[2] previously reported to contain a [2Fe-2S] cluster (23), can also accommodate a [4Fe-4S] cluster, and that this form alone exhibits high affinity DNA binding to *Sc*NsrR-regulated genes (24), consistent with *Sc*NsrR functioning as a repressor. Furthermore, some non-physiological low molecular weight thiols were shown to promote, in the presence of O_2, conversion to a [2Fe-2S] form, likely accounting for the [2Fe-2S] form previously reported.

Here we report studies of the effects of NO on DNA binding by [4Fe-4S] *Sc*NsrR along with spectroscopic and kinetic studies of the cluster reaction with NO. The data reveal that DNA binding is abolished at different stoichiometric ratios of NO to cluster, depending on the promoter sequence. Binding of *Sc*NsrR to the *hmpA2* gene promoter was found to be the most sensitive, with binding abolished at ~2 NO per cluster. Spec-

[2] The abbreviations used are: *Sc*NsrR, *S. coelicolor* NsrR; FeS, iron-sulfur; DNIC, dinitrosyl iron complex; RRE, Roussin's red ester; Bistris propane, 1, 3-bis[tris(hydroxymethyl)methylamino]propane.

SASBMB

troscopic studies revealed a distinct intermediate at the same NO:cluster ratio.

Experimental Procedures

Purification of S. coelicolor NsrR—Wild type and C terminally His-tagged *Sc*NsrR were purified as previously described (24, 25). Protein concentrations were determined using the method of Smith *et al.* (Pierce) (26) with bovine serum albumin as the standard. Cluster content was determined using an extinction coefficient of $\epsilon_{406\ nm} = 13.30\ (\pm 0.19)\ mM^{-1}\ cm^{-1}$ (24).

Analytical Methods—Stock solutions of the NO donor PROLI-NONOate ($t_{1/2} = 1.5$ s; Cayman Chemicals) were prepared in 25 mM NaOH, quantified optically ($\epsilon_{252\ nm}$ 8400 M^{-1} cm^{-1}) and calibrated as previously described (27). For kinetic experiments, an aliquot of PROLI-NONOate was combined with assay buffer (20 mM Tris, 20 mM MES, 100 mM NaCl, 20 mM Bistris propane, 5% glycerol, pH 8.0) and allowed to decompose in a gas tight syringe (Hamilton) to achieve the desired NO concentration before addition to *Sc*NsrR samples.

Electrophoretic Mobility Shift Assays (EMSAs)—DNA fragments carrying the *hmpA1* (SCO7428), *hmpA2* (SCO7094), or *nsrR* (SCO7427) promoters were PCR amplified using *S. coelicolor* genomic DNA and band shift assays carried out as previously described (24), but with [4Fe-4S] NsrR following reaction with increasing concentrations of NO.

Spectroscopy—For reactions with NO, initial experiments resulted in the observation of a white precipitate in the solution at ratios of NO:[4Fe-4S] of >2. We found that the inclusion of glutathione (0.3 mM) in the buffer solution stabilized the solution against precipitation, even at high levels of NO. Therefore, all spectroscopic studies described here were performed in the presence of glutathione unless otherwise indicated.

UV-visible absorbance measurements were made with a Jasco V500 spectrometer and CD spectra were measured with a Jasco J810 spectropolarimeter. CD titrations were repeated in the presence of a 23-bp double stranded oligonucleotide (dsDNA) that included the *hmp1A* binding site. The dsDNA was annealed from two single strands of DNA (5′-AACACGAATATCATCTACCAAT-3′ and complement strand) according to the manufacturer's instructions (Integrated DNA Technologies). DNA was quantitated via $A_{260\ nm}$ and the molecular weight for the dsDNA calculated using OligoCalc (28). CD data were noisier than in the absence of DNA, reflecting difficulties associated with working with viscous solutions of DNA (29). Fluorescence measurements were made using an anaerobic fluorescence cell (1-cm path length) in a PerkinElmer LS55 spectrometer.

EPR spectra were recorded on a Bruker EMX (X-band) EPR spectrometer equipped with an Oxford Instruments liquid helium system and a spherical high-quality ER 4122 (SP 9703) Bruker resonator. Composite EPR spectra were deconvoluted into individual EPR signals by using the procedure of spectra subtraction with variable coefficients (30, 31). The concentrations of the paramagnetic complexes in the samples were determined by relating double integrals of the protein EPR spectrum to that of 1 mM Cu^{2+} in 10 mM EDTA standard, both measured at identical instrumental conditions and in the absence of

microwave power saturation, *i.e.* at 77 K and $P_{MW} = 0.2$ milliwatt. EPR spectra simulation was performed by using WinEPR SimFonia version 1.26 (Bruker Analytik GmbH).

EPR samples (250 μl) were prepared by combining aliquots of protein and PROLI-NONOate to achieve the desired [NO]:[FeS] ratio. Samples were incubated at ambient anaerobic glovebox temperature ($\sim$21 °C) for 5 min prior to loading into the EPR tube and freezing.

Rapid Reaction Kinetics—UV-visible stopped-flow experiments were performed with a Pro-Data upgraded Applied Photophysics Bio-Sequential DX.17 MV spectrophotometer, with a 1-cm path length cell. Absorption changes were detected at a single wavelength (360 or 420 nm), as previously described (32, 33). Prior to use, the stopped-flow system was flushed with $\sim$30 ml of anaerobic assay buffer and experiments were carried out using gas tight syringes (Hamilton). All solutions used for stopped-flow experiments were stored and manipulated inside an anaerobic cabinet (Belle Technology). Rapid kinetic experiments were done in the absence of glutathione because precipitation did not occur in the time window of experiments ($\sim$10 s). Fitting of the overall multi-phase kinetic data at 360 and 420 nm (separately and together) was performed using Dynafit (BioKin, CA) (34), which employs numerical integration of simultaneous first-order differential equations, and verified by fitting individual phases to single or double exponential functions using Origin (version 8, Origin Labs). Where appropriate, apparent second order rate constants were obtained from plots of observed rate constants (k_{obs}) against initial NO concentrations.

Results

Reaction of [4Fe-4S] NsrR with NO Abolishes Binding to NsrR-regulated Promoters at Different Ratios of NO to Cluster—It was recently demonstrated that [4Fe-4S] NsrR binds tightly at an 11-bp inverted repeat sequence in the promoters of *hmpA1* and *hmpA2*, in addition to its own promoter (24). Loss of the cluster to form apoprotein, or conversion to form a [2Fe-2S] form, resulted in loss of DNA binding. Because *Sc*NsrR is an NO sensing regulator that controls only three genes, it was of interest to investigate the effect of NO on the DNA binding properties of the [4Fe-4S] NsrR with the *hmpA1*, *hmpA2*, and *nsrR* promoters. EMSA experiments were conducted with fluorescently (6-FAM)-labeled PCR fragments carrying the promoters, [4Fe-4S] NsrR, and increasing concentrations of NO (see Fig. 1). Prior to the addition of NO, full binding of the promoter DNA was observed (24) and addition of NO resulted in gradual appearance of unbound DNA. Binding of NsrR to *hmpA2* was reduced to 50% at a $\sim$1.4 NO per cluster and was lost entirely by 2.5 NO per cluster (Fig. 1*A*). For *hmpA1* equivalent ratios were $\sim$2.3 (50% binding) and 4.2 (complete loss of binding) (Fig. 1*B*) and for *nsrR* they were $\sim$4.1 (50% binding) and 8.2 (complete loss of binding) (Fig. 1*C*). These data demonstrate that DNA binding is abolished at different ratios of NO:[4Fe-4S], depending on the promoter, and that, for the *hmpA2* promoter, DNA binding is entirely lost at $\sim$2 NO per cluster.
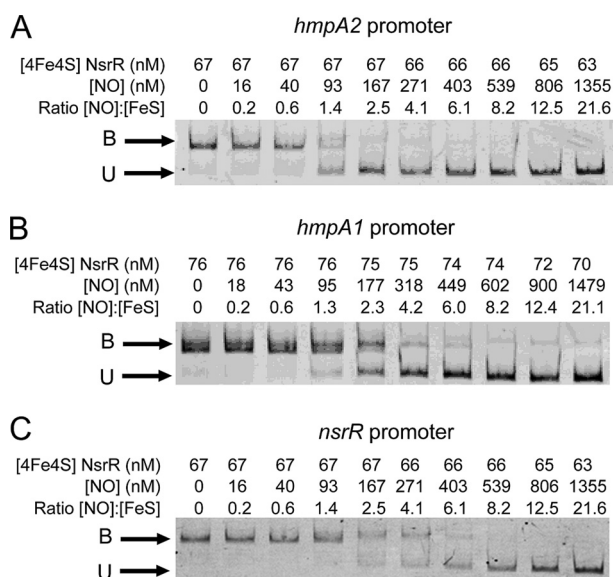
FIGURE 1. **Effect of NO on *Sc*NsrR DNA binding to NsrR-regulated promoters.** *A*, titration of DNA probe (10.6 nM) containing the *hmpA1* promoter with [4Fe-4S] NsrR following reaction with increasing concentrations of NO, as indicated. *B*, as in *A*, except that the DNA probe (5.9 nM) contained the *hmpA2* promoter. *C*, as in *A* except that the DNA probe (4.6 nM) contained the *nsrR* promoter. The binding buffer contained 10 mM Tris, 54 mM KCl, 0.3% (v/v) glycerol, 1.32 mM GSH, pH 7.5.

The Reaction of [4Fe-4S] NsrR with NO: a Multi-NO Reaction with an Intermediate at 2 NO per Cluster—The reaction of [4Fe-4S] NsrR with NO was investigated by measuring changes in the cluster absorption bands following sequential additions of NO under anaerobic conditions (see Fig. 2*A*). Initial increases in intensity were observed at $A_{360\,nm}$, and to a lesser extent in the 500–600 nm region. As the titration progressed, further increases in the 360-nm region were observed as the spectrum changed form. The final spectrum, with principal absorption at ~360 nm and a shoulder at ~430 nm, is consistent with the formation of iron-nitrosyl species, and closely resembles the spectra of the products formed upon nitrosylation of *S. coelicolor* WhiD and *E. coli* FNR. These were assigned to Roussin's red ester (RRE)-like species (see Fig. 3) (32, 33). RRE complexes exhibit a principal absorption band at 362 nm and a shoulder at ~430 nm (27, 35). Importantly, no isosbestic points were observed, suggesting a complex reaction pathway involving several intermediates, as illustrated by the highlighted spectra in Fig. 2*A*, which show the form of the iron nitrosyl species changes during the titration. A plot of $A_{360\,nm} - A_{420\,nm}$ *versus* [NO]:[4Fe-4S] (Fig. 4*A*) shows that the reaction was complete at a stoichiometry of 8–10 NO molecules per cluster, as observed for other NO-sensitive FeS regulators (32, 33). However, for NsrR there is a clear break point in the plot at a stoichiometry of ~2 NO molecules per cluster and a further, less distinct one at ~6 NO per cluster.

Similar changes induced by NO additions were followed by tryptophan fluorescence ($FI_{353\,nm}$) (see Fig. 2*B*). The [4Fe-4S]$^{2+}$ cluster acts as a quencher of protein fluorescence (32) but as NO was added, the fluorescence intensity decreased, indicating that the iron-nitrosyl species formed is a more efficient quencher of fluorescence intensity. As further NO was added, intensity recovered to approximately the starting point, indi-

cating the conversion of the initial iron-nitrosyl species (an intermediate) into a different iron-nitrosyl species (product(s)). A plot of fluorescence intensity changes at 353 nm against the ratio [NO]:[4Fe-4S] (Fig. 4*B*) showed the reaction is complete at 8–10 NO, with the formation of the fluorescence detectable intermediate at 3–4 NO per cluster.

The CD spectra in the near UV-visible region of the FeS cluster arise from the chirality imposed by the protein fold. Hence changes in the CD spectra allow reactions with NO to be followed. Sequential NO addition showed major changes during the course of the titration, reflecting formation of intermediates (see Fig. 2, *C* and *D*). The starting spectrum contained a small positive feature at 330 nm and a major negative feature at 400 nm, as previously reported (24). As NO was added, the intensity of the band at +330 nm increased significantly, whereas the band at −400 nm decreased in intensity and shifted to −380 nm (Fig. 2*C*). As further NO was added, the +330 nm band was lost and the remaining intensity at −380 nm decreased and shifted further to −370 nm. A broad negative feature was also observed at 520 nm (Fig. 2*D*). A plot of CD intensity at 430 nm (Fig. 4*A*) showed that changes were complete at ~6 NO per cluster, with a clear break at ~2 NO. An equivalent plot of CD intensity changes at 330 nm (Fig. 4*B*) very clearly showed the formation of an intermediate at 2 NO per cluster, which subsequently reacts with further NO to give a less distinct intermediate at ~4 NO with the CD response essentially complete at 7–8 NO. All three forms of UV-visible spectroscopy absorption, fluorescence, and CD data show a complex reaction course with [4Fe-4S] NsrR clearly forming intermediates at ~2 NO, and ~4 and ~6 NO molecules, with no further reaction beyond 8–10 NO per cluster.

Reaction with NO Is Not Significantly Affected by [4Fe-4S] NsrR DNA Binding—Because NsrR is a regulatory protein, it will encounter NO when bound to DNA. It was therefore of interest to determine whether the DNA binding affects the reaction of [4Fe-4S] NsrR with NO. CD was used to investigate this because, of the spectroscopic methods above, it gave the most distinctive response to the NO reaction. Thus a CD titration was repeated with [4Fe-4S] NsrR bound to a 23-mer oligomer containing the NsrR-binding sequence of the *hmpA1* promoter, previously found to bind NsrR with highest affinity of all NsrR-regulated promoters (24). In the presence of DNA, the response of the negative feature at −400 nm was essentially identical to that observed in the non-DNA bound form (Fig. 5, *A* and *B*), with breaks in the response at ~2 and ~6 NO per cluster (Fig. 5*C*). The intermediate species detected at +330 nm was also observed to form and decay in a similar way. Maximum intensity occurred at ~2 NO per cluster, slightly shifted compared with the absence of DNA (Fig. 5*D*). Some differences were observed at higher ratios of NO, such that the shoulder observed at ~4 NO in the absence of DNA was not detected in its presence (Fig. 5*D*), but this may be due to the increased noise of the spectra. Overall, the major features of the NO responses are similar for [4Fe-4S] NsrR free in solution or bound to DNA.

EPR Spectroscopy of Nitrosylated [4Fe-4S] NsrR Reveals the Formation of DNIC Species—Reactions of protein-bound FeS clusters with NO were first observed by EPR spectroscopy, through the detection of paramagnetic mononuclear iron dini-
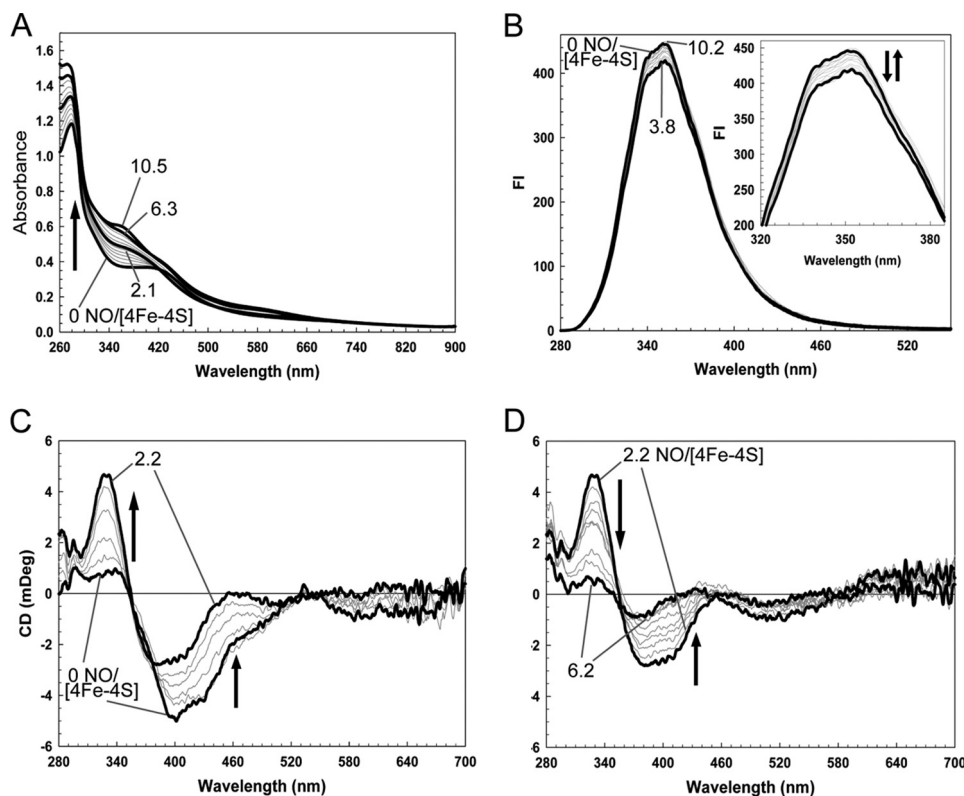
FIGURE 2. **Titrations of [4Fe-4S] *Sc*NsrR with NO.** *A*, absorbance spectra of [4Fe-4S] NsrR following sequential additions of NO up to a [NO]:[FeS] ratio of 10.5 (*black lines* show spectra recorded at ratios of 0, 2.1, 6.3, and 10.5). *B*, fluorescence spectra obtained during a titration equivalent to that in *A*; *inset* shows changes in more detail. *Black lines* show spectra recorded at [NO]:[FeS] ratios of 0 (*lower*) and 3.8 (*upper*). *C* and *D*, CD spectra obtained during a titration equivalent to that in *A*. *Black lines* show spectra recorded at [NO]:[FeS] ratios of 0 and 2.2 in *C*, and 2.2 and 6.2 in *D*. *Arrows* indicate the direction of intensity changes. *Sc*NsrR (28 μM) was in 20 mM Tris, 20 mM MES, 20 mM Bistris propane, 100 mM NaCl, 250 μM GSH, 5% (v/v) glycerol, pH 8.0.
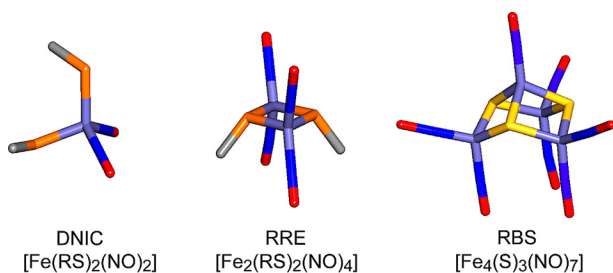

FIGURE 3. **Iron-nitrosyl species that may be formed following nitrosylation of protein-bound FeS clusters.** Structures of DNIC, RRE, and Roussin's black salt (*RBS*) species are illustrated. Thiolate (RS) groups are shown in *orange*, iron in *light blue*, nitrogen in *dark blue*, oxygen in *red*, and sulfide in *yellow*.

trosyl (DNIC) species (36, 37) (Fig. 3). This provides a means of quantifying the amount of DNIC species formed during the course of reactions with NO (32, 33, 38, 39). Therefore, EPR spectroscopy was used to assess the formation of DNIC species upon nitrosylation of [4Fe-4S] NsrR. NO was added to NsrR in increasing NO:[4Fe-4S] ratios from substoichiometric to large excess, allowed to react for 5 min and then frozen for EPR measurements. Prior to the addition of NO, the spectrum was devoid of signals, consistent with the presence of diamagnetic [4Fe-4S]$^{2+}$. On addition of NO signals in the $g = \sim 2$ region, characteristic of the S = ½ DNIC species, were observed increasing in intensity with increasing ratio of NO to cluster (see Fig. 6A). Analysis of the spectra revealed that each can be deconvoluted into three distinct signals that contribute to dif-

ferent extents to the evolving spectra (see Fig. 6B and Table 1). Signal 1 (Sig1) was simulated as a S = ½ species with $g_x = 2.0440$, $g_y = 2.0246$, and $g_z = 2.0000$, and signal 2 (Sig2) as a S = ½ species with $g_x = 2.0426$, $g_y = 2.0332$, and $g_z = 2.0140$. Signal 3 (Sig3) could not be simulated as a single species and even at its maximum was of very low intensity in the observed spectra. Up to a ratio of ~6 NO per cluster, signals 1 and 2 contributed equally to the observed spectrum, but above this ratio, signal 1 decayed away and signal 2 grew further. Signal 2 is characteristic of a Cys-coordinated DNIC, but signal 1 is not similar to previously characterized DNIC species (40) and so may represent another type of iron-nitrosyl species. The relatively small increase in DNIC intensity observed beyond ~10 NO per cluster most likely results from some conversion of multinuclear iron-nitrosyl species (see later) into DNICs. Spin integration of the signals yielded a total maximum concentration equivalent to ~60% of the original [4Fe-4S] concentration; that is, ~15% of the iron originally present as the cluster.

[4Fe-4S] NsrR Cluster Reacts Rapidly with NO—The reaction of [4Fe-4S] NsrR with excess NO (NO:[4Fe-4S] ~32) was followed using stopped-flow absorbance spectroscopy, monitoring $A_{360 \text{ nm}}$ and $A_{420 \text{ nm}}$ as a function of time. These wavelengths correspond to the maxima of the final nitrosylated product and the initial iron-sulfur cluster, respectively (Fig. 7). A rapid, multiphase reaction was observed at both wavelengths, as previously observed for other FeS regulators (32, 33). The data were fitted separately, and together, to exponential functions, giving equivalent results. Analysis revealed the presence
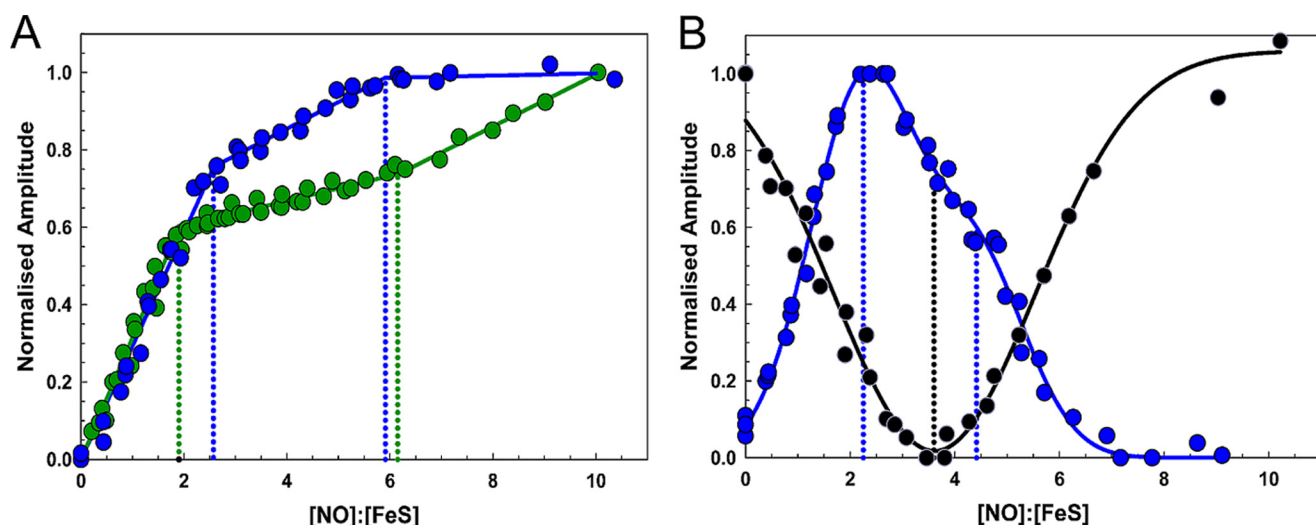
FIGURE 4. **Plots of spectroscopic changes as a function of NO concentration.** *A*, normalized $A_{360\,nm} - A_{420\,nm}$ (*green circles*) and $CD_{430\,nm}$ (*blue circles*), and *B*, normalized $CD_{330\,nm}$ (*blue circles*), and $FI_{350\,nm}$ (*black circles*) plotted *versus* the [NO]:[FeS] ratio. Data are from two independent titrations (data for one of these are shown in Fig. 1).
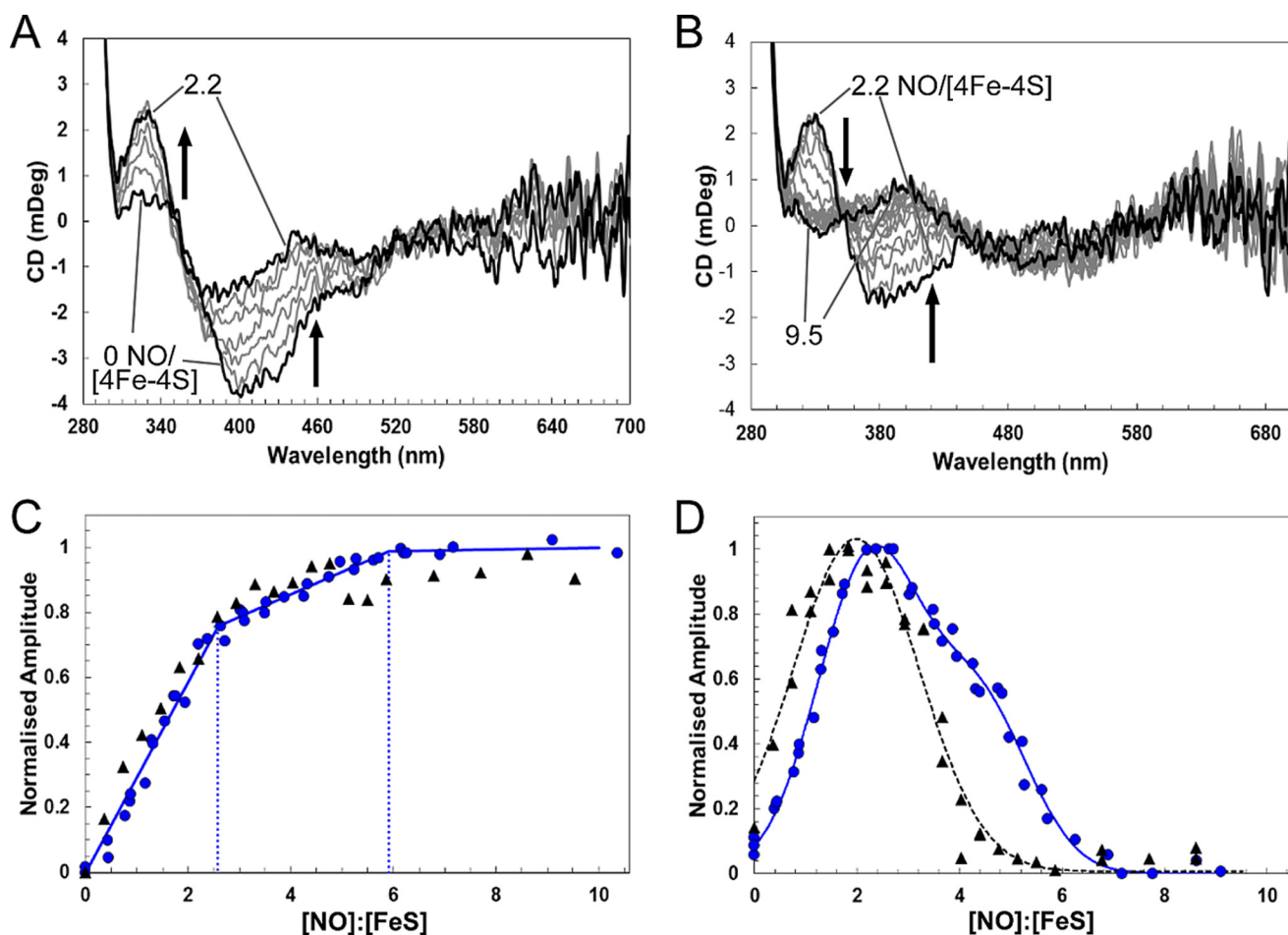
FIGURE 5. **The effect of DNA binding on [4Fe-4S] ScNsrR reaction with NO.** CD spectra of [4Fe-4S] NsrR (16 μM [4Fe-4S] NsrR dimer, 32 μM [4Fe-4S]) following sequential additions of NO during a titration equivalent to that in Fig. 2, *C* and *D,* in the presence of 32 μM dsDNA. *Black lines* show spectra recorded at [NO]:[FeS] ratios of 0 and 2.2 in *A*, and 2.2 and 9.5 in *B*. *Arrows* indicate the direction of intensity changes. *Sc*NsrR was in 10 mM Tris, 54 mM KCl, 0.3% (v/v) glycerol, 1.5 mM GSH, pH 7.5. *C* and *D*, *black triangles* show normalized CD intensity at 430 and 330 nm, respectively, plotted *versus* the [NO]:[FeS] ratio for reaction in the presence of DNA. Equivalent data for reaction in the absence of DNA is replotted (*blue circles*) from Fig. 3 for comparison. Data are from two independent titrations (data for one are shown in *A* and *B*).

of four phases at 360 nm and three phases at 420 nm. The first two phases were detected at both wavelengths, but the remaining phases had different kinetic characteristics (at the two

wavelengths), indicating that they report on different processes. Thus, overall, the third phase was detected at 360 nm, the fourth at 420 nm, and the final phase at 360 nm. Thus, the

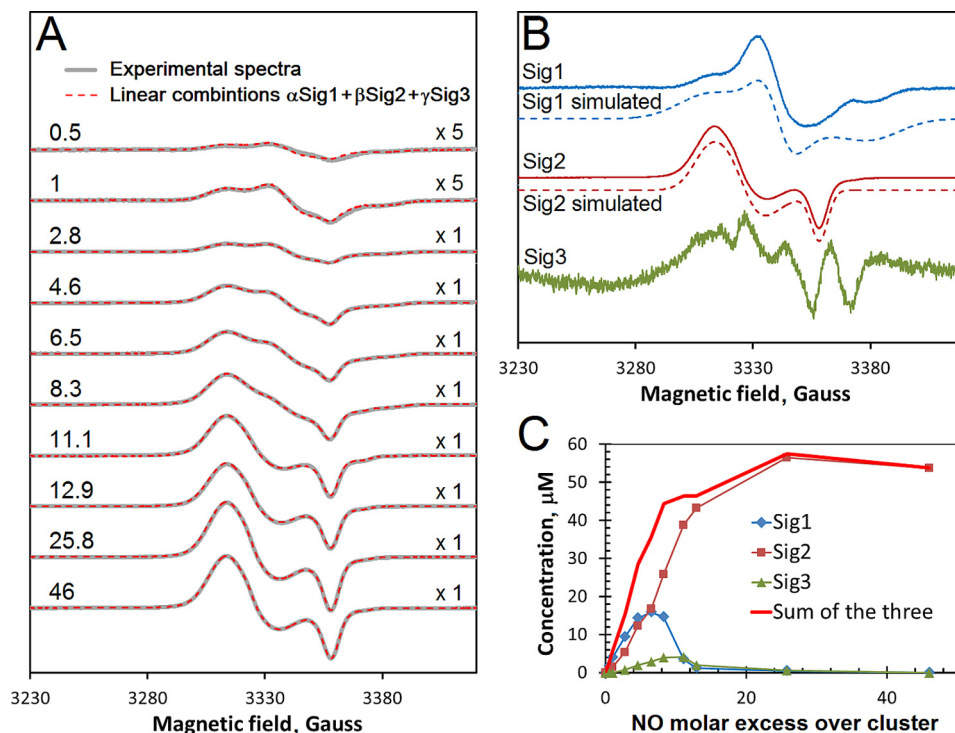FIGURE 6. **EPR analysis of DNIC formation during reaction of [4Fe-4S] NsrR with NO.** *A*, EPR spectra following the addition of NO to 100 μM [4Fe-4S] NsrR (*gray lines*). The spectrum of NsrR prior to NO treatment was subtracted from each. The [NO]:[FeS] ratios are indicated. The first two spectra are magnified by a factor of 5 indicated by the "x" symbol (on the *right*). The experimental data are overlaid with linear combinations of the three EPR signals shown in *B*. The coefficients α, β, and γ used in these linear combinations are given in Table 1. *B*, the three EPR signals (*solid lines*) assumed to be basic components of all spectra shown in *A*, were obtained as described under "Experimental Procedures." Signals 1 and 2 were simulated (*dashed lines*) with the following parameters: Sig1, $g_x = 2.0440$, $g_y = 2.0246$, and $g_z = 2.0000$ ($\Delta H_x = 25$ G, $\Delta H_y = 12$ G, $\Delta H_z = 25$ G); Sig2, $g_x = 2.0426$, $g_y = 2.0332$, and $g_z = 2.0140$ ($\Delta H_x = 14$ G, $\Delta H_y = 14$ G, $\Delta H_z = 7$ G). *C*, concentrations of the species responsible for EPR signals Sig1, Sig2, and Sig3 as functions of the excess of NO over cluster. Spectra were recorded at 77 K. Microwave power and frequency were 3.18 milliwatts and 9.47 GHz, respectively, and field modulation amplitude was 0.3 millitesla. The sample buffer was 50 mM Tris, 2 M NaCl, 5% (v/v) glycerol, pH 8.0.

TABLE 1

Coefficients in the linear combinations (αSig1 + βSig2 + γSig3) of signals 1, 2, and 3 (Fig. 6*B*) used for simulations of the experimental spectra shown in Fig. 6*A*

NO excess	α	β	γ
0.5	0.0315	0.0134	0.0000
1	0.0724	0.0291	0.0000
2.8	0.1725	0.0994	0.0106
4.6	0.2650	0.2275	0.0256
6.5	0.2917	0.3120	0.0369
8.3	0.2700	0.4800	0.0513
11.1	0.0650	0.7227	0.0525
12.9	0.0223	0.8034	0.0258
25.8	0.0092	1.0500	0.0075
46	0.0000	1.0000	0.0000

overall reaction was modeled as a five step reaction, *i.e.* A $\rightarrow$ B $\rightarrow$ C $\rightarrow$ D $\rightarrow$ E $\rightarrow$ F, where the initial (A $\rightarrow$ B) and second (B $\rightarrow$ C) steps are detected at both wavelengths, C $\rightarrow$ D is detected at 360 nm, D $\rightarrow$ E at 420 nm, and the final step E $\rightarrow$ F at 360 nm.

Experiments were repeated at NO:[4Fe-4S] ratios ranging from 2 to 130 and the data fitted as described under "Experimental Procedures," to give observed rate constants. Plots of observed rate constants (k_{obs}) against the NO concentration are shown in Fig. 8. The observed rate constant for the first step (A $\rightarrow$ B) exhibited, initially, a first order dependence on NO (measured at both 360 and 420 nm) (see Fig. 8*A*). The gradient of the dependence gave a second order rate constant of $\sim$4.5 $\times$ 10^6 M^{-1} s^{-1} (see Table 2). At $\sim$150 μM NO, the reaction became

independent of NO, indicating that at higher NO concentrations, the rate determining step switched to a process that does not involve NO.

Step 2 (B $\rightarrow$ C) was found to be linearly dependent on NO (Fig. 8*B*), giving a rate constant an order of magnitude lower than that for step 1 (Table 2). Step 3 (C $\rightarrow$ D) was linear with NO in the range $\sim$100–500 μM with a rate constant lower again by an order of magnitude (Fig. 8*C*, Table 2). At lower concentrations, insufficient amplitude was detected in the few seconds of measurement for the phase to be fitted. At higher NO, the reaction became independent of NO concentration. Steps 4 (D $\rightarrow$ E, Fig. 8*D*) and 5 (E $\rightarrow$ F, Fig. 8*E*) were similar in that they were not detected at low NO but were linearly dependent on NO at intermediate NO concentrations before becoming NO independent at high NO concentrations. The rate constants determined from the linear parts of the plots were sequentially lower by an order of magnitude than that for the previous step (Table 2).

Discussion

Here we provide novel biochemical insight into why the response to NO may not be uniform for all genes controlled by a single (FeS) regulator. DNA band shift experiments revealed that the response of the three *Sc*NsrR-bound promoters to NO was different, with *hmpA2* the most sensitive and *nsrR* the least sensitive. This implies a hierarchy of expression response to NO. Interestingly, previous studies of [4Fe-4S] NsrR binding to
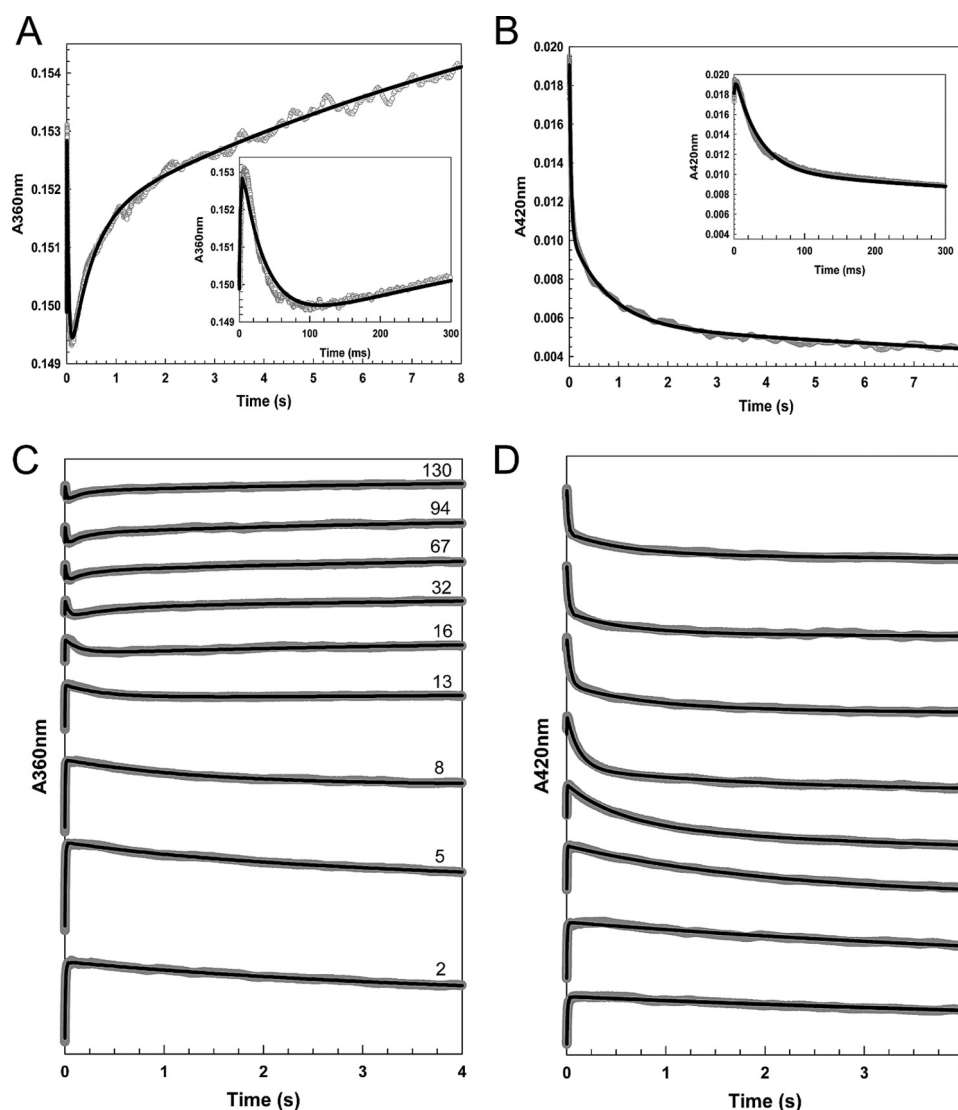
FIGURE 7. **Stopped-flow measurements of the reaction of [4Fe-4S] NsrR with NO.** *A–D*, absorbance at 360 (*A* and *B*) and 420 nm (*C* and *D*) following the addition of NO to NsrR ($\sim$7.6 μM). *A* and *C* show data at 360 and 420 nm, respectively, for the addition of $\sim$32 NO molecules per cluster. *B* and *D* show data at 360 and 420 nm, respectively, for a range of other NO:cluster ratios, as indicated. *Insets* in *A* and *C* show early events in the reaction time course. Fits to each of the observed phases (see "Experimental Procedures") are drawn in *black lines*.

DNA revealed that binding to the *hmpA2* promoter was the weakest (24). However, there is no clear correlation with promoter binding affinity because binding to *hmpA1*, which here was found to have intermediate sensitivity to NO, exhibited the strongest binding to [4Fe-4S] NsrR (24).

Analysis of the reaction of [4Fe-4S] NsrR with NO using spectroscopic and kinetic methods revealed a complex process involving reaction of up to 8–10 NO per cluster, but with a series of intermediates, at $\sim$2, 4, and 6 NO per cluster, formed along the nitrosylation pathway. Thus the *Sc*NsrR nitrosylation reaction is not concerted, *i.e.* NO does not react preferentially to completion with clusters that have already undergone initial reaction, relative to those that have not yet reacted.

These studies follow related investigations of other NO-sensing FeS regulators, including *Mycobacterium tuberculosis* WhiB1, *S. coelicolor* WhiD, and *E. coli* FNR (32, 33). The data reported here for *Sc*NsrR bear similarities to those regulators,

in that reaction involves multiple NO molecules and results in iron-nitrosyl products. However, whereas an intermediate was observed at $\sim$4 NO:[4Fe-4S] for the reaction of [4Fe-4S] FNR (33) and possibly also [4Fe-4S] WhiD (32), no other intermediates, particularly at low ratios of NO to cluster, have been detected previously.

Importantly the spectroscopic observations link to the DNA binding data. Binding of *Sc*NsrR to the *hmpA2* promoter was entirely abolished at $\sim$2 NO per cluster, indicating that the *Sc*NsrR intermediate species detected at this ratio can no longer bind *hmpA2* promoter DNA. For *hmpA1* and *nsrR* promoters, $\sim$4 and $\sim$8 NO molecules per cluster, respectively, were required to abolish DNA binding. For *hmpA1*, this ratio also corresponds to an intermediate observed via spectroscopy, whereas for *nsrR*, it suggests that the full nitrosylation reaction is needed to abolish binding. Previous studies of cluster reactivity of a DNA-bound FeS regulator revealed that the rate of reaction was affected but the overall mechanism was not (29). The
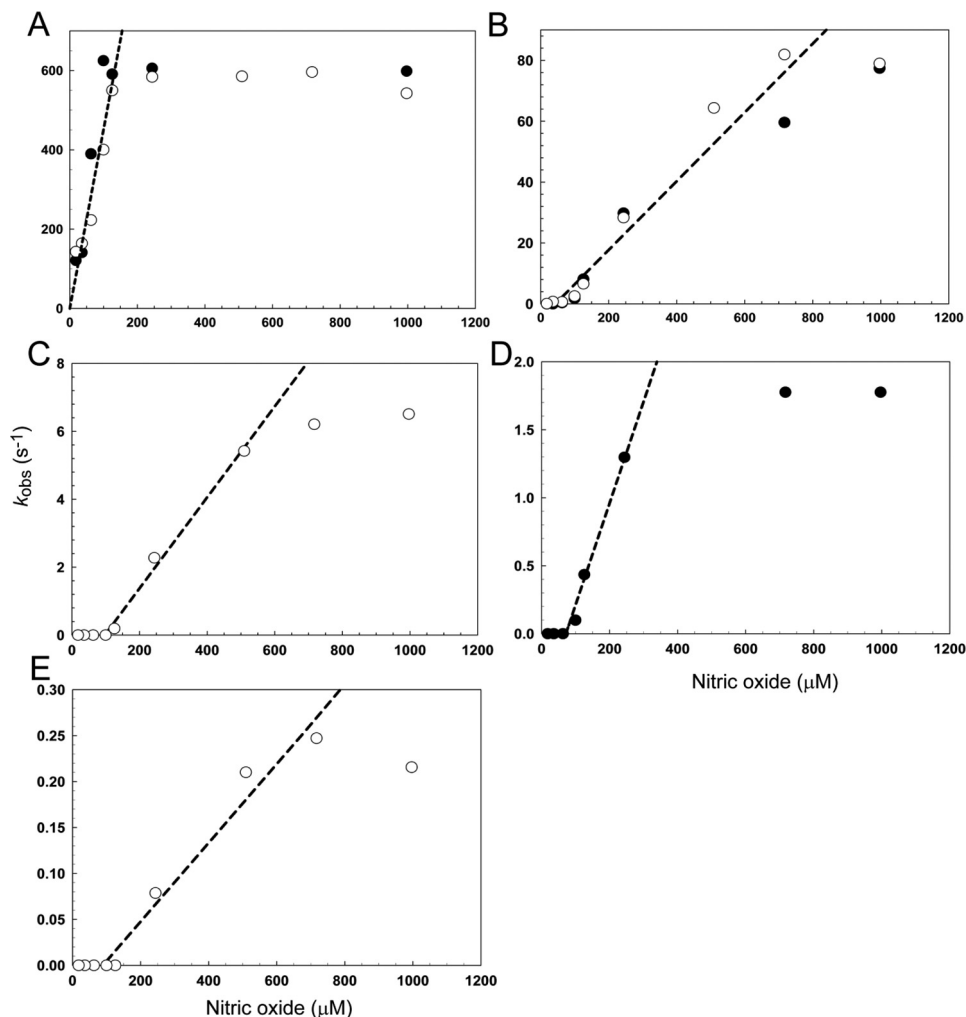
FIGURE 8. **Dependence of the observed rate constant for each step of nitrosylation on NO.** *A–E,* plots of the observed (pseudo-first order) rate constant (k_{obs}), obtained from fits of the kinetic data at 360 (*open circles*) and 420 nm (*filled circles*), over a range of NO concentrations. Note that *panels A–E* correspond to steps 1 to 5, respectively, of the reaction (see text). Least squares linear fits are shown giving apparent second order rate constants (see Table 2). The buffer was 20 mM Tris, 20 mM MES, 20 mM Bistris propane, 100 mM NaCl, 5% (v/v) glycerol, pH 8.0.

TABLE 2

Apparent second order rate constants for the five observed steps in the nitrosylation reaction of [4Fe-4S] *Sc*NsrR and comparison with those of *S. coelicolor* WhiD (32) and *E. coli* FNR (33)

Note that, with the exception of the initial reaction (A → B), the rate constants for NsrR may not represent the same reaction as for WhiD and FNR.

Phase	Step	Rate constant		
		NsrR	WhiD	FNR
			$M^{-1} s^{-1}$	
1	A → B	$4.52 \pm 0.23 \times 10^6$	4.40×10^5	2.81×10^5
2	B → C	$1.13 \pm 0.11 \times 10^5$	1.38×10^4	1.89×10^4
3	C → D	$1.34 \pm 0.10 \times 10^4$	8.34×10^3	4.61×10^3
4	D → E	$7.48 \pm 0.69 \times 10^3$	0.90×10^3	0.75×10^3
5	E → F	$0.43 \pm 0.05 \times 10^3$		

clear correlation between observed intermediates and DNA binding behavior of *Sc*NsrR suggests that it is unlikely that DNA binding significantly affects the mechanism of the NsrR nitrosylation reaction, and this conclusion is consistent with CD measurements showing that DNA-bound *Sc*NsrR reacted with NO similarly to *Sc*NsrR free in solution. Although the structures of the intermediate species formed cannot be determined from the current data, a key conclusion from this work is that complete reaction with the FeS cluster of this NO-sensing

regulator is not a requirement and DNA binding can be switched off at low ratios of NO to cluster.

The kinetic data revealed a rapid, complex multiphase reaction that was modeled most simply as a five step reaction. The first step of the reaction with NO (A → B; first order with respect to NO) is very likely the binding of one NO molecule to the cluster. The second step (B → C; first order with respect to NO), generating the intermediate with clear spectroscopic characteristics, results from the binding of a second NO, which could be at the same iron or elsewhere on the cluster. As in previous studies of FeS regulators, it is difficult to assign the identity of this species and those resulting from subsequent steps of the reaction because the form of the iron, and how it changes during the reaction as the cluster breaks down, cannot be determined from these data and are, in fact, extremely difficult to identify unambiguously. Each step shows a linear dependence on NO (at least initially where they are detected), indicating that they correspond to the sequential binding of NO to iron. Individual steps could involve the binding of more than one NO, but this would involve independent binding of NO to different irons of the cluster, giving an overall first order dependence.

The rate constant for the initial step of the *Sc*NsrR NO reaction is an order of magnitude greater than that detected for WhiD from the same organism (Table 2) (32). Furthermore, the rate constant for the slowest step of the NO reaction is at least ~3 orders of magnitude greater than that for the slowest step of the reaction with O_2 (24). These observations are consistent with a role for *Sc*NsrR as the first line of defense against NO, *i.e.* from kinetic analyses NsrR would be predicted to preferentially react with NO in the *S. coelicolor* cytoplasm containing both NsrR and WhiD.

The kinetic data also support the conclusion that the *Sc*NsrR cluster nitrosylation is not a concerted reaction. Under conditions of excess NO, five phases were detected, but at low NO:cluster ratios, only the early phases of the reaction were observed, consistent with there being sufficient NO to achieve the formation of only the first intermediates of the full reaction. The effect of this is that plots of observed rate constants *versus* NO concentration have an unusual appearance for the mid/latter phases, in that there are zero values at low NO (Fig. 8).

In the case of WhiD, the cluster is coordinated by four Cys residues, whereas NsrR is coordinated by three Cys residues and one oxygenic residue that previous DNA binding and spectroscopic data suggested might be Glu-85 (24). Previous studies of the effect of low molecular weight thiols suggested that the oxygenic ligand can be readily displaced to generate an all thiol-coordinated cluster (24), and so the unique iron site of the cluster is the most likely site of initial NO binding. The kinetic data showed that the observed rate constant for the initial binding becomes independent of NO above a particular concentration (Fig. 8*A*). At this point, the slow step of the reaction does not involve NO and we propose that, at high NO concentrations, the dissociation of the existing (oxygenic) ligand to the iron, permitting binding of NO, is the rate-limiting step. The rate constant for this process, from the plot in Fig. 8*A*, is estimated to be ~600 s^{-1}.

The form of the final absorbance spectrum of the nitrosylated cluster is similar to those observed previously with WhiD and FNR (32, 33). Although the nature of the iron-nitrosyl species cannot be determined solely from its absorbance properties, these are consistent with RRE-like $[Fe^I_2(NO)_4(Cys)_2]$ species, as proposed for WhiD and FNR (32, 33), rather than DNIC species, which have distinct absorbance properties (27, 35). However, the spectra could also arise from more complex iron-nitrosyl species such as those related to Roussin's black salt (see Fig. 3) (39). For WhiD and FNR, only minor amounts of DNICs ($[Fe^I(NO)_2(Cys)_2]$) were detected (<4% total iron) (32, 33). In the case of *Sc*NsrR, significantly more DNIC species, up to 15% of the total iron, was detected by EPR; however, this is still a relatively minor component of the products.

Concluding Remarks—The data presented here reveal novel aspects of NO sensing by an FeS regulatory protein, with distinct responses of DNA binding to NO depending on the sequence of the promoter. Intermediates of cluster nitrosylation, particularly that detected at ~2 NO per cluster, correlate well with DNA binding behavior, pointing to their physiological importance. Further investigations will be needed to try to establish the precise nature of these intermediates.

References

1. Seth, D., Hausladen, A., Wang, Y. J., and Stamler, J. S. (2012) Endogenous protein S-nitrosylation in *E. coli*: regulation by OxyR. *Science* **336,** 470–473
2. Weiss, B. (2006) Evidence for mutagenesis by nitric oxide during nitrate metabolism in *Escherichia coli. J. Bacteriol.* **188,** 829–833
3. Broillet, M. C. (1999) *S*-Nitrosylation of proteins. *Cell Mol. Life Sci.* **55,** 1036–1042
4. Kwon, Y. M., and Weiss, B. (2009) Production of 3-nitrosoindole derivatives by *Escherichia coli* during anaerobic growth. *J. Bacteriol.* **191,** 5369–5376
5. Crack, J. C., Green, J., Thomson, A. J., and Le Brun, N. E. (2012) Iron-sulfur cluster sensor-regulators. *Curr. Opin. Chem. Biol.* **16,** 35–44
6. Drapier, J. C. (1997) Interplay between NO and [Fe-S] clusters: relevance to biological systems. *Methods* **11,** 319–329
7. Szabó, C., Ischiropoulos, H., and Radi, R. (2007) Peroxynitrite: biochemistry, pathophysiology and development of therapeutics. *Nat. Rev. Drug Discov.* **6,** 662–680
8. Bruckdorfer, R. (2005) The basics about nitric oxide. *Mol. Aspects Med.* **26,** 3–31
9. Crack, J. C., Green, J., Thomson, A. J., and Le Brun, N. E. (2014) Iron-sulfur clusters as biological sensors: the chemistry of reactions with molecular oxygen and nitric oxide. *Acc. Chem. Res.* **47,** 3196–3205
10. Spiro, S. (2007) Regulators of bacterial responses to nitric oxide. *FEMS Microbiol. Rev.* **31,** 193–211
11. Corker, H., and Poole, R. K. (2003) Nitric oxide formation by *Escherichia coli*: dependence on nitrite reductase, the NO-sensing regulator FNR, and flavohemoglobin Hmp. *J. Biol. Chem.* **278,** 31584–31592
12. Vine, C. E., and Cole, J. A. (2011) Unresolved sources, sinks, and pathways for the recovery of enteric bacteria from nitrosative stress. *FEMS Microbiol. Lett.* **325,** 99–107
13. Kers, J. A., Wach, M. J., Krasnoff, S. B., Widom, J., Cameron, K. D., Bukhalid, R. A., Gibson, D. M., Crane, B. R., and Loria, R. (2004) Nitration of a peptide phytotoxin by bacterial nitric oxide synthase. *Nature* **429,** 79–82
14. Partridge, J. D., Bodenmiller, D. M., Humphrys, M. S., and Spiro, S. (2009) NsrR targets in the *Escherichia coli* genome: new insights into DNA sequence requirements for binding and a role for NsrR in the regulation of motility. *Mol. Microbiol.* **73,** 680–694
15. Kommineni, S., Lama, A., Popescu, B., and Nakano, M. M. (2012) Global transcriptional control by NsrR in *Bacillus subtilis. J. Bacteriol.* **194,** 1679–1688
16. Heurlier, K., Thomson, M. J., Aziz, N., and Moir, J. W. (2008) The nitric oxide (NO)-sensing repressor NsrR of *Neisseria meningitidis* has a compact regulon of genes involved in NO synthesis and detoxification. *J. Bacteriol.* **190,** 2488–2495
17. Stevanin, T. M., Read, R. C., and Poole, R. K. (2007) The *hmp* gene encoding the NO-inducible flavohaemoglobin in *Escherichia coli* confers a protective advantage in resisting killing within macrophages, but not *in vitro*: links with swarming motility. *Gene* **398,** 62–68
18. Schwartz, C. J., Giel, J. L., Patschkowski, T., Luther, C., Ruzicka, F. J., Beinert, H., and Kiley, P. J. (2001) IscR, an Fe-S cluster-containing transcription factor, represses expression of *Escherichia coli* genes encoding

Fe-S cluster assembly proteins. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 14895–14900

19. Rajagopalan, S., Teter, S. J., Zwart, P. H., Brennan, R. G., Phillips, K. J., and Kiley, P. J. (2013) Studies of IscR reveal a unique mechanism for metal-dependent regulation of DNA binding specificity. *Nat. Struct. Mol. Biol.* **20,** 740–747

20. Tucker, N. P., Le Brun, N. E., Dixon, R., and Hutchings, M. I. (2010) There's NO stopping NsrR, a global regulator of the bacterial NO stress response. *Trends Microbiol.* **18,** 149–156

21. Isabella, V. M., Lapek, J. D., Jr., Kennedy, E. M., and Clark, V. L. (2009) Functional analysis of NsrR, a nitric oxide-sensing Rrf2 repressor in *Neisseria gonorrhoeae. Mol. Microbiol.* **71,** 227–239

22. Yukl, E. T., Elbaz, M. A., Nakano, M. M., and Moënne-Loccoz, P. (2008) Transcription factor NsrR from *Bacillus subtilis* senses nitric oxide with a 4Fe-4S cluster. *Biochemistry* **47,** 13084–13092

23. Tucker, N. P., Hicks, M. G., Clarke, T. A., Crack, J. C., Chandra, G., Le Brun, N. E., Dixon, R., and Hutchings, M. I. (2008) The transcriptional repressor protein NsrR senses nitric oxide directly via a [2Fe-2S] cluster. *PLoS ONE* **3,** e3623

24. Crack, J. C., Munnoch, J., Dodd, E. L., Knowles, F., Al Bassam, M. M., Kamali, S., Holland, A. A., Cramer, S. P., Hamilton, C. J., Johnson, M. K., Thomson, A. J., Hutchings, M. I., and Le Brun, N. E. (2015) NsrR from *Streptomyces coelicolor* is a nitric oxide-sensing [4Fe-4S] cluster protein with a specialized regulatory function. *J. Biol. Chem.* **290,** 12689–12704

25. Crack, J. C., Green, J., Thomson, A. J., and Le Brun, N. E. (2014) Techniques for the production, isolation, and analysis of iron-sulfur proteins. *Methods Mol. Biol.* **1122,** 33–48

26. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150,** 76–85

27. Cruz-Ramos, H., Crack, J., Wu, G., Hughes, M. N., Scott, C., Thomson, A. J., Green, J., and Poole, R. K. (2002) NO sensing by FNR: regulation of the *Escherichia coli* NO-detoxifying flavohaemoglobin, Hmp. *EMBO J.* **21,** 3235–3244

28. Kibbe, W. A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35,** W43–46

29. Crack, J. C., Stapleton, M. R., Green, J., Thomson, A. J., and Le Brun, N. E. (2014) Influence of association state and DNA binding on the O_2-reactivity of [4Fe-4S] fumarate and nitrate reduction (FNR) regulator. *Biochem. J.* **463,** 83–92

30. Svistunenko, D. A. (2005) Reaction of haem containing proteins and enzymes with hydroperoxides: the radical view. *Biochim. Biophys. Acta* **1707,** 127–155

31. Svistunenko, D. A., Davies, N., Brealey, D., Singer, M., and Cooper, C. E. (2006) Mitochondrial dysfunction in patients with severe sepsis: an EPR interrogation of individual respiratory chain components. *Biochim. Biophys. Acta* **1757,** 262–272

32. Crack, J. C., Smith, L. J., Stapleton, M. R., Peck, J., Watmough, N. J., Buttner, M. J., Buxton, R. S., Green, J., Oganesyan, V. S., Thomson, A. J., and Le Brun, N. E. (2011) Mechanistic insight into the nitrosylation of the [4Fe-4S] cluster of WhiB-like proteins. *J. Am. Chem. Soc.* **133,** 1112–1121

33. Crack, J. C., Stapleton, M. R., Green, J., Thomson, A. J., and Le Brun, N. E. (2013) Mechanism of [4Fe-4S](Cys)$_4$ cluster nitrosylation is conserved among NO-responsive regulators. *J. Biol. Chem.* **288,** 11492–11502

34. Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* **237,** 260–273

35. Costanzo, S., Menage, S., Purrello, R., Bonomo, R. P., and Fontecave, M. (2001) Re-examination of the formation of dinitrosyl-iron complexes during reaction of *S*-nitrosothiols with Fe(II). *Inorg. Chim. Acta* **318,** 1–7

36. Vanin, A. F., Kiladze, S. V., and Kubrina, L. N. (1977) Factors influencing formation of dinitrosyl complexes of non-heme iron in the organs of animals *in vivo. Biofizika* **22,** 850–855

37. Tsai, M. L., Tsou, C. C., and Liaw, W. F. (2015) Dinitrosyl iron complexes (DNICs): from biomimetic synthesis and spectroscopic characterization toward unveiling the biological and catalytic roles of DNICs. *Acc. Chem. Res.* **48,** 1184–1193

38. Tinberg, C. E., Tonzetich, Z. J., Wang, H., Do, L. H., Yoda, Y., Cramer, S. P., and Lippard, S. J. (2010) Characterization of iron dinitrosyl species formed in the reaction of nitric oxide with a biological Rieske center. *J. Am. Chem. Soc.* **132,** 18168–18176

39. Tonzetich, Z. J., Wang, H., Mitra, D., Tinberg, C. E., Do, L. H., Jenney, F. E., Jr., Adams, M. W., Cramer, S. P., and Lippard, S. J. (2010) Identification of protein-bound dinitrosyl iron complexes by nuclear resonance vibrational spectroscopy. *J. Am. Chem. Soc.* **132,** 6914–6916

40. Lewandowska, H., Kalinowska, M., Brzóska, K., Wójciuk, K., Wójciuk, G., and Kruszewski, M. (2011) Nitrosyl iron complexes: synthesis, structure and biology. *Dalton Trans* **40,** 8273–8289

Differentiated, Promoter-specific Response of [4Fe-4S] NsrR DNA Binding to Reaction with Nitric Oxide

Jason C. Crack, Dimitri A. Svistunenko, John Munnoch, Andrew J. Thomson, Matthew I. Hutchings and Nick E. Le Brun

Access the most updated version of this article at doi: 10.1074/jbc.M115.693192

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 40 references, 14 of which can be accessed free at
http://www.jbc.org/content/291/16/8663.full.html#ref-list-1

SCIENTIFIC REP🅾RTS

Characterization of a putative NsrR homologue in *Streptomyces venezuelae* reveals a new member of the Rrf2 superfamily

John T. Munnoch[1], Ma Teresa Pellicer Martinez[2], Dimitri A. Svistunenko[3], Jason C. Crack[2], Nick E. Le Brun[2] & Matthew I. Hutchings[1]

Members of the Rrf2 superfamily of transcription factors are widespread in bacteria but their functions are largely unexplored. The few that have been characterized in detail sense nitric oxide (NsrR), iron limitation (RirA), cysteine availability (CymR) and the iron sulfur (Fe-S) cluster status of the cell (IscR). In this study we combined ChIP- and dRNA-seq with *in vitro* biochemistry to characterize a putative NsrR homologue in *Streptomyces venezuelae.* ChIP-seq analysis revealed that rather than regulating the nitrosative stress response like *Streptomyces coelicolor* NsrR, Sven6563 binds to a conserved motif at a different, much larger set of genes with a diverse range of functions, including a number of regulators, genes required for glutamine synthesis, NADH/NAD(P)H metabolism, as well as general DNA/RNA and amino acid/protein turn over. Our biochemical experiments further show that Sven6563 has a [2Fe-2S] cluster and that the switch between oxidized and reduced cluster controls its DNA binding activity *in vitro.* To our knowledge, both the sensing domain and the putative target genes are novel for an Rrf2 protein, suggesting Sven6563 represents a new member of the Rrf2 superfamily. Given the redox sensitivity of its Fe-S cluster we have tentatively named the protein RsrR for Redox sensitive response Regulator.

Filamentous *Streptomyces* bacteria produce bioactive secondary metabolites that account for more than half of all known antibiotics as well as anticancer, anti-helminthic and immunosuppressant drugs[1,2]. More than 600 *Streptomyces* species are known and each encodes between 10 and 50 secondary metabolites but only 25% of these compounds are produced *in vitro.* As a result, there is huge potential for the discovery of new natural products from *Streptomyces* and their close relatives. This is revitalizing research into these bacteria and *Streptomyces venezuelae* has recently emerged as a new model for studying their complex life cycle, in part because of its unusual ability to sporulate to near completion when grown in submerged liquid culture. This means the different tissue types involved in the progression to sporulation can be easily separated and used for tissue specific analyses such as RNA sequencing and chromatin immunoprecipitation and sequencing (RNA- and ChIP-seq)[3,4]. *Streptomyces* species are complex bacteria that grow like fungi, forming a branching, feeding substrate mycelium in the soil that differentiates upon nutrient stress into reproductive aerial hyphae that undergo cell division to form spores[5]. Differentiation is closely linked to the production of antibiotics, which are presumed to offer a competitive advantage when nutrients become scarce in the soil.

Streptomyces bacteria are well adapted for life in the complex soil environment with more than a quarter of their ~9 Mbp genomes encoding one and two-component signaling pathways that allow them to rapidly sense and respond to changes in their environment[6]. They are facultative aerobes and have multiple systems for dealing with redox, oxidative and nitrosative stress. Most species can survive for long periods in the absence of O_2, most likely by respiring nitrate, but the molecular details are not known[7]. They deal effectively with nitric oxide (NO) generated either endogenously through nitrate respiration[7] or in some cases from dedicated bacterial NO synthase

[1]School of Biological Sciences, University of East Anglia, Norwich, Norwich Research Park, United Kingdom. [2]Centre for Molecular and Structural Biochemistry, School of Chemistry, University of East Anglia, Norwich, Norwich Research Park, United Kingdom. [3]School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, United Kingdom. Correspondence and requests for materials should be addressed to N.E.L.B. (email: n.le-brun@uea.ac.uk) or M.I.H. (email: m.hutchings@uea.ac.uk)
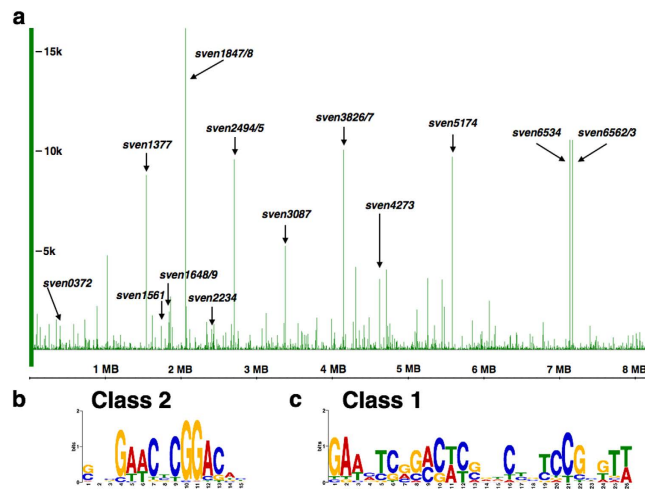
Figure 1. Defining the regulon and binding site for RsrR. Top panel (a) shows the whole genome ChIP-seq analysis with class 1 sites labeled in black. The frequency of each base sequenced is plotted with genomic position on the x-axis and frequency of each base sequenced on the y-axis for *S. venezualae* (NC_018750). Bottom panel (b) shows the class 1 and 2 web logos generated following MEME analysis of the ChIP-seq data.

(bNOS) enzymes[8] or by other NO generating organisms in the soil[9]. We recently characterized NsrR, which is the major bacterial NO stress sensor, in *Streptomyces coelicolor* (ScNsrR). NsrR is a dimeric Rrf2 family protein with one [4Fe-4S] cluster per monomer that reacts rapidly with up to eight molecules of NO[10,11]. Nitrosylation of the Fe-S cluster results in derepression of the *nsrR*, *hmpA1* and *hmpA2* genes[11], which results in transient expression of HmpA NO dioxygenase enzymes that convert NO to nitrate[12–14]. The Rrf2 superfamily of bacterial transcription factors is still relatively poorly characterized, but many have C-terminal cysteine residues that are known or predicted to coordinate Fe-S clusters. Other characterized Rrf2 proteins include RirA which senses iron limitation most likely through an Fe-S cluster[15] and IscR which senses the Fe-S cluster status of the cell[16].

In this work we report the characterization of the *S. venezuelae* Rrf2 protein Sven6563 that is annotated as an NsrR homologue. In fact, it shares only 27% primary sequence identity with ScNsrR and is not genetically linked to an *hmpA* gene (Supplementary Figure S1). We purified the protein from *E. coli* under anaerobic conditions and found that it is a dimer with each monomer containing a reduced [2Fe-2S] cluster that is rapidly oxidized but not destroyed by oxygen. Thus, the [2Fe-2S] cofactor is different to the [4Fe-4S] cofactors in the *S. coelicolor* and *Bacillus subtilis* NsrR proteins. The [2Fe-2S] cluster of Sven6563 switches easily between oxidized and reduced states and we provide evidence that this switch controls its DNA binding activity, with holo-RsrR showing highest affinity for DNA in its oxidised state. We have tentatively named the protein RsrR for Redox sensitive response Regulator. ChIP-seq and ChIP-exo analysis allowed us to define the RsrR binding sites on the *S. venezuelae* genome with RsrR binding to class 1 target genes with an 11-3-11 bp inverted repeat motif and class 2 target genes with a single repeat or half site. Class 1 target genes suggest a primary role in regulating NADH/NAD(P) H and glutamate/glutamine metabolism rather than nitrosative stress. The *sven6562* gene, which is divergent from *rsrR*, is the most highly induced transcript, up 5.41-fold (log2), in the Δ*rsrR* mutant and encodes a putative NAD(P)+ binding repressor in the NmrA family. Other class 1 target genes are not significantly affected by loss of RsrR suggesting additional levels of regulation, possibly including the divergently expressed Sven6562 (NmrA). Taken together our data suggest that RsrR is a new member of the Rrf2 family and extends the known functions of this superfamily, potentially sensing redox via a [2Fe-2S] cofactor in a mechanism that has only previously been observed in SoxR proteins.

Results

Identifying RsrR target genes in *S. venezualae*. We previously reported a highly specialized function for the NO-sensing NsrR protein in *S. coelicolor*. ChIP-seq against a 3xFlag-ScNsrR protein showed that it only regulates three genes, two of which encode NO dioxygenase HmpA enzymes, and the *nsrR* gene itself[11]. To investigate the function of RsrR, the putative NsrR homologue in *S. venezuelae*, we constructed an *S. venezuelae* Δ*rsrR* mutant expressing an N-terminally 3xFlag-tagged protein and performed ChIP-seq against this strain (accession number GSE81073). The sequencing reads from the wild-type (control) sample were subtracted from the experimental sample before ChIP peaks were called (Fig. 1a). Using an arbitrary cut-off of ≥500 sequencing reads we identified 117 enriched target sequences (Supplementary data S1). We confirmed these peaks by visual inspection of the data using Integrated Genome Browser[17] and used MEME[18] to identify a conserved motif in all 117 ChIP peaks (Fig. 1b). In 14 of the 117 peaks this motif is present as an inverted 11-3-11 bp repeat, which is characteristic of full-length Rrf2 binding sites[16,19], and we called these class 1 targets (Fig. 1c). In the other 103 peaks it is present as a single motif or half site and we call these class 2 targets (Fig. 1b). The divergent genes *sven3827/8* contain a single class 1 site and the 107 bp intergenic region between *sven6562* and *rsrR* contains two class 1 binding sites separated by a single base pair. It seems likely that RsrR autoregulates and also regulates the divergent *sven6562*, which encodes a LysR family regulator with an NmrA-type ligand-binding domain. These domains are predicted to sense redox poise by binding NAD(P)+ but not NAD(P)H[20]. The positions of the two

Flanking gene[a]		Distance[b]	Dist. TSS[c]	Fold change[e]	Annotation	Additional description
Left (−1)	Right (+1)					
	sven0372[d]	7	−99	−0.73	Two-component system histidine kinase	Involved in a two-component system signal transduction set
	sven0519[d]	**−993**		**0.53**	**Sulfate permease**	**Involved in sulfate uptake**
	sven0772	**−408**		**N/A**	**Xaa-Pro aminopeptidase**	**Peptidase releasing N-terminal amino acid next to a proline**
	sven1561[d]	103	36	−0.11	Glutamine synthase	Carries out the reaction: Glutamate + NH_4 −> Glutamine
	sven1670	17		−0.28	Pyridoxamine 5′-phosphate oxidase	Involved in steps of the vitamin B6 metabolism pathway
	sven1686	−41		N/A	Citrate lyase beta chain	—
	sven1847[d]	6		−0.89	3-oxoacyl-[acyl-carrier protein] reductase	Carries out: NADP+ dependant reduction of 3-oxoacyl-[ACP]
	sven1902	**−1643**	**−1689**	**−0.03**	**Glutamine synthase adenylyltransferase**	**Regulates glutamine synthase activity by adenylation**
	sven2494	90	0	−1.69	Hypothetical protein	—
	sven2540	**221**		**N/A**	**Glucose fructose oxidoreductase**	**D-glucose + D-fructose <> D-gluconolactone + D-glucitol**
	sven3087	51	51	−0.02	Acetyltransferase	Transfers an acetyl group
	sven3827[d]	26	−10	0.15	SAICAR synthetase	Involved in purine metabolism
sven3934		16		−0.21	Enhanced intracellular survival protein	—
	sven4022	**−772**		**−0.55**	**Hypothetical protein**	**NAD(P)-binding Rossmann-like domain**
	sven4273	5		0.01	NADH-ubiquinone oxidoreductase chain I	Involved in the electron transfer chain, binds a [4Fe-4S]
	sven5088	**−77**		**−0.15**	**Epimerase/dehydratase**	**NADH dependant isomerase enzyme**
	sven5174[d]	−119		−0.18	Quinone oxidoreductase	H_2 + menaquinone <> menaquinol
	sven6227	**73**		**−5.21**	**NADH-FMN oxidoreductase**	**$FMNH_2 + NAD^+ <> FMN + NADH + H^+$**
sven6534		−100		0.97	Trypsin-like peptidase domain	A serine protease that hydrolyses proteins
sven6562[d]	sven6563[d]	72, −35	36	5.41[F], N/A	*nmrA/rsrR*	DNA binding proteins, NADP/[2Fe-2S] binding

Table 1. Combined ChIP-Seq and RNA-Seq data for selected RsrR targets. [a]Genes flanking the ChIP-seq peak. [b]Distance to the translational start codon (bp). [c]Distance to the transcriptional start site (bp). [d]EMSA reactions have been carried out successfully and specifically on these targets. [e]Relative expression (Log2) fold change WT vs. *RsrR::apr* mutant. [f]Expression values defined for targets with >100 mapped reads. Class 2 targets are highlighted in red.

RsrR binding sites relative to the transcript start sites (TSS) of *sven6562* and *rsrR* suggests that RsrR represses transcription of both genes by blocking the RNA polymerase binding site (Supplementary Figure S2). Following investigation of RNA-seq expression data (Supplementary data S1) comparing the wild-type and Δ*rsrR* strains the only ChIP-seq associated class 1 target with a significantly altered expression profile is *sven6562* which is ~5.41-fold (log2) induced by loss of RsrR. We hypothesis that other class 1 targets for which we have RNA-seq data are not significantly affected because they are subject to additional levels of regulation, including perhaps by Sven6562 itself although this remains to be seen.

Other class 1 targets include the *nuo* (NADH dehydrogenase) operon *sven4265-78* (*nuoA-N*) which contains an internal class 1 site upstream of *nuoH,* the putative NADP+ dependent dehydrogenase Sven1847 and the quinone oxidoreductase Sven5173 which converts quinone and NAD(P)H to hydroquinone and NAD(P)+ (Table 1). These data suggest a role for RsrR in regulating NAD(P)H metabolism. In addition to the genes involved directly in NADH/NAD(P)H metabolism, class 2 targets include 21 putative transcriptional regulators, genes involved in both primary and secondary metabolism, RNA/DNA replication and modification genes, transporters (mostly small molecule), proteases, amino acid (particularly glutamate and glutamine) metabolism, and a large number of genes with of unknown function (Supplementary data S1).
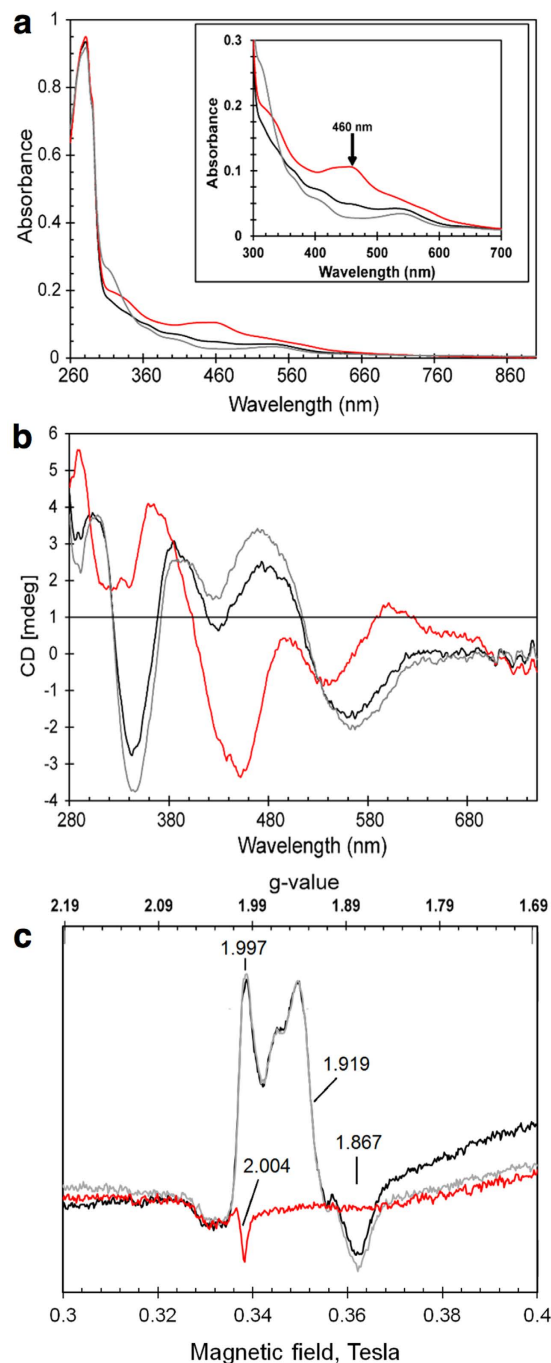
Figure 2. Spectroscopic characterization of RsrR. UV-visible absorption (**a**), CD (**b**) and EPR spectra (**c**) of 309 μM [2Fe-2S] RsrR (~75% cluster-loaded). Black lines – as isolated, red lines – oxidised, grey lines reduced proteins. In (**a,b**), initial exposure to ambient O_2 for 30 min was followed by 309 μM sodium dithionite treatment; in (**c**) – as isolated protein was first anaerobically reduced by 309 μM sodium dithionite and then exposed to ambient O_2 for 50 min. A 1 mm pathlength cuvette was used for optical measurements. Inset in (**a**) shows details of the iron-sulfur cluster absorbance in the 300–700 nm region.

Purified RsrR contains a redox active [2Fe-2S] cluster. The genes bound by RsrR do not include any NO detoxification genes and this suggested it is not an NsrR homologue but instead has an alternative function. To learn more about the protein we purified it from *E. coli* under strictly anaerobic conditions. The anaerobic RsrR solution is pink in colour but rapidly turns brown when exposed to O_2, suggesting the presence of a redox-active cofactor. Consistent with this, the UV-visible absorbance spectrum of the as-isolated protein revealed broad weak bands in the 300–640 nm region but following exposure to O_2, the spectrum changed significantly, with a more intense absorbance band at 460 nm and a pronounced shoulder feature at 330 nm (Fig. 2a). The form of the reduced and oxidised spectra are similar to those previously reported for [2Fe-2S] clusters that are coordinated
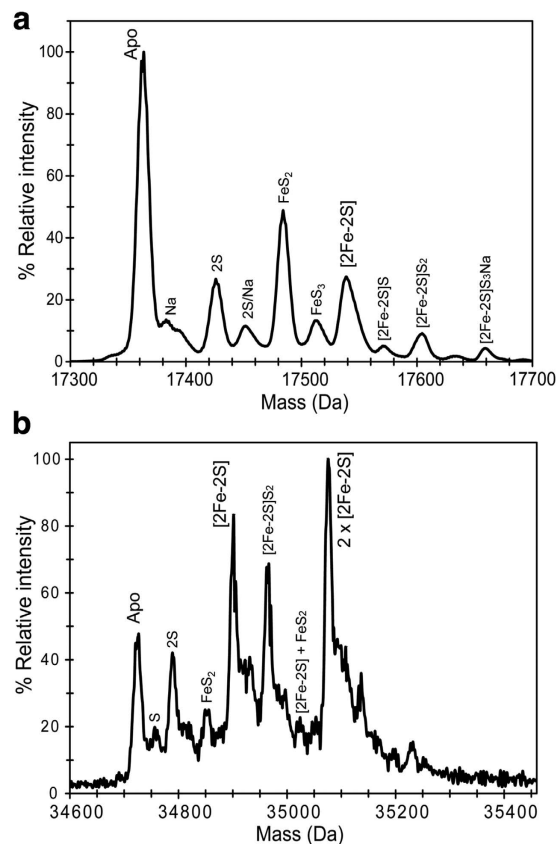
Figure 3. Native mass spectrometry of RsrR. (**a**,**b**) Positive ion mode ESI-TOF native mass spectrum of ~21 μM [2Fe-2S] RsrR in 250 mM ammonium acetate pH 8.0, in the RsrR monomer (**a**) and dimer (**b**) regions. Full m/z spectra were deconvoluted with Bruker Compass Data analysis with the Maximum Entropy plugin.

by three Cys residues and one His[21,22]. The anaerobic addition of dithionite to the previously air-exposed sample (at a 1:1 ratio with [2Fe-2S] cluster as determined by iron content) resulted in a spectrum very similar to that of the as-isolated protein (Fig. 2a), demonstrating that the cofactor undergoes redox cycling.

Because the electronic transitions of iron-sulfur clusters become optically active as a result of the fold of the protein in which they are bound, CD spectra reflect the cluster environment[23]. The near UV-visible CD spectrum of RsrR (Fig. 2b) for the as-isolated protein contained three positive (+) features at 303, 385 and 473 nm and negative (−) features at 343 and 559 nm. When the protein was exposed to ambient O_2 for 30 min, significant changes in the CD spectrum were observed, with features at (+) 290, 365, 500, 600 nm and (−) 320, 450 and 534 nm (Fig. 2b). The CD spectra are similar to those reported for Rieske-type [2Fe-2S] clusters[21,24,25], which are coordinated by two Cys and two His residues. Anaerobic addition of dithionite (1 equivalent of [2Fe-2S] cluster) resulted in reduction back to the original form (Fig. 2b) consistent with the stability of the cofactor to redox cycling.

The absorbance data above indicates that the cofactor is in the reduced state in the as-isolated RsrR protein. [2Fe-2S] clusters in their reduced state are paramagnetic (S = ½) and therefore should give rise to an EPR signal. The EPR spectrum for the as-isolated protein contained signals at g = 1.997, 1.919 and 1.867 (Fig. 2c). These g-values and the shape of the spectrum are characteristic of a $[2Fe-2S]^{1+}$ cluster. The addition of excess sodium dithionite to the as-isolated protein did not cause any changes in the EPR spectrum (Fig. 2c) indicating that the cluster was fully reduced as isolated. Exposure of the as-isolated protein to ambient O_2 resulted in an EPR-silent form, with only a small free radical signal typical for background spectra, consistent with the oxidation of the cluster to the $[2Fe-2S]^{2+}$ form (Fig. 2c), and the same result was obtained upon addition of the oxidant potassium ferricyanide (data not shown).

To further establish the cofactor that RsrR binds, native ESI-MS was employed. Here, a C-terminal His-tagged form of the protein was ionized in a volatile aqueous buffered solution that enabled it to remain folded with its cofactor bound. The deconvoluted mass spectrum contained several peaks in regions that corresponded to monomer and dimeric forms of the protein, (Supplementary Figure S3). In the monomer region (Fig. 3a), a peak was observed at 17,363 Da, which corresponds to the apo-protein (predicted mass 17363.99 Da), along with adduct peaks at +23 and +64 Da due to Na^+ (commonly observed in native mass spectra) and most likely two additional sulfurs (Cys residues readily pick up additional sulfurs as persulfides, respectively[26]. A peak was also observed at +176 Da, corresponding to the protein containing a [2Fe-2S] cluster. As for the apo-protein, peaks corresponding to Na^+ and sulfur adducts of the cluster species were also observed (Fig. 3a). A significant peak was also detected at +120 Da that corresponds to a break down product of the [2Fe-2S] cluster (from which one iron is missing, FeS_2).
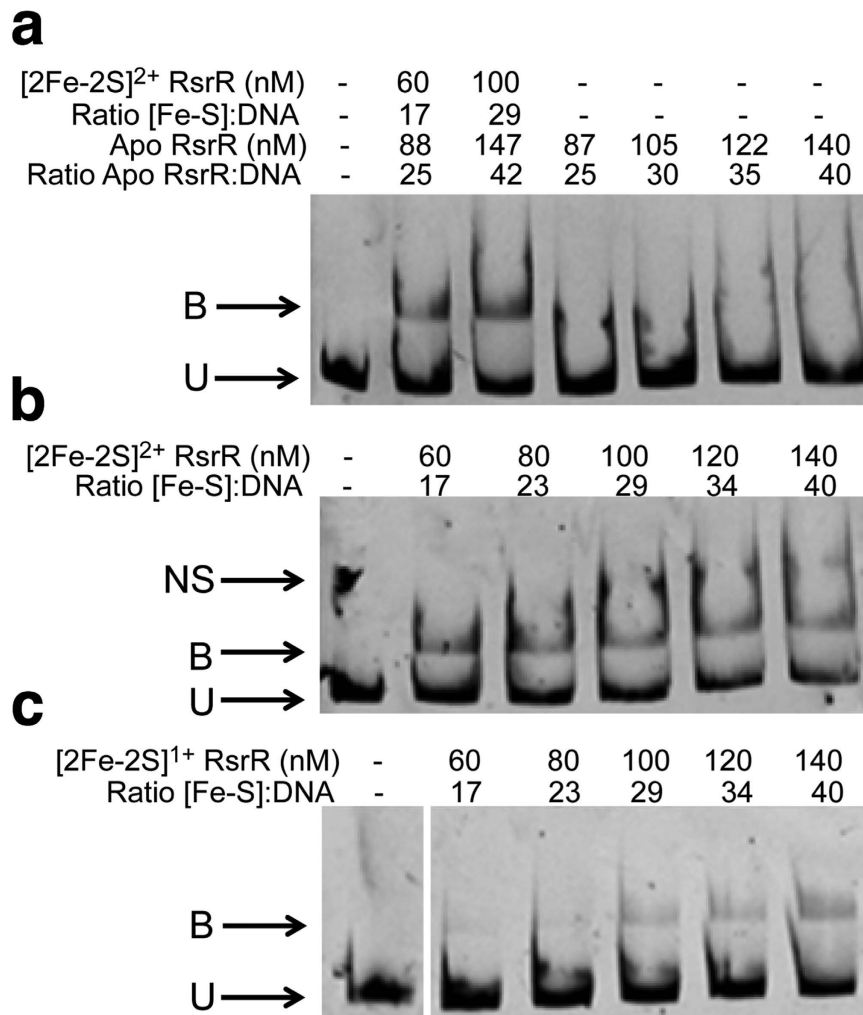
Figure 4. Cluster- and oxidation state-dependent DNA binding by [2Fe-2S] RsrR. EMSAs showing DNA probes unbound (U), bound (B), and non-specifically bound (NS) by (**a**) $[2Fe-2S]^{2+}$ and apo-RsrR (**b**) $[2Fe-2S]^{2+}$ RsrR and (**c**) $[2Fe-2S]^{1+}$ RsrR. Ratios of [2Fe-2S] containing RsrR (Holo) and [RsrR] (apo) to DNA are indicated for (**a**) while the concentration of [2Fe-2S] RsrR only is reported in (**b,c**). DNA concentration was 3.5 nM for the $[2Fe-2S]^{2+/1+}$ and apo-RsrR experiments. For (**a,b**) the reaction mixtures were separated at 30 mA for 50 min and the polyacrylamide gels were pre-run at 30 mA for 2 min prior to use. For (**c**) the reaction mixtures were separated at 30 mA for 1 h 45 min and the polyacrylamide gel was pre-run at 30 mA for 50 min prior to use using the de-gassed running buffer containing 5 mM sodium dithionite. For (**a**) both holo and apo protein concentrations are represented as the sample contained both forms due to incomplete cluster loading. The concentrations reported are of the [2Fe-2S] concentration.

In the dimer region, the signal to noise is significantly reduced but peaks are still clearly present (Fig. 3b). The peak at 34,726 Da corresponds to the RsrR homodimer (predicted mass 34727.98 Da), and the peak at +352 Da corresponds to the dimer with two [2Fe-2S] clusters. A peak at +176 Da is due to the dimer containing one [2Fe-2S] cluster. A range of cluster breakdown products similar to those detected in the monomer region were also observed (Fig. 3b). Taken together, the data reported here demonstrate that RsrR contains a [2Fe-2S] cluster that can be reversibly cycled between oxidised (+2) and reduced (+1) states.

Cluster and oxidation state dependent binding of RsrR *in vitro*. To determine which forms of RsrR are able to bind DNA, we performed EMSA experiments using the intergenic region between the highly enriched ChIP target *sven1847/8* as a probe. Increasing ratios of [2Fe-2S] RsrR to DNA resulted in a clear shift in the mobility of the DNA from unbound to bound, see Fig. 4a. Equivalent experiments with cluster-free (apo) RsrR did not result in a mobility shift, demonstrating that the cluster is required for DNA-binding activity. These experiments were performed aerobically and so the [2Fe-2S] cofactor was in its oxidised state. To determine if oxidation state affects DNA binding activity, EMSA experiments were performed with $[2Fe-2S]^{2+}$ and $[2Fe-2S]^{1+}$ forms of RsrR. The oxidised cluster was generated by exposure to air and confirmed by UV-visible absorbance. The reduced cluster was obtained by reduction with sodium dithionite, confirmed by UV-visible absorbance, and the reduced state was maintained using EMSA running buffer containing an excess of dithionite. The resulting EMSAs, Fig. 4b,c,
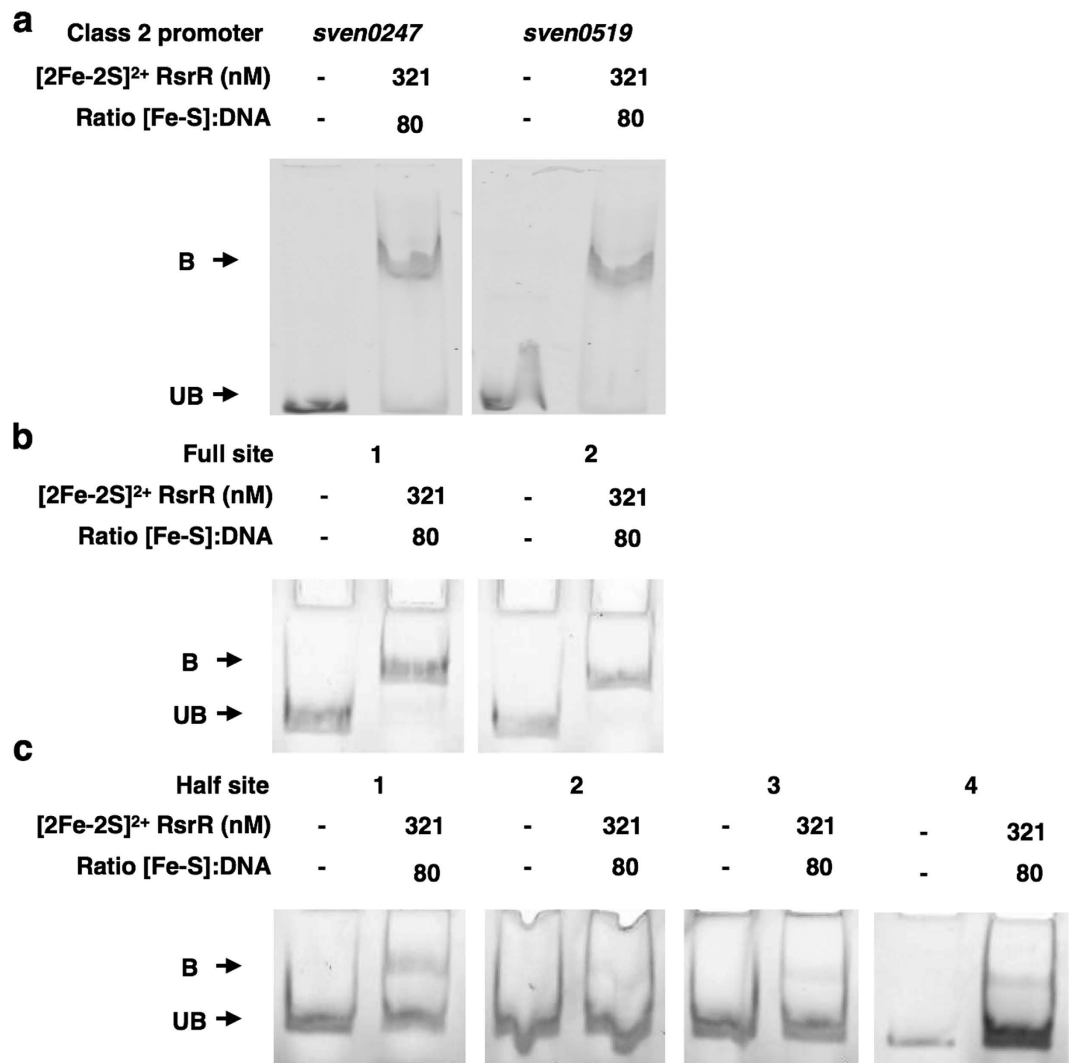
Figure 5. Oxidised RsrR binding to full site (class 1) and half site (class 2) RsrR targets. EMSAs showing DNA probes unbound (U) and bound (B) by [2Fe-2S]$^{2+}$. Ratios of [2Fe-2S] RsrR and [RsrR] to DNA are indicated. DNA concentration was 4 nM for each probe. EMSA's using class 2 DNA probes *sven0247* and *sven0519* (**a**), class 1 probes from the RsrR *rsrR* binding region (**b**) and the four possible half sites from the *rsrR* class 1 sites (**c**) were used. For (**a**) the reaction mixtures were separated at 30 mA for 1 h and the polyacrylamide gel was pre-run at 30 mA for 2 min prior to use. For (**b,c**) the reaction mixtures were separated at 30 mA for 30 min and the polyacrylamide gels were pre-run at 30 mA for 2 min prior to use. A representation of the *rsrR* promoter breakdown is also available in Supplementary Figure S3b.

show that DNA-binding occurred in both cases but that the oxidised form bound significantly more tightly. Tight binding could be restored to the reduced RsrR samples by allowing it to re-oxidise in air (data not shown). We cannot rule out that the apparent low affinity DNA binding observed for the reduced sample results from partial re-oxidation of the cluster during the electrophoretic experiment. Nevertheless, the conclusion is unaffected: oxidised, [2Fe-2S]$^{2+}$ RsrR is the high affinity DNA-binding form and these results suggest a change in the redox state of the [2Fe-2S] cluster controls the activity of RsrR, something which has only previously been observed for SoxR, a member of the MerR superfamily[27].

Oxidised [2Fe-2S] RsrR binds strongly to class 1 and 2 binding sites *in vitro*. To further investigate the DNA binding activities of [2Fe-2S]$^{2+}$ RsrR, EMSAs were performed on DNA probes containing the two class 2 RsrR binding sites at *sven0247* and *sven519* (Fig. 5a). Both probes were shifted by oxidized [2Fe-2S] RsrR showing that RsrR binds to both class 1 and 2 probes *in vitro*. To further test the idea of RsrR recognizing full and half sites, we constructed a series of probes based on the divergent *nmrA-rsrR* promoters carrying both or each individual natural class 1 sites (Fig. 5b) and artificial half sites (Fig. 5c). The combinations of artificial half sites are illustrated in Supplemental Figure S3 in regards to the original promoter region. The results show that RsrR binds strongly to both full class 1 binding sites at the *nmrA-rsrR* promoters (Fig. 5b) but only weakly to artificial half sites (Fig. 5c).

This suggests that although MEME only calls half sites in most of the RsrR target genes identified by ChIP-seq these class 2 targets must contain sufficient sequence information in the other half to enable strong binding by RsrR.

Mapping RsrR binding sites *in vivo* using ChIP-exo and differential RNA-seq.   MEME analysis of the ChIP-seq data detected only 14 class 1 (11-3-11 bp inverted repeat) sites out of the 117 target sites bound by RsrR on the *S. venezuelae* chromosome. However, ChIP-Seq and EMSAs show that RsrR can bind to target genes whether they contain class 1 or class 2 sites. This differs from *E. coli* NsrR which binds only weakly to target sites containing putative half sites (class 2)[28]. To gain more information about RsrR recognition sequences and the positions of these binding sites at target promoters we combined differential RNA-seq (dRNA-seq, accession number GSE81104), which maps the start sites of all expressed transcripts, with ChIP-exo (accession number GSE80818) which uses Lambda exonuclease to trim excess DNA away from ChIP complexes leaving only the DNA which is actually bound and protected by RsrR. For dRNA-seq, total RNA was prepared from cultures of wild type *S. venezuelae* and for the Δ*rsrR* mutant grown for 16 hours. ChIP-exo was performed on the Δ*rsrR* strain producing Flag-tagged RsrR, also at 16 hours. ChIP-exo identified 630 binding sites which included the 117 targets identified previously using ChIP-seq. The ChIP-exo peaks are on average only ~50 bp wide giving much better resolution of the RsrR binding sites at each target. MEME analysis using all 630 ChIP-exo sequences identified the class 2 binding motif in every sequence and we identified transcript start sites (TSS) for 261 of the 630 RsrR target genes using our dRNA-seq data (Supplementary data S1). Figure 6 shows a graphical representation of class 1 targets that have clearly defined TSS, indicating the centre of the ChIP peak, the associated TSS and any genes within the ~200 bp frame. Based on the RsrR binding site position at putative target genes RsrR likely acts as both a transcriptional activator and repressor and we have shown that RsrR represses transcription of *sven6562* which is a class 1 target with two 11-3-11 bp binding site in the intergenic region between *sven6562* and *rsrR*. The functional significance of RsrR binding to the other class 1 and 2 target genes identified here by ChIP-seq and ChIP-exo remains to be seen but they are not significantly affected by loss of RsrR under the conditions used in our experiments.

Discussion

In this work we have identified and characterized a new member of the Rrf2 protein family, which was mis-annotated as an NsrR homologue in the *S. venezuelae* genome. ChIP analyses show that RsrR binds to 630 sites on the *S. venezuelae* genome which compares to just three target sites for *S. coelicolor* NsrR and their DNA recognition sequences are very different. RNA-seq data shows a dramatic 5.3 fold (log2) change in the expression of the divergent gene from *rsrR*, *sven6562*, but under normal laboratory conditions no other direct RsrR targets are significantly induced or repressed by loss of RsrR. Approximately 2.7% of the RsrR targets contain class 1 binding sites which consist of a MEME identified 11-3-11 bp inverted repeat. Class 1 target genes include *sven6562* and are involved in either signal transduction and/or NAD(P)H metabolism which perhaps points to a link to redox poise and recycling of NAD(P)H to NAD(P) *in vivo*. The >600 class 2 target genes contain only half sites with a single repeat but exhibit strong binding by RsrR *in vitro*. Our EMSA experiments show that RsrR binds weakly to artificial half sites and this suggests additional sequence information is present at class 2 binding sites that increases the strength of DNA binding by RsrR. Six of the class 2 targets are involved in glutamate and glutamine metabolism including: *sven1561*, encoding a Glutamine Synthase (GS) that carries out the ATP dependent conversion of glutamate and ammonium to glutamine[29], *sven1902*, encoding a GS adenylyltransferase that carries out the adenylation and deadenylation of GS, reducing or increasing GS activity respectively[30]. *sven3711*, encoding a protein which results in the liberation of glutamate from glutamine[31]. *sven4418*, encoding a glutamine fructose-6-phosphate transaminase that carries out the reaction: L-glutamine and D-fructose 6-phosphate to L-glutamate and D-glucosamine 6-phosphate[32]. *sven4888*, encoding a glutamate-1-semialdehyde aminotransferase, which carries out the PLP dependent, reversible reaction of L-glutamate to 1-semialdehyde 5-aminolevulinate[33]. Finally, *sven7195*, encoding a glutamine-dependent asparagine synthase which carries out the ATP dependent transfer of NH_3 from glutamine to aspartate, forming glutamate and asparagine[34]. Glutamate and glutamine are precursors for the production of mycothiol, the actinobacterial equivalent of glutathione, which acts as a cellular reducing agent. Mycothiol also acts as a cellular reserve of cysteine and in the detoxification of redox species and antibiotics[35]. Glutamate is important, as a non-essential amino acid, because it links nitrogen and carbon metabolism in bacteria[36]. Additionally, glutamate acts as a proton sink through its decarboxylation to GABA, which especially under acidic conditions, favorably removes protons from the intracellular milieu[37].

Our data show that the purified RsrR protein contains a [2Fe-2S] cluster, which is stable in the presence of O_2 and can be reversibly cycled between reduced (+1) and oxidized (+2) states. The oxidised $[2Fe-2S]^{2+}$ form binds strongly to both class 1 and class 2 binding sequences *in vitro*, whereas the reduced $[2Fe-2S]^{1+}$ form exhibits significantly weaker binding. The binding we did observed is likely due to partial oxidation of the RsrR Fe-S cluster during the EMSA electrophoresis. The cluster free form of RsrR does not bind to DNA at all. Given these observations and the stability of the Fe-S cluster to aerobic conditions, we propose that the activity of RsrR is modulated by the oxidation state of its cluster, becoming activated for DNA binding through oxidation and inactivated through reduction. Exposure to O_2 is sufficient to cause oxidation, but other oxidants may also be important *in vivo*. The properties of RsrR described here are reminiscent of the *E. coli* [2Fe-2S] cluster containing transcription factor SoxR, which controls the regulation of another regulator, SoxS, through the oxidation state of its cluster[38].

Due to the number of RsrR regulated transcription factors it is likely that its target genes are subject to multiple levels of regulation. For example, the *sven6562* gene, which is divergent from *rsrR*, encodes a LysR family regulator with an N terminal NmrA-type NAD(P)+ binding domain. NmrA proteins are thought to control redox poise in fungi by sensing the levels of NAD(P), which they can bind, and NAD(P)H, which they cannot[39]. This is intriguing and we propose a model in which reduction of holo-RsrR induces expression of Sven6562 which
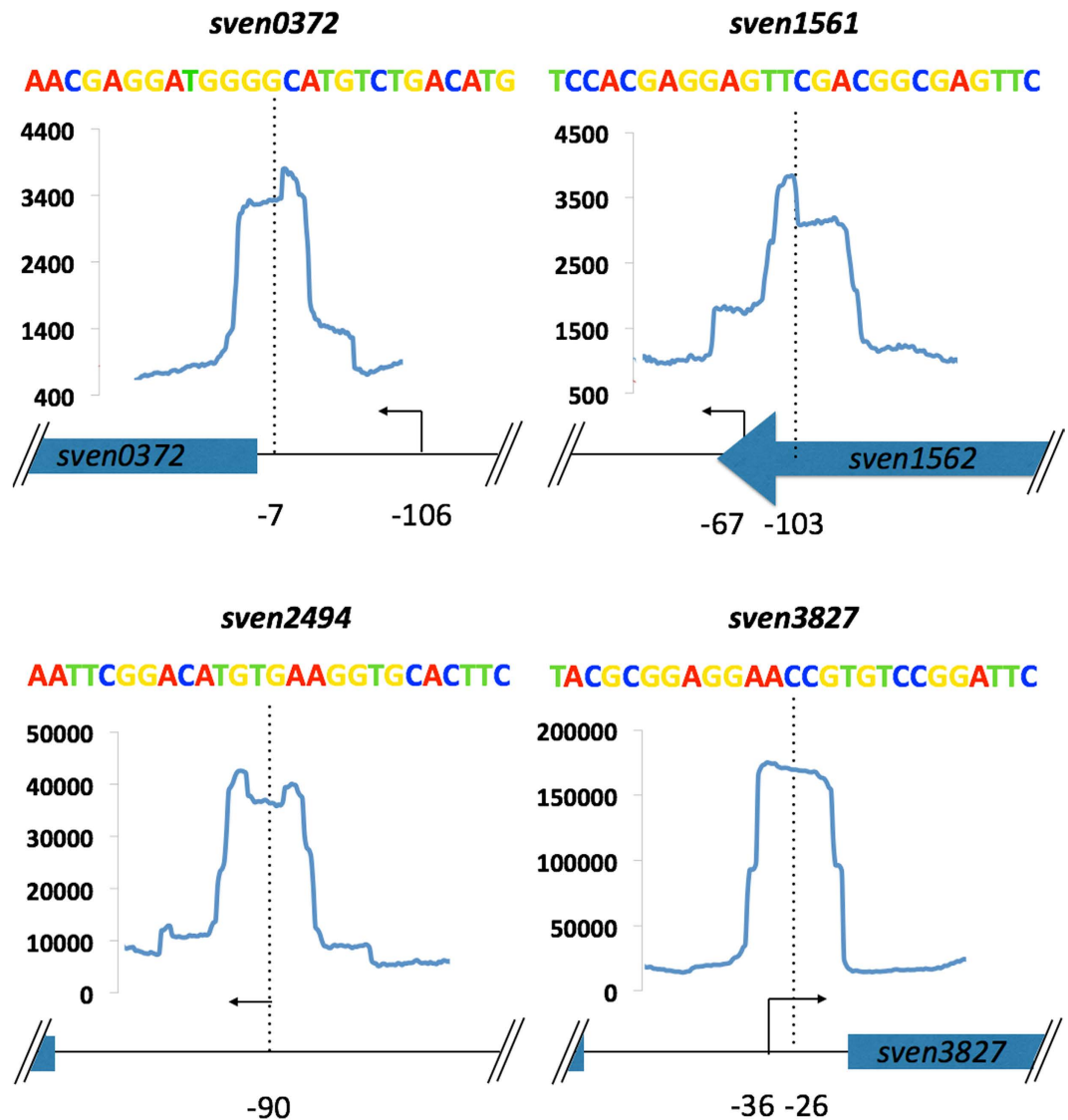
Figure 6. Graphical representation of combined ChIP-Seq, ChIP-exo and dRNA-seq for four class 1 targets. Each target has the relative position of ChIP-exo (blue line) peak centre (dotted line) and putative transcriptional start site (TSS - solid arrow) indicated with the distance in bp (black numbers) relative to the down stream start codon of target genes. The y-axis scale corresponds to number of reads for ChIP data with each window corresponding to 200 bp with each ChIP-peak being ~50 bp wide. Above each is the relative binding site sequence coloured following the weblogo scheme (A – red, T – green, C – blue and G – yellow) from the MEME results.

in turn senses redox poise via the ratio of NAD(P)+/NAD(P)H/NAD(P)H and then modulates expression of its own regulon which likely overlaps with that of RsrR. Clearly there is much to learn about this system and it will be important to define the role of Sven6562 in *S. venezuelae* in the future. We did not observe any phenotype for the Δ*rsrR* mutant and it is no more sensitive to redox active compounds or oxidative stress than wild-type *S. venezuelae* (not shown). However, this is not surprising given the number of systems in bacteria that deal with reactive nitrogen and oxygen species and redox stress. In *Streptomyces* species these include catalases, peroxidases[40] and superoxide dismutases[41] and associated regulators such as OxyR[42], SigR[43], OhrR[44], Rex[20] and SoxR[45]. Thus, our data suggests Sven6563, tentatively renamed here as RsrR, is a new member of the Rrf2 family and this work extends our knowledge about this neglected but widespread superfamily of bacterial transcription factors.

Materials and Methods

Bacterial strains, plasmids, oligonucleotides and growth conditions. Bacterial strains and plasmids are listed in Table 2 and oligonucleotides are listed in Table 3. For ChIP-seq experiments, *S. venezuelae* strains were grown at 30 °C in MYM liquid sporulation medium[46] made with 50% tap water and supplemented with 200 μl trace element solution[47] per 100 ml and adjusted to a final pH of 7.3. Disruption of *rsrR* was carried out following the PCR-targeting method[48] as described previously[49,50]. Primers JM0109 and JM0110 were used to PCR amplify the apramycin disruption cassette from pIJ773. Cosmid SV-5-F05 was used as the template

Strain/plasmid	Description	Source
E. coli		
TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ$^-$	Invitrogen
BW25113 (pIJ790)	*E. coli* BW25113 containing λRED recombination plasmid pIJ790	48,58
ET12567 (pUZ8002)	*E. coli* Δdam dcm strain containing helper plasmid pUZ8002	59,60
BL21	F$^-$ ompT gal dcm lon hsdS$_B$(r$_B^-$ m$_B^-$) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	61
Streptomyces		
S. venezualae	*S. venezuelae* ATCC 10712 WT strain	4
rsrR::apr	*S. venezuelae* with a ReDirect disrupted *sven6563::apr*	
rsrR::apr 3xFlag RsrR	*rsrR::apr* with a pMS82 encoded N-terminal 3xFlag tagged *rsrR* with 300 bp of upstream flanking DNA (promoter)	This work
Plasmids		
pIJ773	pBluescript KS (+), aac(3)IV, oriT (RK2), FRT sites	48
SV-5-F05	Supercos-1-cosmid with (a 52181 bp) fragment containing *sven6562/3*	4
pMS82	*ori,* pUC18, *hyg, oriT,* RK2, int ΦBT1	62
pGS-21a	Genscript overexpression and purification vector (SD0121)	Genscript
pJM026	SV-5-F05 containing *sven6563::apr oriT*	This work
pJM027	pMS82, *rsrR* gene plus 300 bp upstream DNA with a c-terminal synthetic linker and 3xFLAG tag	This work
pJM028	pGS-21a, full length *rsrR* cloned NdeI/XhoI	This work
pJM029	pJM028 with a c-terminal 6xHis tag NdeI/XhoI	This work
pJM030	pJM028 with a c-terminal synthetic linker as with (flag), 2xFLAG tag and a 6xHis tag, cloned NdeI/XhoI	This work

Table 2. Strains and plasmids used during this study.

cosmid. The disruption cosmid (pJM026) was checked by PCR using primers JM0111 and JM0112. Antibiotic marked, double crossover exconjugants, were identified as previously described and confirmed once more with JM0111 and JM0112. The 3x Flag tag copy of *rsrR* was synthesized by Genescript and subcloned into pMS82 using HindIII/KpnI and confirmed by PCR using primers JM0113 and JM0114.

ChIP (chromatin immunoprecipitation) – seq and exo. ChIP-Seq was carried out as previously described[51] with the below modifications. A 3xFlag tagged RsrR was used as with our previous work[11]. Following sonication and lysate clearing M2 affinity beads (Sigma-Aldrich #A2220) were prepared by washing in ½IP buffer following manufacturers instructions. The cleared lysate was incubated with 40 μl of washed M2 beads and incubated for 4 h at 4C in a vertical rotor. The lysate was removed and the beads pooled into one 1.5 microfuge tube and washed in 0.5 IP buffer. The beads were transferred to a fresh microfuge tube and washed a further 3 times removing as much buffer as possible without disturbing the beads. The DNA-protein complex was eluted from the beads with 100 μl elution buffer (50 mM Tris-HCl pH7.6, 10 mM EDTA, 1% SDS) by incubating at 65 °C overnight. Removing the ~100 μl elution buffer, an extra 50 μl of elution buffer was added and further incubated at 65 °C for 5 min. To extract the DNA 150 μl eluate, 2 μl proteinase K (10 mg/ml) was added and incubated 1.5 h at 55 °C. To the reaction 150 μl phenol-chloroform was added. Samples were vortexed and centrifuged at full speed for 10 min. The aqueous layer was extracted and purified using the Qiaquick column from Qiagen with a final elution using 50 μl EB buffer (Qiagen). The concentration of samples were determined using Quant-iT™ PicoGreen® dsDNA Reagent (Invitrogen) or equivalent kit or by nanodrop measurement. DNA sequencing of ChIP-Seq samples was carried out by GATC Biotech. ChIP-exo following sonication of lysates was carried out by Peconic LLC (State College, PA) adding an additional exonuclease treatment to the process as previously described[52]. Data analysis was carried out using CLC workbench 8 followed by a manual visual inspection of the data.

dRNA - seq. Mycelium was harvested at experimentally appropriate time points and immediately transferred to 2 ml round bottom tubes, flash frozen in liquid N$_2$, stored at −80 °C or used immediately. All apparatus used was treated with RNaseZAP (Sigma) to remove RNases for a minimum of 1 h before use. RNaseZAP treated mortar and pestles were used, the pestle being placed and cooled on a mixture of dry ice and liquid N$_2$ with liquid N$_2$ being poured into the bowl and over the mortar. Once the bowl had cooled the mycelium samples were added directly to the liquid N$_2$ and thoroughly crushed using the mortar leaving a fine powder of mycelium. Grindings were transferred to a pre-cooled 50 ml Falcon tube and stored on dry ice. Directly to the tube, 2 ml of TRI reagent (Sigma) was added to the grindings and mixed. Samples are then thawed while vortexing intermittently at room temperature for 5–10 min until the solution cleared. To 1 ml of TRI reagent resuspension, 200 μl of chloroform was added and vortexed for 15 seconds at room temperature then centrifuged for 10 min at 13,000 rpm. The upper, aqueous phase (clear colourless layer) was removed into a new 2 ml tube. The remainder of the isolation protocol follows the RNeazy Mini Kit (Qiagen) instructions carrying out both on and off column DNase treatments. On column treatments were carried out following the first RW1 column wash. DNaseI (Qiagen) was added (10 μl enzyme, 70 μl RDD buffer) to the column and stored at RT for 1 h. The column was washed again with RW1 then treated as described in the manufacturer's instructions. Once eluted from the column, samples

Name	Description	Sequence
JM0062	M13_Fwd sequence labelled with 6′Fam for EMSA reactions using M13Fam nested primers	CTAAAACGACGGCCAGT
JM0063	M13_Rev sequence labelled with 6′Fam for EMSA reactions using M13Fam nested primers	CAGGAAACAGCTATGAC
JM0109	RsrR (*sven6563*) forward disruption primer (Redirect)	CCAGTCCCCTCCCCCACGGACCTGCTGCGTCGCACCATGATTCCGGGGATCCGTCGACC
JM0110	RsrR (*sven6563*) reverse disruption primer (Redirect)	CACCGAACAGCCAAGCCCCCCTCAGCAAGCCCTCCCTCATGTAGGCTGGAGCTGCTTC
JM0111	RsrR (*sven6563*) forward test primer	ACGCGGCGACCACGTCGTGG
JM0112	RsrR (*sven6563*) reverse test primer	GCCCGTACGGTAGACCGCCG
JM0113	pMS82 cloning forward test primer	GCAACAGTGCCGTTGATCGTGCTATG
JM0114	pMS82 cloning reverse test primer	GCCAGTGGTATTTATGTCAACACCGCC
JM0117	M13Fam nested *sven1847* for primer sequence for EMSA reactions	CTAAAACGACGGCCAGTTCTCCTCGCCCGCCCCGTCG
JM0118	M13Fam nested *sven1847* rev primer sequence for EMSA reactions	CAGGAAACAGCTATGACCCGTCCGGCGCCCCGGGTGG
JM0119	M13Fam nested *sven3827* for primer sequence for EMSA reactions	CTAAAACGACGGCCAGTCTCGCCCACTCGCCGTACCG
JM0120	M13Fam nested *sven3827* rev primer sequence for EMSA reactions	CAGGAAACAGCTATGACCATCACGAGATCGCCCGCCT
JM0121	M13Fam nested *sven4273* for primer sequence for EMSA reactions	CTAAAACGACGGCCAGTGAGAACATCGCCTTCGGCAA
JM0122	M13Fam nested *sven4273* rev primer sequence for EMSA reactions	CAGGAAACAGCTATGACGCGGGGCGCCGTCGTCTTCT
JM0123	M13Fam nested *sven5174* for primer sequence for EMSA reactions	CTAAAACGACGGCCAGTCGCGTTCCGGACCCGTACAAAGAAT
JM0124	M13Fam nested *sven5174* rev primer sequence for EMSA reactions	CAGGAAACAGCTATGACACCTGAATCTCGCATGACCCTCCGA
JM0125	M13Fam nested *sven0372* for primer sequence for EMSA reactions	CTAAAACGACGGCCAGTTGGTGACCGGGTCCGAACGGTCCGTAA
JM0126	M13Fam nested *sven0372* rev primer sequence for EMSA reactions	CAGGAAACAGCTATGACAACAGGGAGAGCTGGTCGACCATCC
JM0127	M13Fam nested *sven1561* for primer sequence for EMSA reactions	CTAAAACGACGGCCAGTCCCAGCTACGAGGTGGCGAAGCAGG
JM0128	M13Fam nested *sven1561* rev primer sequence for EMSA reactions	CAGGAAACAGCTATGACGGTCTGGGTGTCGAAGAAGGTGGTG
JM0129	M13Fam nested *sven6563* for primer sequence for EMSA reactions	CTAAAACGACGGCCAGTTCGTCGAAGGTCGGGGAGTT
JM0130	M13Fam nested *sven6563* rev primer sequence for EMSA reactions	CAGGAAACAGCTATGACCGTGCAGCTCAGCGAGCCGG
JM0131	M13Fam nested *sven0247* for primer sequence for EMSA reactions	CTAAAACGACGGCCAGTTCGTCATGATCGTGTGGCGGCTGCG
JM0132	M13Fam nested *sven0247* rev primer sequence for EMSA reactions	CAGGAAACAGCTATGACAGCACCAGCCGCTCGTCGAACGCGG
JM0133	M13Fam nested *sven0519* for primer sequence for EMSA reactions	CTAAAACGACGGCCAGTAGACGATGATCAACGTGAAGGTGTCCG
JM0134	M13Fam nested *sven0519* rev primer sequence for EMSA reactions	CAGGAAACAGCTATGACAAGGTCGCGACGCACACCATGATCAT
JM0141	M13Fam nested *sven6562/3* Site 1-4 primer sequence for EMSA reactions	CTAAAACGACGGCCAGTCAAACTCGGATACCCGATGTCCGAGATAATACTCGGATAGTCTGTGTCCGAGTCAAGTCATAGCTGTTTCCTG
JM0142	M13Fam nested *sven6562/3* Site 1-2 primer sequence for EMSA reactions	CTAAAACGACGGCCAGTGCAAACTCGGATACCCGATGTCCGAGATAATGTCATAGCTGTTTCCTG
JM0143	M13Fam nested *sven6562/3* Site 3-4 primer sequence for EMSA reactions	CTAAAACGACGGCCAGTTAATACTCGGATAGTCTGTGTCCGAGTCAAAGTCATAGCTGTTTCCTG
JM0144	M13Fam nested *sven6562/3* Site 1 primer sequence for EMSA reactions	CTAAAACGACGGCCAGTGCAAACTCGGATACCCGGTCATAGCTGTTTCCTG
JM0145	M13Fam nested *sven6562/3* Site 2 primer sequence for EMSA reactions	CTAAAACGACGGCCAGTCCGATGTCCGAGATAATGTCATAGCTGTTTCCTG
JM0146	M13Fam nested *sven6562/3* Site 3 primer sequence for EMSA reactions	CTAAAACGACGGCCAGTTAATACTCGGATAGTCTGTCATAGCTGTTTCCTG
JM0147	M13Fam nested *sven6562/3* Site 4 primer sequence for EMSA reactions	CTAAAACGACGGCCAGTTCTGTGTCCGAGTCAAAGTCATAGCTGTTTCCTG

Table 3. List of primers used in this study. Primers JM0119-JM0134 were used to produce EMSA DNA templates that were successfully shifted using purified RsrR and mentioned in the text but the data is not shown as part of the work.

were treated using TURBO DNA-free Kit (Ambion) following manufacturer's instructions to remove residual DNA contamination.

RNA-seq was carried out by vertis Biotechnologie. Data analysis was carried out using the Tuxedo protocol[53] for analysis of gene expression and TSSAR webservice for dRNA transcription start site analysis[54]. In addition a manual visual processing approach was carried out for each.

Purification of RsrR. L Luria-Bertani medium (10×500 mL) was inoculated with freshly transformed BL21 (DE3) *E. coli* containing a pGS-21a vector with the *prsrR-His* insert. 100 μg/mL ampicillin and 20 μM ammonium ferric citrate were added and the cultures were grown at 37 °C, 200 rpm until $OD_{600\,nm}$ was 0.6–0.9. To facilitate *in vivo* iron-sulfur cluster formation, the flasks were placed on ice for 18 min, then induced with 100 μM IPTG and incubated at 30 °C and 105 rpm. After 50 min, the cultures were supplemented with 200 μM ammonium ferric citrate and 25 μM L-Methionine and incubated for a further 3.5 h at 30 °C. The cells were harvested by centrifugation at $10000 \times g$ for 15 min at 4 °C. Unless otherwise stated, all subsequent purification steps were performed under anaerobic conditions inside an anaerobic cabinet ($O_2 < 2$ ppm). Cells pellets were resuspended in 70 mL of buffer A (50 mM TRIS, 50 mM $CaCl_2$, 5% (v/v) glycerol, pH 8) and placed in a 100 mL beaker. 30 mg/mL of lysozyme and 30 mg/mL of PMSF were added and the cell suspension thoroughly homogenized by syringe, removed from the anaerobic cabinet, sonicated twice while on ice, and returned to the anaerobic cabinet. The cell suspension

was transferred to O-ring sealed centrifuge tubes (Nalgene) and centrifuged outside of the cabinet at 40,000 × g for 45 min at 1 °C.

The supernatant was passed through a HiTrap IMAC HP (1 × 5 mL; GE Healthcare) column using an ÄKTA Prime system at 1 mL/min. The column was washed with Buffer A until $A_{280\,nm} < 0.1$. Bound proteins were eluted using a 100 mL linear gradient from 0 to 100% Buffer B (50 mM TRIS, 100 mM $CaCl_2$, 200 mM L- Cysteine, 5% glycerol, pH 8). A HiTrap Heparin (1 × 1 mL; GE Healthcare) column was used to remove the L- Cysteine, using buffer C (50 mM TRIS, 2 M NaCl, 5% glycerol, pH 8) to elute the protein. Fractions containing RsrR-His were pooled and stored in an anaerobic freezer until needed. RsrR-His protein concentrations were determined using the method of Bradford (Bio-Rad Laboratories)[55], with BSA as the standard. Cluster concentrations were determined by iron assay[56], from which an extinction coefficient, ε, at 455 nm was determined as $3450 \pm 25\,M^{-1}\,cm^{-1}$, consistent with values reported for [2Fe-2S] clusters with His coordination[21].

Preparation of Apo- RsrR.

Apo-RsrR -His was prepared from as isolated holoprotein by aerobic incubation with 1 mM EDTA overnight.

Spectroscopy and mass spectrometry.

UV-visible absorbance measurements were performed using a Jasco V500 spectrometer, and CD spectra were measured with a Jasco J810 spectropolarimeter. EPR measurements were performed at 10 K using a Bruker EMX EPR spectrometer (X-band) equipped with a liquid helium system (Oxford Instruments). Spin concentrations in the protein samples were estimated by double integration of EPR spectra with reference to a 1 mM Cu(II) in 10 mM EDTA standard. For native MS analysis, His-tagged RsrR was exchanged into 250 mM ammonium acetate, pH 8, using PD10 desalting columns (GE Life Sciences), diluted to ~21 μM cluster and infused directly (0.3 mL/h) into the ESI source of a Bruker micrOTOF-QIII mass spectrometer (Bruker Daltonics, Coventry, UK) operating in the positive ion mode. Full mass spectra (m/z 700–3500) were recorded for 5 min. Spectra were combined, processed using the ESI Compass version 1.3 Maximum Entropy deconvolution routine in Bruker Compass Data analysis version 4.1 (Bruker Daltonik, Bremen, Germany). The mass spectrometer was calibrated with ESI-L low concentration tuning mix in the positive ion mode (Agilent Technologies, San Diego, CA).

Electrophoretic Mobility Shift Assays (EMSAs).

DNA fragments carrying the intergenic region between *sven1847* and *sven1848 of the S. venzualae* chromosome were PCR amplified using *S. venezualae* genomic DNA with 5′ 6-FAM modified primers (Table 2). The PCR products were extracted and purified using a QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. Probes were quantitated using a NanoDrop ND2000c. The molecular weights of the double stranded FAM labelled probes were calculated using OligoCalc[57].

EMSA reactions (20 μl) were carried out on ice in 10 mM Tris, 60 mM KCl, pH 7.52. Briefly, 1 μL of DNA was titrated with varying aliquots of RsrR. 2 μL of loading dye (containing 0.01% (w/v) bromophenol blue), was added and the reaction mixtures were immediately separated at 30 mA on a 5% (w/v) polyacrylamide gel in 1 X TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA), using a Mini Protean III system (Bio-Rad). Gels were visualized (excitation, 488 nm; emission, 530 nm) on a molecular imager FX Pro (Bio-Rad). Polyacrylamide gels were pre-run at 30 mA for 2 min prior to use. For investigations of [2Fe-2S]$^{1+}$ RsrR DNA binding, in order to maintain the cluster in the reduced state, 5 mM of sodium dithionite was added to the isolated protein and the running buffer (de-gassed for 50 min prior to running the gel). Analysis by UV-visible spectroscopy confirmed that the cluster remained reduced under these conditions.

References

1. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **75,** 311–335 (2012).
2. Challis, G. L. & Hopwood, D. a. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc. Natl. Acad. Sci. USA* **100,** 14555–14561 (2003).
3. Glazebrook, M. a, Doull, J. L., Stuttard, C. & Vining, L. C. Sporulation of *Streptomyces venezuelae* in submerged cultures. *J. Gen. Microbiol.* **136,** 581–588 (1990).
4. Pullan, S. T., Chandra, G., Bibb, M. J. & Merrick, M. Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. *BMC Genomics* **12,** 175 (2011).
5. Flärdh, K. & Buttner, M. J. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.* **7,** 36–49 (2009).
6. Rodríguez, H., Rico, S., Díaz, M. & Santamaría, R. I. Two-component systems in *Streptomyces*: key regulators of antibiotic complex pathways. *Microb. Cell Fact.* **12,** 127 (2013).
7. van Keulen, G., Alderson, J., White, J. & Sawers, R. G. The obligate aerobic actinomycete *Streptomyces coelicolor* A3(2) survives extended periods of anaerobic stress. *Environ. Microbiol.* **9,** 3143–3149 (2007).
8. Johnson, E. G. *et al.* Plant-pathogenic *Streptomyces* species produce nitric oxide synthase-derived nitric oxide in response to host signals. *Chem. Biol.* **15,** 43–50 (2008).
9. Sasaki, Y. *et al.* Nitrogen oxide cycle regulates nitric oxide levels and bacterial cell signaling. *Sci. Rep.* **6,** 22038 (2016).
10. Crack, J. C. *et al.* Differentiated, promoter-specific response of [4Fe-4S] NsrR DNA-binding to reaction with nitric oxide. *J. Biol. Chem.* **291,** 8663–8672 (2016).
11. Crack, J. *et al.* NsrR from Streptomyces coelicolor is a Nitric Oxide-Sensing [4Fe-4S] Cluster Protein with a Specialized Regulatory Function. *J. Biol. Chem.* **290,** 12689–12704 (2015).
12. Gardner, P. R. *et al.* Hemoglobins dioxygenate nitric oxide with high fidelity. *J. Inorg. Biochem.* **100,** 542–550 (2006).
13. Poole, R. K. & Hughes, M. N. New functions for the ancient globin family: Bacterial responses to nitric oxide and nitrosative stress. *Mol. Microbiol.* **36,** 775–783 (2000).
14. Forrester, M. T. & Foster, M. W. Protection from nitrosative stress: a central role for microbial flavohemoglobin. *Free Radic. Biol. Med.* **52,** 1620–1633 (2012).
15. Hibbing, M. E. & Fuqua, C. Antiparallel and interlinked control of cellular iron levels by the Irr and RirA regulators of Agrobacterium tumefaciens. *J. Bacteriol.* **193,** 3461–3472 (2011).

16. Santos, J. a., Pereira, P. J. B. & Macedo-Ribeiro, S. What a difference a cluster makes: The multifaceted roles of IscR in gene regulation and DNA recognition. *Biochim. Biophys. Acta - Proteins Proteomics* 1–12, doi: 10.1016/j.bbapap.2015.01.010 (2015).

17. Nicol, J. W., Helt, G. A., Blanchard, S. G., Raja, A. & Loraine, A. E. The Integrated Genome Browser: Free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **25,** 2730–2731 (2009).

18. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37,** W202–W208 (2009).

19. Partridge, J. D., Bodenmiller, D. M., Humphrys, M. S. & Spiro, S. NsrR targets in the *Escherichia coli* genome: New insights into DNA sequence requirements for binding and a role for NsrR in the regulation of motility. *Mol. Microbiol.* **73,** 680–694 (2009).

20. Brekasis, D. & Paget, M. S. B. A novel sensor of NADH/NAD+ redox poise in *Streptomyces coelicolor* A3 (2). *EMBO J.* **22,** 4856–4865 (2003).

21. Kimura, S., Kikuchi, A., Senda, T., Shiro, Y. & Fukuda, M. Tolerance of the Rieske-type [2Fe-2S] cluster in recombinant ferredoxin BphA3 from *Pseudomonas* sp. KKS102 to histidine ligand mutations. *Biochem. J.* **388,** 869–878 (2005).

22. Lin, J., Zhou, T., Ye, K. & Wang, J. Crystal structure of human mitoNEET reveals distinct groups of iron sulfur proteins. *Proc. Natl. Acad. Sci. USA* **104,** 14640–14645 (2007).

23. Stephens, P. J. *et al.* Circular dichroism and magnetic circular dichroism of iron-sulfur proteins. *Biochemistry* **17,** 4770–4778 (1978).

24. Link, T. A. *et al.* Comparison of the 'Rieske' [2Fe-2S] center in the bc1 complex and in bacterial dioxygenases by circular dichroism spectroscopy and cyclic voltammetry. *Biochemistry* **35,** 7546–7552 (1996).

25. Couture, M. M. J. *et al.* Characterization of BphF, a Rieske-type ferredoxin with a low reduction potential. *Biochemistry* **40,** 84–92 (2001).

26. Zhang, B. *et al.* Reversible cycling between cysteine persulfide-ligated [2Fe-2S] and cysteine-ligated [4Fe-4S] clusters in the FNR regulatory protein. *Proc. Natl. Acad. Sci.* **109,** 15734–15739 (2012).

27. Fujikawa, M., Kobayashi, K. & Kozawa, T. Direct oxidation of the [2Fe-2S] cluster in SoxR protein by superoxide: Distinct differential sensitivity to superoxide-mediated signal transduction. *J. Biol. Chem.* **287,** 35702–35708 (2012).

28. Chhabra, S. & Spiro, S. Inefficient translation of *nsrR* constrains behavior of the NsrR regulon in *Escherichia coli*. *Microbiology* **161,** 2029–2038 (2015).

29. Gill, H. S. & Eisenberg, D. The crystal structure of phosphinothricin in the active site of glutamine synthetase illuminates the mechanism of enzymatic inhibition. *Biochemistry* **40,** 1903–1912 (2001).

30. Jiang, P., Mayo, A. E. & Ninfa, A. J. *Escherichia coli* glutamine synthetase adenylyltransferase (ATase, EC 2.7.7.49): Kinetic characterization of regulation by PII, PII-UMP, glutamine, and α-ketoglutarate. *Biochemistry* **46,** 4133–4146 (2007).

31. Huang, Y. T. *et al.* *In vitro* characterization of enzymes involved in the synthesis of nonproteinogenic residue (2S,3S)-B-methylphenylalanine in glycopeptide antibiotic mannopeptimycin. *Chem Bio Chem* **10,** 2480–2487 (2009).

32. Yamazaki, K. In *Handb. Glycosyltransferases Relat. Genes* 1465–1479 (2014).

33. Grimm, B. Primary structure of a key enzyme in plant tetrapyrrole synthesis: glutamate 1-semialdehyde aminotransferase. *Proc. Natl. Acad. Sci. USA* **87,** 4169–4173 (1990).

34. Tesson, A. R., Soper, T. S., Ciustea, M. & Richards, N. G. J. Revisiting the steady state kinetic mechanism of glutamine-dependent asparagine synthetase from *Escherichia coli*. *Arch. Biochem. Biophys.* **413,** 23–31 (2003).

35. Newton, G. L. & Fahey, R. C. Regulation of mycothiol metabolism by sigma(R) and the thiol redox sensor anti-sigma factor RsrA. *Mol. Microbiol.* **68,** 805–809 (2008).

36. Berg, J., Tymoczko, J. & Stryer, L. *Biochemistry* (2007).

37. Feehily, C. & Karatzas, K. A. G. Role of glutamate metabolism in bacterial responses towards acid and other stresses. *J. Appl. Microbiol.* **114,** 11–24 (2013).

38. Lee, K. L., Singh, A. K., Heo, L., Seok, C. & Roe, J. H. Factors affecting redox potential and differential sensitivity of SoxR to redox-active compounds. *Mol. Microbiol.*, doi: 10.1111/mmi.13068 (2015).

39. Lamb, H. K. *et al.* The negative transcriptional regulator NmrA discriminates between oxidized and reduced dinucleotides. *J. Biol. Chem.* **278,** 32107–32114 (2003).

40. Mishra, S. & Imlay, J. Why do bacteria use so many enzymes to scavenge hydrogen peroxide? *Arch. Biochem. Biophys.* **525,** 145–160 (2012).

41. Youn, H. D., Kim, E. J., Roe, J. H., Hah, Y. C. & Kang, S. O. A novel nickel-containing superoxide dismutase from *Streptomyces spp*. *Biochem. J.* **318** (Pt 3), 889–896 (1996).

42. Hahn, J., Oh, S. & Roe, J. Role of OxyR as a peroxide-sensing positive regulator in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **184,** 5214–5222 (2002).

43. Kim, M. S. *et al.* Conservation of thiol-oxidative stress responses regulated by SigR orthologues in actinomycetes. *Mol. Microbiol.* **85,** 326–344 (2012).

44. Oh, S. Y., Shin, J. H. & Roe, J. H. Dual role of OhrR as a repressor and an activator in response to organic hydroperoxides in *Streptomyces coelicolor*. *J. Bacteriol.* **189,** 6284–6292 (2007).

45. Shin, J. H., Singh, A. K., Cheon, D. J. & Roe, J. H. Activation of the SoxR regulon in *Streptomyces coelicolor* by the extracellular form of the pigmented antibiotic actinorhodin. *J. Bacteriol.* **193,** 75–81 (2011).

46. Stuttard, C. Temperate Phages of *Streptomyces venezuelae*: Lysogeny and Host Specificity Shown by Phages SV1 and SV2. *Microbiology* **128,** 115–121 (1982).

47. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. Practical Streptomyces Genetics. *John Innes Cent. Ltd.* **529,** doi: 10.4016/28481.01 (2000).

48. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. 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 (2003).

49. Thompson, B. J. *et al.* Investigating lipoprotein biogenesis and function in the model Gram-positive bacterium *Streptomyces coelicolor*. *Mol. Microbiol.* **77,** 943–957 (2010).

50. Hutchings, M. I., Hong, H.-J. & Buttner, M. J. The vancomycin resistance VanRS two-component signal transduction system of *Streptomyces coelicolor*. *Mol. Microbiol.* **59,** 923–935 (2006).

51. Al-Bassam, M. M., Bibb, M. J., Bush, M. J., Chandra, G. & Buttner, M. J. Response regulator heterodimer formation controls a key stage in *Streptomyces* development. *PLos Genet.* **10,** e1004554 (2014).

52. Reja, R., Vinayachandran, V., Ghosh, S. & Pugh, B. F. Molecular mechanisms of ribosomal protein gene coregulation. *Genes Dev* **29,** 1942–1954 (2015).

53. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7,** 562–578 (2012).

54. Amman, F. *et al.* TSSAR: TSS annotation regime for dRNA-seq data. *BMC Bioinformatics* **15,** 89 (2014).

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

56. Crack, J., Green, J., Le Brun, N. & Thomson, A. Detection of Sulfide Release from the Oxygen-sensing [4Fe-4S] Cluster of FNR. *J. Biol. Chem.* **281,** 18909–18913 (2006).

57. Kibbe, W. A. OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Res.* **35,** 43–46 (2007).

58. Datsenko, K. a & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97,** 6640–6645 (2000).

59. MacNeil, D. J. *et al.* Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111,** 61–68 (1992).
60. Kang, J. G. *et al.* RsrA, an anti-sigma factor regulated by redox change. *EMBO J*. **18,** 4292–4298 (1999).
61. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189,** 113–130 (1986).
62. Gregory, M. A., Till, R. & Smith, M. C. M. Integration Site for *Streptomyces* Phage φBT1 and Development of Site-Specific Integrating Vectors. *Society* **185,** 5320–5323 (2003).

Acknowledgements

Author Contributions

J.T.M. carried out all of the molecular microbiology experiments, some of the biochemical experiments, analysed the data and co-wrote the manuscript. M.T.P.C. carried out the bulk of the biochemical experiments, analysed the data and co-wrote the manuscript. J.C.C. analyzed data and co-wrote the manuscript. D.A.S. performed EPR experiments and analysed data. N.E.L.B. and M.I.H. conceived and coordinated the study, analyzed data and co-wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Munnoch, J. T. *et al.* Characterization of a putative NsrR homologue in *Streptomyces venezuelae* reveals a new member of the Rrf2 superfamily. *Sci. Rep.* **6**, 31597; doi: 10.1038/srep31597 (2016).

SCIENTIFIC REP⚙RTS

Cosmid based mutagenesis causes genetic instability in *Streptomyces coelicolor*, as shown by targeting of the lipoprotein signal peptidase gene

John T. Munnoch[1], David A. Widdick[1,2], Govind Chandra[2], Iain C. Sutcliffe[3], Tracy Palmer[4] & Matthew I. Hutchings[1]

Bacterial lipoproteins are extracellular proteins tethered to cell membranes by covalently attached lipids. Deleting the lipoprotein signal peptidase (*lsp*) gene in *Streptomyces coelicolor* results in growth and developmental defects that cannot be restored by reintroducing *lsp*. This led us to hypothesise that *lsp* is essential and that the *lsp* mutant we isolated previously had acquired compensatory secondary mutations. Here we report resequencing of the genomes of wild-type M145 and the *cis*-complemented Δ*lsp* mutant (BJT1004) to map and identify these secondary mutations but we show that they do not increase the efficiency of disrupting *lsp* and are not *lsp* suppressors. We provide evidence that they are induced by introducing the cosmid St4A10Δ*lsp,* as part of ReDirect PCR mutagenesis protocol, which transiently duplicates a number of important cell division genes. Disruption of *lsp* using a suicide vector (which does not result in gene duplication) still results in growth and developmental delays and we conclude that loss of Lsp function results in developmental defects due to the loss of all lipoproteins from the cell membrane. Significantly, our results also indicate the use of cosmid libraries for the genetic manipulation of bacteria can lead to phenotypes not necessarily linked to the gene(s) of interest.

Bacterial lipoproteins are essential for building and maintaining the cell envelope and they also provide a key interface with the external environment[1–3]. Most lipoprotein precursors are exported as unfolded polypeptides via the Sec (general secretory) pathway but others can be exported via the twin arginine transport (Tat) pathway, which is typically utilised for the transport of fully folded proteins[4–6]. The signal peptides of lipoproteins closely resemble other types of bacterial Sec and Tat signal peptide but they contain a characteristic lipobox motif, typically L_{-3}-A/S_{-2}-G/A_{-1}-C_{+1}, relative to the signal cleavage site, in which the cysteine residue is essential and invariant. The lipobox motif allows putative lipoproteins to be easily identified in bacterial genome sequences[3,7]. Following translocation, lipoprotein precursors are firstly modified by covalent attachment of a diacylglycerol molecule, derived from a membrane phospholipid, to the thiol of the conserved lipobox cysteine residue via a thioether linkage. This reaction is catalysed by an enzyme named Lgt (Lipoprotein diacylglycerol transferase) and results in a diacylated lipoprotein. Lsp (Lipoprotein signal peptidase) then cleaves the signal sequence immediately upstream of the lipidated cysteine to leave it at the +1 position. These early steps in lipoprotein biogenesis are highly conserved and unique to bacteria making them potential targets for antibacterial drug development[2,8]. In Gram-negative bacteria and Gram-positive Actinobacteria, lipoproteins can be further modified by addition of an amide-linked fatty acid to the amino group of the diacylated cysteine residue at the mature N-terminus. This final step is catalysed by the enzyme Lnt (Lipoprotein n-acyltransferase) and results in triacylated lipoproteins. In Gram-negative proteobacteria, Lnt modification is a pre-requisite for the recognition of lipoproteins by the Lol

[1]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, UK. [2]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich, UK. [3]Department of Applied Sciences, Northumbria University, Newcastle, UK. [4]Division of Molecular Microbiology, College of Life Sciences, University of Dundee, Dundee, UK. Correspondence and requests for materials should be addressed to M.I.H. (email: M.hutchings@uea.ac.uk)
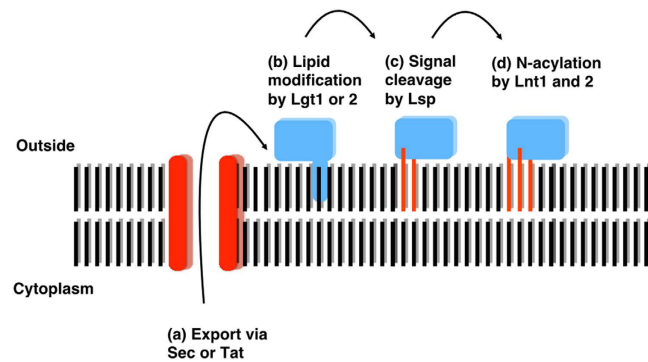
Figure 1. Lipoprotein biogenesis in *Streptomyces coelicolor*. Approximately 80% of precursor lipoproteins in *S. coelicolor* are translocated via the general secretory (Sec) pathway with around 20% being translocated by the twin arginine transport (Tat) pathway (**a**). Following translocation across the cytoplasmic membrane they are diacylated on the thiol of the lipobox (+1) cysteine residue by Lgt1 or Lgt (**b**) and then the signal sequence is cleaved by Lsp immediately upstream of that modified cysteine (**c**). Lnt1 then adds a third acyl chain to the amino group on the +1 cysteine to produce a triacylated lipoprotein (**d**). Lnt2 is not essential for triacylation *in vitro* but appears to increase its efficiency. The function of the N-acyl modification is not yet known.

machinery, which transports lipoproteins to the outer membrane[2,9] but its function in monoderm Gram-positive bacteria is not known. Members of the Gram-positive phyla Firmicutes and Mollicutes also N-acylate lipoproteins despite lacking Lnt homologues and *S. aureus* can diacylate or triacylate individual lipoproteins in an environmentally dependent manner[10–14]. These studies suggest that triacylation of lipoproteins in Gram-positive bacteria has an important role in their natural environment but is dispensable *in vitro*. Loss of Lnt activity in *Streptomyces* bacteria has no obvious effect on fitness or lipoprotein localisation *in vitro* but it does have a moderate effect on virulence in the plant pathogen *Streptomyces scabies*, supporting the idea that it has environmental importance[15].

We previously characterised all four steps of the lipoprotein biogenesis pathway in *Streptomyces* spp. (Fig. 1)[5,15], which is one of the best studied genera in the Gram-positive phylum Actinobacteria. Our key findings are (i) that Tat exports ~20% of lipoprotein precursors in streptomycetes; (ii) they N-acylate lipoproteins using two non-essential Lnt enzymes; (iii) *Streptomyces coelicolor* encodes two functional copies of Lgt which cannot be removed in the same strain; (iv) *lsp* mutants can be isolated at low frequencies but they acquire spontaneous secondary mutations which might be *lsp* suppressors. It was recently reported that Lgt is essential in *Mycobacterium tuberculosis*, which is also a member of the phylum Actinobacteria, and that *lgt* deletion in the fast-growing species *Mycobacterium smegmatis* is accompanied by spontaneous secondary mutations[16]. Natural product antibiotics that target the lipoprotein biogenesis pathway include globomycin, made by *Streptomyces globisporus*[2] and antibiotic TA made by *Myxococcus xanthus*[1,16]. Both inhibit Lsp activity and are lethal to *Escherichia coli* but TA resistance arises through spontaneous *IS3* insertion into the *lpp* gene, which encodes an abundant lipoprotein that attaches the *E. coli* outer membrane to the peptidoglycan cell wall[16,17]. Over-expressing *lsp* also confers TA resistance in both *E. coli* and *M. xanthus*, and the latter encodes additional Lsp homologues within the TA biosynthetic gene cluster[17].

Deletion of *S. coelicolor lsp* results in very small and flat colonies that are delayed in sporulation and these *lsp* mutants could not be fully complemented even by reintroducing the *lsp* gene to its native locus. Although both *cis* and *in trans* complementation restored lipoprotein biogenesis and sporulation it did not restore the wild-type growth rate[5]. There are two likely reasons for this: either *lsp* is essential and Δ*lsp* mutants acquire secondary suppressor mutations or the ReDirect PCR targeting method that we used to delete the *lsp* gene resulted in mutations independent of *lsp*. ReDirect is the name given to a protocol in which the Lambda Red system is used to PCR target genes of interest in a *Streptomyces* cosmid library in *E. coli* and the mutated cosmids are conjugated into *Streptomyces* species to select for mutants. Here we provide evidence to support the second hypothesis by demonstrating that introduction of a cosmid carrying an ~40 kb region of the *S. coelicolor* chromosome, including *lsp*, from *E. coli* to *S. coelicolor* leads to growth and developmental defects. We further show that *lsp* is non-essential but that deletion of the *lsp* gene results in small colonies that over-produce actinorhodin, as observed previously. These phenotypes must therefore be due to the loss of lipoproteins from the cytoplasmic membrane of *S. coelicolor*.

Results

Mapping secondary mutations in the *cis* complemented Δ*lsp* strain BJT1004.

We previously reported that the *S. coelicolor* Δ*lsp* mutant BJT1001 cannot be complemented even by restoring *lsp* to its native locus[5]. Since *cis* complementation should effectively restore the genome to wild-type this suggests that other spontaneous mutations have occurred during the deletion of *lsp*. To test this we resequenced the genomes of the isogenic parent strain *S. coelicolor* M145 and the *cis*-complemented Δ*lsp* strain BJT1004 using two independent companies and compared them with each other and with the published M145 sequence to identify mutations (Supplementary Table S1). Across all four sequences we identified a total of 51 single nucleotide polymorphisms (SNPs) as well as a chromosomal rearrangement in BJT1004 that is not present in the parent strain M145 (Fig. 2a,b and Supplementary Text S1 and S2). Of the 51 SNPs, 13 are unique to one of the BJT1004 sequences,
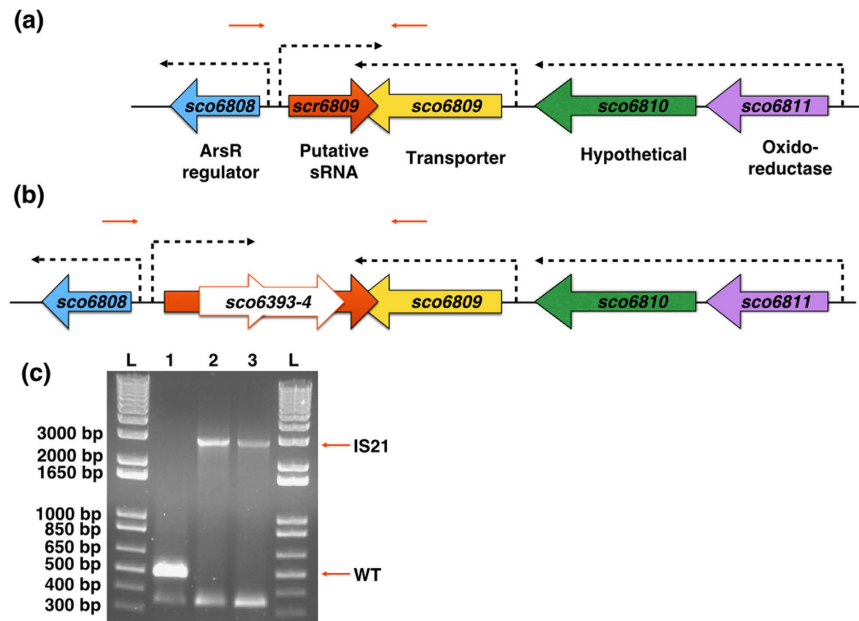
Figure 2. *IS21* **insertion into** *scr6809.* The *sco6811-08* region of the *S. coelicolor* M145 genome contains 4 genes (*sco6811* (Purple), *sco6810* (green), *sco6809* (yellow) and *sco6809* (blue) and a putative sRNA *scr6809* (large red) along with 3 putative promoters (broken arrows). Representations of the WT loci (**a**) and that sequenced from BJT1004 (**b**) indicate where an IS21 element (*sco6393* and sco6394) has inserted within *scr6809*. PCR verification of the IS21 insertion with primers JM0093 and JM0094 (small red arrows) was carried out (**c**) using M145 (lane 1), BJ1001 (lane 2) and BJT1004 (lane 3) genomic DNA. Lanes marked L contain the size ladders (Invitrogen 1 kb plus DNA ladder), lane 1 contains the PCR product using wild-type M145 DNA (514 bp), lane 2 contains the PCR product using Δ*lsp* strain BJT1001 DNA and lane 3 contains the PCR product using genomic DNA from the cis complemented Δ*lsp* strain BJT1004 (both 2884 bp.

with four residing inside coding sequences. Only one SNP occurs in both BJT1004 sequences and this is in the intergenic region between *sco5331* and *sco5332* and does not affect a coding sequence. In the single chromosomal rearrangement, the *IS21* insertion element (*sco6393-4*) has inserted into the intergenic region between *sco6808* and *sco6809* and this was confirmed by PCR (Fig. 2a–c). Although this could affect *sco6808* expression, deletion of *sco6808* has no effect on growth or development under standard laboratory conditions (Fig. 3a). This, and the intergenic position of *IS21* in BJT1004, led us to hypothesise that *IS21* might disrupt a non-coding RNA. Examination of RNA sequence data for *S. coelicolor* M145 confirmed the presence of a 189 nt transcript initiating 107 bp upstream of *sco6808* and reading into the last 82 nucleotides of the *sco6809* gene (data from GSM1121652 and GSM1121655 RNA sequencing; Supplementary Text S1 and 2; Fig. 2a,b). Following convention we named this putative small RNA *scr6809* for *S. coelicolor* RNA 6809. Deletion of *scr6809*, without disrupting the coding sequences of either *sco6808* or *sco6809*, resulted in pleiotropic effects, including colonies that look like wild-type and colonies defective in growth, aerial hyphae formation, sporulation or antibiotic overproducers. Restreaking Δ*scr6809* colonies with wild-type appearance gave rise to a range of colony morphologies, including growth and developmental defects (Fig. 3b). Δ*scr6809* colonies that were already defective in growth, development or antibiotic production maintained those phenotypes after restreaking. A previous report showed that a Δ*sco6808* mutant exhibits accelerated production of actinorhodin and undecylprodigiosin as well as precocious spore formation on R5 medium[18]. There was no observable difference between the wild-type and Δ*sco6808* strains under the growth conditions used here but disruption of *sco6808* in strain BJT1004 improved sporulation most likely because *scr6809* has been restored to its genome (Fig. 3a).

To determine whether *IS21* disruption of *scr6809* is induced by deletion of *lsp*, we isolated ten more non-clonal *lsp* mutants. Of these 10 mutants, both single cross-over (n = 7) and double cross-over (n = 3) strains were isolated following the introduction of cosmid St4A10Δ*lsp* into wild-type M145. Microscopy images of these strains showed a range of colony morphologies (Fig. 4a) and PCR results of the *lsp* loci confirm their genotypes, either single or double cross-over (Fig. 4b). These results suggest that secondary mutations are occurring in single cross-over strains containing intact *lsp* genes. The intergenic region between *sco6808* and *sco6809* was amplified to determine if these results were due to the disruption of *scr6809*. The size of the PCR products matched the predicted wild-type size for each strain and indicated that none of these *lsp* mutants contain an *IS21* insertion suggesting that the original observation is not specific to *lsp* mutants (Fig. 4c). Consistent with this conclusion, the frequency with which *lsp* mutants could be isolated was not increased in BJT1004 relative to the wild-type strain suggesting that none of the mapped mutations in BJT1004 are *lsp* specific suppressors. All attempts to over-express *scr6809* in *S. coelicolor* M145, *S. scabies* 87-22 and *S. venezualae* ATCC 10712 resulted in no observable phenotype but as the same vectors failed to complement the Δ*scr6809* strain, this suggests that functional *scr6809* was not expressed from vector pJM017. Cumulatively these results show that deletion of *lsp* is not solely
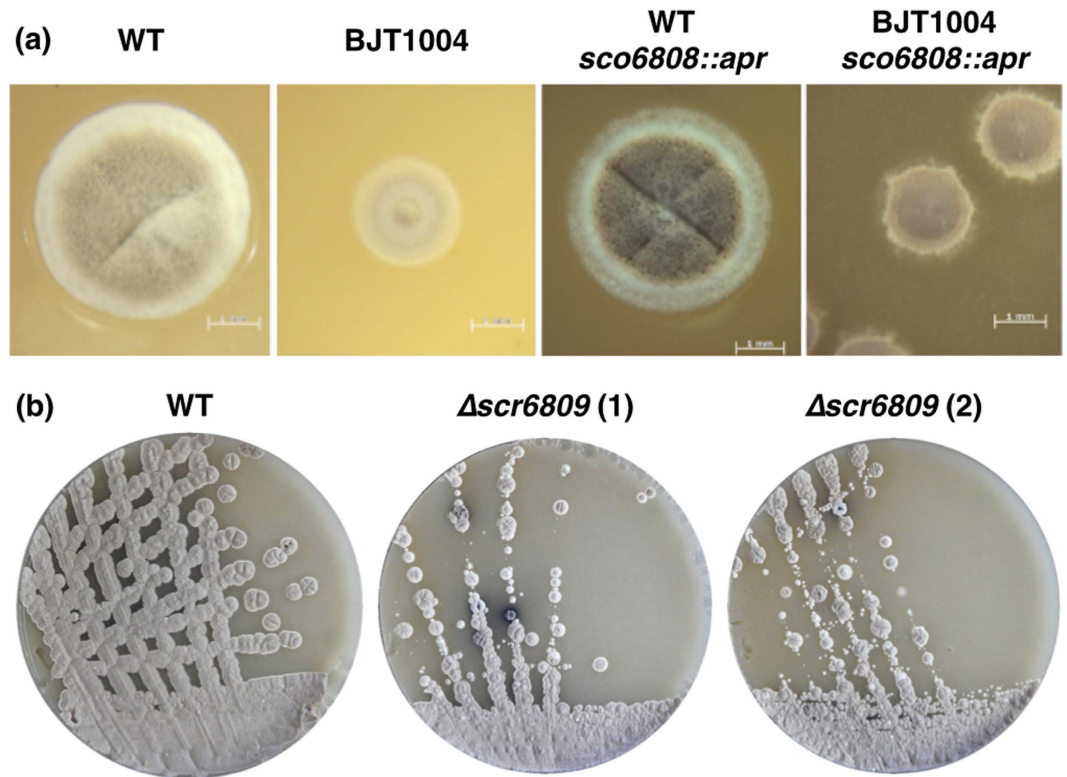
Figure 3. Analysis of the *IS21* disrupted genomic region in *S. coelicolor*. Colony morphology (**a**) shows that deletion of *sco6808* has no obvious effect on growth or development in wild-type M145 but does partially restore sporulation in BJT1004 (recovery of *scr6809*). Disruption of *scr6809* in M145 results in a range of pleiotropic morphological and developmental phenotypes (**b**).

responsible for the secondary mutations and this led us to hypothesise that these mutations accumulate as a result of duplicating cell division genes when we introduced cosmid St4A10Δ*lsp*. These results also suggest a role for *scr6809* in development, but this was not pursued in this work.

Introduction of wild-type St4A10 results in a pleiotropic phenotype. The ReDirect PCR-targeting method[19,20] used to delete the *lsp* gene utilized cosmid St4A10, which contains a ~40 kb region of the *S. coelicolor* genome spanning genes *sco2069–2104* (Supplementary Table S2). Conjugation of St4A10Δ*lsp* into *S. coelicolor* transiently duplicates all the genes on that cosmid (except *lsp*) and because this region includes cell division genes (*ftsZ*, *ftsQ*, *ftsW*, *ftsI* and *ftsL*) and essential cell wall synthesis genes (*murG*, *murD*, *murX*, *murF* and *murE*) we reasoned that transient over-expression of these genes, rather than deletion of *lsp*, is responsible for at least some of the spontaneous secondary mutations and the resulting pleiotropic phenotype. To test this idea we introduced an origin of transfer into the wild-type St4A10 cosmid backbone and then conjugated this cosmid into wild-type *S. coelicolor* M145. We used growth in the presence of kanamycin to select for single cross-over events where the whole cosmid is integrated into the chromosome, thus duplicating the *S. coelicolor* genes on St4A10. Analysis of these single cross-over strains revealed them to be genetically unstable, with many initially appearing similar to the observed Δ*lsp* phenotype, i.e. forming small colonies that are delayed in sporulation (Fig. 5). However, they do not over-produce the blue antibiotic actinorhodin which was an obvious characteristic of *S. coelicolor* Δ*lsp*. In addition, colonies arising from the M145::St4A10 strain acquired more significant developmental issues upon prolonged maintenance and restreaking onto MS agar containing kanamycin (not shown). This suggests that this strain accumulates spontaneous secondary mutations as a direct result of carrying two copies of the genes on cosmid St4A10 and also supports our hypothesis that the observed Δ*lsp* phenotype is at least in part due to duplication of the genes on cosmid St4A10. This is consistent with the fact that complementation of Δ*lsp* restored lipoprotein biogenesis and all detectable lipoproteins to the cell membrane but did not restore wild-type colony morphology[5].

Targeted deletion of *lsp* results in a small colony phenotype. To test how much deletion of *lsp* gene alone contributed to the phenotype of BJT1001 (the Δ*lsp* strain generated using ReDirect) we undertook a targeted disruption of *lsp* in wild-type *S. coelicolor* M145 using a suicide vector, which does not duplicate or affect any other coding sequences. The *lsp* suicide vector, pJM016 (Table 1), was introduced into wild-type *S. coelicolor* by conjugation and ex-conjugants were selected by growing on MS agar plates containing apramycin. Following introduction of the pJM016, two colony types were observed (Fig. 6), one with wild-type appearance and the other with small colonies that over-produce actinorhodin, reminiscent of the *lsp* mutant BJT1001. PCR testing
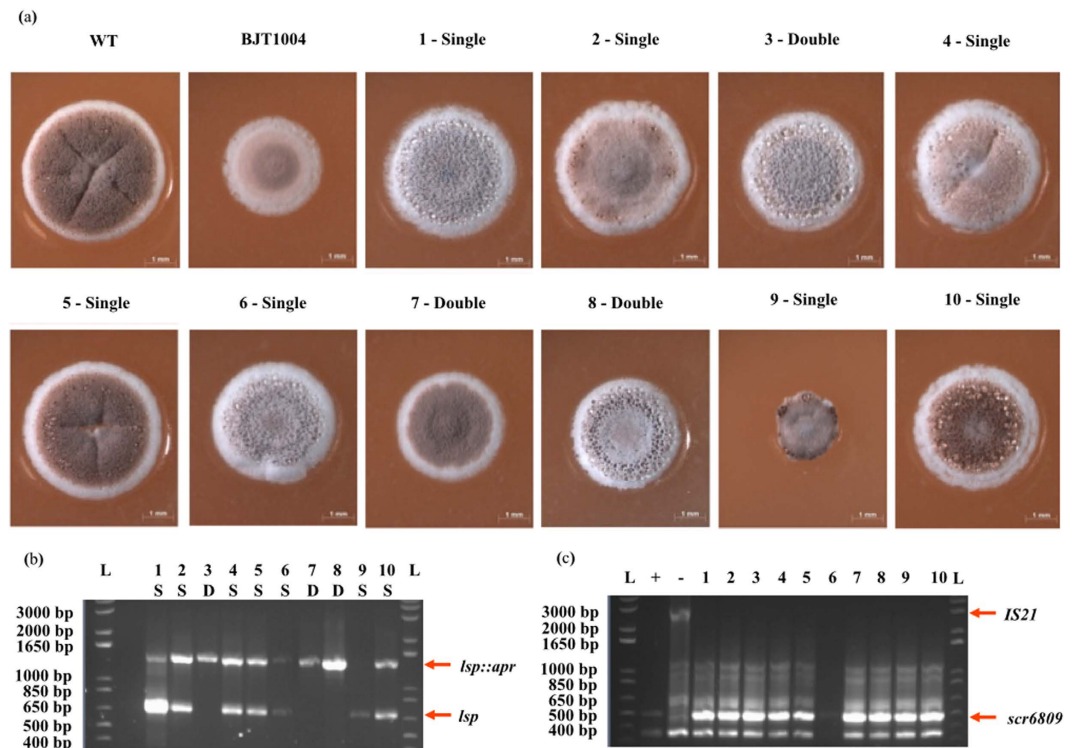
Figure 4. New *lsp* mutants generated using ReDirect do not contain the *IS21* mutation. Colony morphology of mutants *lsp::apr* 1–10 (corresponding to strains JTM008.01–JTM008.10), both single (n = 7, colonies 1–2, 4–7 and 9–10) and double cross-overs (n = 3, colonies: 3 and 7–8) show a range of phenotypes (**a**). PCR of the *lsp* loci (**b**) indicates colonies are either a single (wild-type and/or mutant band) or double (mutant band only) cross-overs (WT = 687 bp, mutant = 1447 bp). PCR of the *scr6809* loci (**c**) indicates that strains 1–10 have intact *scr6809* with no Insertion (WT = 514 bp, IS21 insertion = 2884 bp) using wild-type M145 and BJT1004 as controls (labeled "+" and "−" respectively).
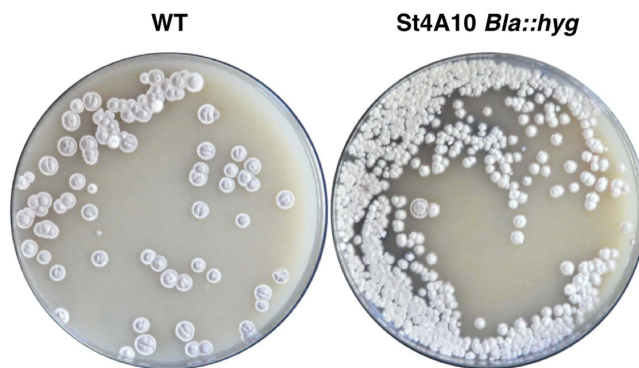

Figure 5. Introduction of wild-type St4A10 causes a pleiotropic phenotype. Conjugation of M145 with St4A10 *bla::hyg* results in non-wildtype phenotypes similar to those observed in the St4A10 *lsp::apr* single cross-overs.

of the genomic DNA of both morphotypes revealed that those with the wild-type colony morphology are indeed wild-type strains with a fully functioning *lsp* gene whereas those with a small colony phenotype that over-produce actinorhodin have disruptions in *lsp* caused by pJM016. PCR amplification and sequencing of the loci in the small colony variants revealed an interesting and unexpected recombination event had occurred. The vector and almost all of the *lsp* gene have been removed such that all that remains is the apramycin resistance cassette (Supplementary Text S4 and S5). These data confirm that *lsp* is not essential in *S. coelicolor* and that loss of Lsp function (and the resulting loss of lipoproteins from the cell membrane) results in a growth and developmental defect and overproduction of the blue antibiotic actinorhodin, as observed previously[5].

Strain				
E. coli	**Genotype/description**	**Plasmid (held)**	**Resistance**	**Source**
TOP10	F– *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX74 recA1 araD139* Δ(ara leu) 7697 *galU galK rpsL* (StrR) *endA1 nupG*	—	—	Invitrogen
BW25113	F-, DE(*araD-araB*)567, *lacZ*4787(del)::*rrnB*-3, LAM-, *rph*-1, DE(*rhaD-rhaB*)568, *hsdR*514	pIJ790	CmR	Datsenko & Wanner[30]
ET12567	dam- dcm- hsdM-	pUZ8002	CmR/TetR	MacNeil *et al.*[31]
Streptomyces	Genotype/description	Plasmid (used)	Resistance	Source
M145	*S. coelicolor* wild type strain, SCP1-, SCP2-	—	—	Hopwood *et al.*[32]
BJT1000	M145 *lsp::apr*	—	AprR	Thompson *et al.*[5]
BJT1001	M145 *lspFLP*	—	—	Thompson *et al.*[5]
BJT1004	BJT1000 + Sco *lsp* cis	—	—	Thompson *et al.*[5]
JTM005	M145 *sco6808::apr*	pJM010	AprR	This work
JTM007	M145 *scr6809::apr*	pJM012	AprR	This work
JTM008	M145 *lsp::apr*	pJM013	AprR	This work
JTM009	M145 St4A10 *bla::hyg*	pJM014	KanR/HygR	This work
JTM012	BJT1004 *sco6808::apr*	pJM010	AprR	This work
JTM015	BJT1004 *lsp::apr*	pJM013	AprR	This work
JTM018	M145 *lsp* suicide vector	pJM016	AmpR/AprR	This work
Plasmids	Genotype/description		Resistance	Source
pIJ773	*aac(3)IV oriT* (contains apramycin (apr) resistance cassette)		AprR	Gust *et al.*[26]
pIJ10700	contains hygromycin resistance cassette, FRT *oriT-hyg* FRT MkII		HygR	Gust *et al.*[26]
pIJ790	*araC-Parab, Y, β, exo, cat, repA1001ts, oriR101*		CmR	Gust *et al.*[26]
pUZ8002	RK2 derivative with a mutation in *oriT*			Kieser *et al.*[25]
pMS82	*ori*, pUC18, *hyg, oriT, RK2, int ΦBT1*		HygR	Gregory *et al.*, 2003
pIJ10257	*oriT, ΦBT1 attB-int, Hygr, ermEp*, pMS81 backbone		HygR	Hong *et al.*, 2005
pGEM-T-Eazy	*bla, lacZα*		AmpR	Promega
St1A2	Supercos-1-cosmid with (39829 bp) fragment containing (*sco6808* and *scr6809*)		KanR/AmpR	Redenbach *et al.*[20]
St4A10	Supercos-1-cosmid with (43147 bp) fragment containing (*sco2074-lsp*)		KanR/AmpR	Redenbach *et al.*[20]
pJM010	St1A2 containing *sco6808::apr oriT* (St1A2Δ*sco6808*)		KanR/AmpR/AprR	This work
pJM012	St1A2 containing *scr6809::apr oriT* (St1A2Δ*scr6809*)		KanR/AmpR/AprR	This work
pJM013	St4A10 containing sco2074::apr oriT (St4A10Δ*lsp*)		KanR/AmpR/AprR	This work
pJM014	St4A10 containing *bla::hyg oriT* (St1A2*bla::hyg*)		KanR/HygR	This work
pJM015	pMS82 containing full length *sco2074* and promoter (300 bp of upstream DNA)		HygR	This work
pJM016	*lsp* suicide vector, pGEM-T-Eazy, 411 bp fragment of the *lsp* gene with a BamHI site. The *aac(3)IV* containing BamHI fragment from a pIJ773 was sub cloned in.		AmpR/AprR	This work
pJM017	pMS82, KpnI/HindIII insert containing, pMC500 MCS and terminators with *scr6809* (*see Text S3*)		HygR	This work
Primer	Sequence		Description	Source
JM0083	GTCTATGGTTGACGGGTGACTGTCATAGATCTGCAGATGATTCCGGGGATCCGTCGACC		*sco6808* forward disruption primer (ReDirect)	This work
JM0084	GTCATCTTCCGAACGGAGATGGAGGGAGATCCGGAATCATGTAGGCTGGAGCTGCTTC		*sco6808* reverse disruption primer (ReDirect)	This work
JM0085	CGGAGGCCGCCTGTCCTAGC		*sco6808* forward test primer	This work
JM0086	AACGCGCACTCGCTGCGGTC		*sco6808* reverse test primer	This work
JM0091	TCCGACATCTGCAGATCTATGACAGTCACCCGTCAACCAATTCCGGGGATCCGTCGACC		*scr6809* forward disruption primer (ReDirect)	This work
JM0092	TGGTACACGGCACCGACTCCGGCTGCCAGAAAGCCATAGTGTAGGCTGGAGCTGCTTC		*scr6809* reverse disruption primer (ReDirect)	This work
JM0093	CAGACGCAGGCCTCGCCATC		*scr6809* forward test primer	This work

Continued

Strain				
E. coli	Genotype/description	Plasmid (held)	Resistance	Source
JM0094	CCCATCGCTACGGCCGCCT	*scr6809* reverse test primer		This work
JM0093	AATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTCAGGCGCCGGGGGCGGTG	*bla (bla::hyg)* gene forward disruption primer (ReDirect) for supercos-1		This work
JM0096	CCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTAAGTTCCCGCCAGCCTCGCA	*bla (bla::hyg)* gene forward disruption primer (ReDirect) for supercos-1		This work
JM0097	AAGCAGCAGATTACGCGCAG	*bla (bla::hyg)* gene forward test primer (ReDirect) for supercos-1		This work
JM0098	GTGCGCGGAACCCCTATTTG	*bla (bla::hyg)* gene forward test primer (ReDirect) for supercos-1		This work
JM0099	TCGTGCTCAGTCAAGGACCTAGGCTGAGGGACTCACGTGATTCCGGGGATCCGTCGACC	*lsp (sco2074)* forward disruption primer (ReDirect)		This work
JM0100	GACAACCAGTCCCTGTGGACAGCCGGACCGGAGGGGTCATGTAGGCTGGAGCTGCTTC	*lsp (sco2074)* reverse disruption primer (ReDirect)		This work
JM0113	GCAACAGTGCCGTTGATCGTGCTATG	pMS82 cloning forward test primer		This work
JM0114	GCCAGTGGTATTTATGTCAACACCGCC	pMS82 cloning reverse test primer		This work
JM0115	GGATCCCTGTTCGCGGTCGCCCTGTTCGCGTACCT	Forward primer amplifies a 411 bp fragment of the *lsp* gene, adding a bamHI site upstream.		This work
JM0116	GATGCCGCCGCACACGATCGCCGAGTCGG	Reverse primer amplifies a 411 bp fragment of the *lsp* gene		This work
JM0150	TCGTGCTCAGTCAAGGACCT	Sco Lsp Test For		Thompson *et al.*[5]
JM0151	GACAACCAGTCCCTGTGGAC	Sco Lsp Test Rev		Thompson *et al.*[5]
JM0154	AAGCTTCGACGAGGCGGACACAGGCAG	*lsp (sco2074)* comp (pMS82) for HindII		This work
JM0155	GGTACCTCAGTCCTTGTGGACGGTCCCGTC	*lsp (sco2074)* comp (pMS82) Rev KpnI		This work

Table 1. Strains, plasmids and primers.

Discussion

The results presented here show that the pleiotropic nature of *S. coelicolor* Δ*lsp* strain BJT1001 resulted from the introduction of cosmid St4A10Δ*lsp* and this was most likely caused by over-expression of the cell division and cell wall biosynthesis genes carried on that cosmid (Supplementary Table S2). We have further shown that the resulting secondary mutations do not make it easier to delete *lsp* suggesting they are not *lsp*-specific suppressors. Genetic manipulation has always been challenging in *Streptomyces* bacteria and the ReDirect PCR targeting method 13 years ago was a significant development but our work is a cautionary tale to others to consider the effects of using large insert cosmid libraries in the genetic manipulation of bacteria. Fortunately, recent advances in CRISPR/Cas9 editing of *Streptomyces* genomes[21] negate the need for a cosmid library and techniques such as this will further accelerate research into the basic biology of *Streptomyces* and other filamentous actinomycetes. This is vital because the secondary metabolites derived from these bacteria are an under utilised reservoir from which new anti-infectives and other drugs can and must be developed. Moreover, our identification of a new small RNA *scr6809* and the demonstration that its deletion results in a range of growth and developmental defects add to the growing appreciation of the significance of small RNAs in streptomycetes[22–24].

Materials and Methods

Bacterial strains and culture conditions. All primers, plasmids and strains used are listed in Table 1. Strains were routinely grown as previously described[5] following the recipes of Kieser *et al.*[25]. *E. coli* was grown in LB or LB –NaCl for Hygromycin selection and *S. coelicolor* M145 and its derivatives were grown on Soya Flour Mannitol (SFM) medium to study growth and development or LB culture for genomic isolations.
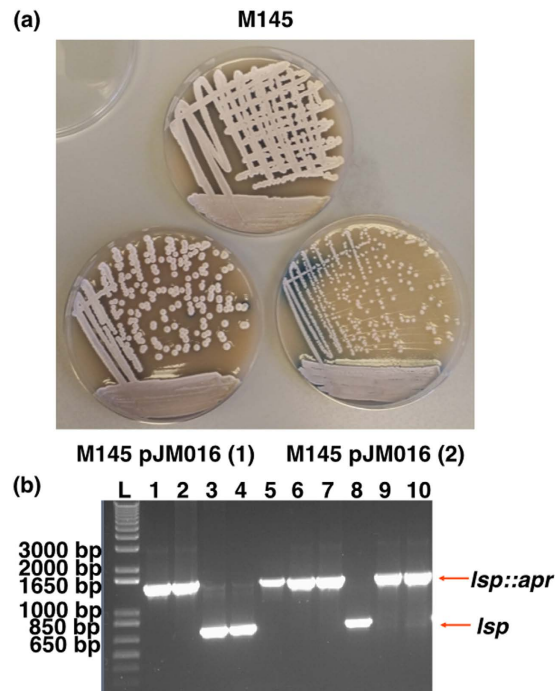
Figure 6. Targeted disruption of *lsp* using a suicide vector results in a small colony phenotype that over-produces actinorhodin. To test how much of the BJT1001 phenotype is due to loss of *lsp* we disrupted the *lsp* gene using a suicide vector which does not affect or duplicate any other target genes. Plate images (**a**) show two distinct phenotypes following insertion of the suicide vector (pJM016) into M145, either a wildtype appearance (M145 pJM016 (1), n = 3 corresponding to JTM018.03-04 and 08) or a small colony phenotype over producing actinorhodin (M145 pJM016 (2), n = 7, corresponding to strains JTM018.01-2, 05-07 and 09-10) similar to our original observation of the *lsp* phenotype[5], alongside *lsp* loci PCR results (**b**). All strains with intact *lsp* show a wild-type phenotype while those with disrupted *lsp* have the reported *lsp* phenotype.

Gene deletions and complementation.   Gene deletions were carried out following the ReDirect method of PCR-targeting[26] as previously described Hutchings *et al.*[27]. Disruption of *lsp* (*sco2074::apr*) on cosmid St4A10 (pJM013, St4A10Δ*lsp*) using the pIJ773 apramycin disruption-cassette and *sco6808* (*sco6808::*a*pr*) and *scr6809* (*scr6809::apr*) on cosmid St1A2 (pJM010-St1A2Δ*sco6808* and pJM012-St1A2Δ*sco6808* respectively) using primers JM0101-2, JM0083-84 and JM0091-2 respectively were confirmed by PCR using primers JM0150-1, JM0085-6 and JM0093-4 respectively. Introduction of the wild-type cosmid St4A10 was facilitated by introducing an *oriT* by disruption and replacement of the Supercos-1 backbone *bla* resistance gene (pJM014–St4A10*bla::hyg)* using primers JM0095-6 and the hygromycin disruption cassette from pIJ10701, confirmed using primers JM0099-100. The *lsp* suicide vector pJM016 was produced by introducing a 411 bp fragment of the *lsp* gene with an N-terminal *Bam*HI site, amplified with primers JM0117-8 and cloned into pGEM T-Eazy. The *Bam*HI site was then used to subclone the *Bam*HI fragment from a pIJ773 digest, containing an *apr* disruption cassette. An overexpression construct, pJM017 was synthesised by Genscript to include the pMC500 MCS and terminators[28] with *scr6809* (sequence included in Supplementary Text S5). All constructs were subsequently conjugated into *S. coelicolor* following the method described by Gust *et al.*[26].

Genomic DNA isolation.   Genomic DNA was isolated from M145 and BJT1004 following the Pospiech and Neumann (1995) salting out method as described by Keiser *et al.* (2000). Mycelium from a 30 ml culture was resuspended in 5 ml SET buffer containing 1 mg/ml lysozyme and incubated at 37 °C 30–60 min. To this lysate, 140 μl of proteinase K solution (20 mg ml^{-1}) was added, mixed, then 600 μl of 10% SDS added, mixed and incubated at 55 °C for 2 h, with occasional mixing throughout. After this incubation 2 ml of 5 M NaCl was added, mixed and left to cool to 37 °C before adding 5 ml chloroform, mixed at 20 °C for 30 min. Samples were centrifuged at 4500 × g for 15 min at 20 °C. The supernatant was removed to a fresh tube and DNA precipitated by adding 0.6 volumes of 100% isopropanol. Tubes were mixed by inversion and after at least 3 min DNA spooled out using a sterile Pasteur pipette. The DNA was rinced in 70% ethanol, air dried and dissolved in 1–2 ml TE buffer (10 mM Tris-HCl pH 7.8, 1 mM EDTA) at 55 °C.

Genome resequencing and secondary mutation identification.   The isolated DNA from our wild-type *S. coelicolor* M145 parent strain and BJT1004 were sent to both GATC Biotech and The Genome Analysis Centre (TGAC) for 35 bp paired end HiSeq Illumina sequencing. Assembly mapping and SNP identification was carried out with MIRA (Chevreux *et al.*, 2004) using the reference genome NC_003888 (Bentley *et al.*[29]) as a scaffold for mapping each of the resequenced genomes. Putative SNPs were detected in each sample

independently reporting the SNP position, the nucleotide change, the number of reads that sequence the region, those containing wild-type or mutated nucleotides and a percentage change. Each set of results was then compared by eye to determine the likely hood that a SNP was real by number of reads containing the mutation and its presence in each sample. Larger mutations (rearrangements) were identified in the same fashion.

Microscopy. Brightfield images were acquired using a Zeiss M2 Bio Quad SV11 stereomicroscope. Samples were illuminated from above using a halogen lamp images captured with an AxioCam HRc CCD camera. The AxioVision software (Carl Zeiss, Welwyn Garden City, UK) was used for image capture and processing.

References

1. Hutchings, M. I., Palmer, T., Harrington, D. J. & Sutcliffe, I. C. Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em. *Trends Microbiol.* **17,** 13–21 (2009).
2. Buddelmeijer, N. The molecular mechanism of bacterial lipoprotein modification–How, when and why? *FEMS Microbiol. Rev.* **39,** 246–261 (2015).
3. Sutcliffe, I. C., Harrington, D. J. & Hutchings, M. I. A phylum level analysis reveals lipoprotein biosynthesis to be a fundamental property of bacteria. *Protein Cell* **3,** 163–170 (2012).
4. Gralnick, J. A., Vali, H., Lies, D. P. & Newman, D. K. Extracellular respiration of dimethyl sulfoxide by *Shewanella oneidensis* strain MR-1. *Proc. Natl. Acad. Sci. USA* **103,** 4669–4674 (2006).
5. Thompson, B. J. *et al.* Investigating lipoprotein biogenesis and function in the model Gram-positive bacterium *Streptomyces coelicolor. Mol. Microbiol.* **77,** 943–957 (2010).
6. Shruthi, H., Babu, M. M. & Sankaran, K. TAT-pathway-dependent lipoproteins as a niche-based adaptation in prokaryotes. *J. Mol. Evol.* **70,** 359–370 (2010).
7. Rahman, O., Cummings, S. P., Harrington, D. J. & Sutcliffe, I. C. Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. *World J. Microbiol. Biotechnol.* **24,** 2377–2382 (2008).
8. Okuda, S. & Tokuda, H. Lipoprotein Sorting in Bacteria. *Annu. Rev. Microbiol.* **65,** 239–259 (2011).
9. Narita, S. I., Matsuyama, S. I. & Tokuda, H. Lipoprotein trafficking in *Escherichia coli. Arch. Microbiol.* **182,** 1–6 (2004).
10. Jan, G., Fontenelle, C., Le Hénaff, M. & Wróblewski, H. Acylation and immunological properties of *Mycoplasma gallisepticum* membrane proteins. *Res. Microbiol.* **146,** 739–750 (1995).
11. Asanuma, M. *et al.* Structural evidence of α-aminoacylated lipoproteins of *Staphylococcus aureus. FEBS J.* **278,** 716–728 (2011).
12. Nakayama, H., Kurokawa, K. & Lee, B. L. Lipoproteins in bacteria: Structures and biosynthetic pathways. *FEBS J.* **279,** 4247–4268 (2012).
13. Kurokawa, K. *et al.* The Triacylated ATP Binding Cluster Transporter Substrate-binding Lipoprotein of *Staphylococcus aureus* Functions as a Native Ligand for Toll-like Receptor 2. *J. Biol. Chem.* **284,** 8406–8411 (2009).
14. Serebryakova, M. V. *et al.* The acylation state of surface lipoproteins of mollicute *Acholeplasma laidlawii. J. Biol. Chem.* **286,** 22769–22776 (2011).
15. Widdick, D. a. *et al.* Dissecting the complete lipoprotein biogenesis pathway in *Streptomyces scabies. Mol. Microbiol.* **80,** 1395–1412 (2011).
16. Tschumi, A. *et al.* Functional analyses of mycobacterial lipoprotein diacylglyceryl transferase and comparative secretome analysis of a mycobacterial *lgt* mutant. *J. Bacteriol.* **194,** 3938–3949 (2012).
17. Xiao, Y. & Wall, D. Genetic redundancy, proximity, and functionality of lspA, the target of antibiotic TA, in the *Myxococcus xanthus* producer strain. *J. Bacteriol.* **196,** 1174–1183 (2014).
18. Yang, Y. H. *et al.* Finding new pathway-specific regulators by clustering method using threshold standard deviation based on DNA chip data of *Streptomyces coelicolor. Appl. Microbiol. Biotechnol.* **80,** 709–717 (2008).
19. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. 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 (2003).
20. Redenbach, M. *et al.* A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol. Microbiol.* **21,** 77–96 (1996).
21. Cobb, R. E., Wang, Y. & Zhao, H. High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **4,** 723–728 (2014).
22. Vockenhuber, M. P. *et al.* Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor. RNA Biol.* **8,** 468–477 (2011).
23. Moody, M. J., Young, R. A., Jones, S. E. & Elliot, M. A. Comparative analysis of non-coding RNAs in the antibiotic-producing *Streptomyces* bacteria. *BMC Genomics* **14,** 558 (2013).
24. Romero, D. A. *et al.* 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.* **94,** 963–987 (2014).
25. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. Practical *Streptomyces* Genetics. *John Innes Cent. Ltd.* 529, doi: 10.4016/28481.01 (2000).
26. Gust, B., Kieser, T. & Chater, K. PCR targeting system in *Streptomyces coelicolor* A3(2). *John Innes Cent.* **3,** 1–39 (2002).
27. Hutchings, M. I., Hong, H. J., Leibovitz, E., Sutcliffe, I. C. & Buttner, M. J. The sigma(E) cell envelope stress response of *Streptomyces coelicolor* is influenced by a novel lipoprotein, CseA. *J. Bacteriol.* **188,** 7222–7229 (2006).
28. Gao, C., Mulder, D., Yin, C. & Elliot, M. a. Crp Is a Global Regulator of Antibiotic Production in *Streptomyces. MBio* **3,** 1–12 (2012).
29. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417,** 141–147 (2002).
30. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97,** 6640–6645 (2000).
31. MacNeil, D. J. *et al.* Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111,** 61–68 (1992).
32. Hopwood, D. A. *et al. Genetic manipulation of Streptomyces: a laboratory manual.* (1985).

Acknowledgements

Author Contributions

J.T.M. designed and carried out the experiments, J.T.M., I.C.S., T.P. and M.I.H. designed experiments, J.T.M., D.A.W., I.C.S., T.P. and M.I.H. analysed data and wrote the manuscript, D.A.W. prepared DNA for sequencing, G.C. analysed sequencing data.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Munnoch, J. T. *et al.* Cosmid based mutagenesis causes genetic instability in *Streptomyces coelicolor*, as shown by targeting of the lipoprotein signal peptidase gene. *Sci. Rep.* **6**, 29495; doi: 10.1038/srep29495 (2016).