

Nucleoid-Associated Proteins of *Streptomyces coelicolor*: Discovery and Functions

Elizabeth Helen Bradshaw

2012

Thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

© This copy of the Thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that neither quotation from the thesis, nor any information derived there from may be published without the author's prior, written consent.

Acknowledgements

I would like to first acknowledge the John Innes Centre for being such a friendly and stimulating environment to work in and for giving me the chance to rotate through three different wonderful labs and attend the excellent Friday seminars.

The greatest thanks go to Michael McArthur for giving me the opportunity to work in his lab on such an exciting area and for his advice and encouragement throughout my project. Also thanks to Mervyn Bibb, Mark Buttner and Peter Shaw who formed my supervisory committee and gave me such valuable feedback and guidance. I would like to thank my rotation supervisors Anne Osbourn, Jo Dicks, Tony Maxwell and George Lomonossoff for introducing me to the research that goes on at the rest of the institute and Nick Brewin for organising so many events for the rotation students.

An enormous thankyou goes to Gerhard Saalbach, Govind Chandra, Kim Findlay and Francis Mulholland for all their technical help and patience. Thanks to everyone in Molecular Microbiology and Biological Chemistry for making me feel so welcome in the lab and for helping me learn all the techniques I needed during my project - I could never have managed this without all your day-to-day assistance and wisdom. I deeply appreciate how much our departmental administrators and the staff in stores and the media kitchen do to make it possible for me to focus my energies on science.

I would like to thank the friends I have made at the John Innes Centre and the University of East Anglia, as well as the collection of reprobates in the Grad bar, for moral support and for keeping me sane by making sure there were at least a few hours each week that I stopped thinking about my PhD. Finally I would like to thank Sam and my family for their constant support and for keeping me motivated on the days when I felt like all my experiments were going wrong.

Abstract

The regulation of gene expression from biosynthetic gene clusters is a key concern in natural product discovery as these clusters are often transcriptionally silent (“cryptic”) under normal laboratory conditions, making the initial characterisation and heterologous expression of the products they encode problematic. The role of the architectural nucleoid-associated proteins (NAPs) in regulation of the expression of these products has been neglected within the Actinomycetes. NAPs are small, highly abundant proteins which govern both gene expression and nucleoid structure on a genome-wide scale.

A method for surveying the proteome of the *S. coelicolor* nucleoid was developed which generated a list of 25 proteins with a high probability of being NAPs. This list included known NAPs such as HupA, HupS, Lsr2 and siHF and the known global regulators CRP and BldD, as well as a number of interesting novel NAP candidates from a variety of protein classes suitable for further investigation. One of these proteins, SCO5592, was investigated as a candidate global RNA-binding regulator. It was found to comprise a KH domain with an N-terminal extension and formed oligomers of 10 - 12 subunits in solution.

The mutant phenotypes of both *S. coelicolor* paralogs of HU (HupA and HupS) were examined and showed opposite effects on growth, spore formation and actinorhodin production. The mutant phenotypes of both *S. coelicolor* paralogs of the H-NS-like proteins Lsr2 (SCO3375 and SCO4076) were milder and showed a greater degree of overlap than the HU mutants, however both showed a moderate abnormality in spore morphology, suggesting that they have a role in spore nucleoid segregation.

Phylogenetic analysis of the HU and Lsr2 proteins revealed that the lysine-rich tail domain of HupS was acquired after the HupA and HupS core domains had begun to diverge and that lysine-rich sequences have evolved multiple times.

General abbreviations

3D	Three-dimensional
A	Adenine
aa	Amino acids
AFM	Atomic force microscopy
Apra	Apramycin
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
bp	Base pairs
C	Cytosine
cAMP	Cyclic adenosine monophosphate
Carb	Carbenicillin
cDNA	Complementary DNA
CFUs	Colony-forming units
ChIP	Chromatin immunoprecipitation
Cm	Chloramphenicol
C-terminal	Carboxy-terminal
ddH ₂ O	Double-distilled water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA	Difco Nutrient Agar
DNase	Deoxyribonuclease
dNTP	Deoxynucleotide triphosphates
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EF	Elongation factor
EMSA	Electrophoretic mobility shift assay
FISH	Fluorescent <i>in situ</i> hybridisation
FPLC	Fast protein liquid chromatography
G	Guanine
gDNA	Genomic DNA

GE	Gel electrophoresis
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HGT	Horizontal gene transfer
h	Hours
Hyg	Hygromycin
HPI	Hours post-inoculation
IPTG	Isopropyl β -D-1-thiogalactopyranoside
Kan	Kanamycin
kb	Kilobase pairs
kDa	Kilodalton
<i>lacZ</i>	Gene encoding β -galactosidase
LC-MS/MS	Liquid chromatography with tandem mass spectrometry
m/z	Mass to charge ratio
MALDI-ToF	Matrix-assisted laser desorption ionisation time-of-flight
MCS	Multiple cloning site
min	Minute
mRNA	Messenger RNA
ms	Milliseconds
MS	Mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut-off
Nal	Nalidixic acid
NAP	Nucleoid-associated protein
NMR	Nuclear magnetic resonance
N-terminal	Amino-terminal
nt	Nucleotides
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
psi	Pounds per square inch
qPCR	Quantitative PCR
RNA	Ribonucleic acid
RNAP	RNA polymerase

RNS	Reactive nitrogen species
ROS	Reactive oxygen species
RBS	Ribosome binding site
rRNA	Ribosomal RNA
RNase	Ribonuclease
RT	Room temperature
s	Second
SDS	Sodium dodecyl sulphate
<i>sp.</i>	Species
sRNA	Small non-coding RNA
ssDNA	Single-stranded DNA
T	Thymine
TFA	Trifluoroacetic acid
UTR	Untranslated region
UV	Ultraviolet
V	Volts
WT	Wild-type

Contents

Chapter 1 : Introduction

1.1	Levels of organisation in the nucleoid.....	19
1.1.1	Compaction.....	20
1.1.2	Gross nucleoid morphology and positioning.....	21
1.1.3	Supercoiling, topoisomerases and topological domains.....	23
1.1.4	Macrodomains and genome layout.....	26
1.1.5	Contribution of transcription and transcription factories.....	28
1.1.6	Local structural considerations.....	29
1.1.7	<i>Streptomyces coelicolor</i> chromosome layout.....	31
1.2	Nucleoid-associated proteins (NAPs)	33
1.2.1	DNA-bending proteins: HU, IHF, Mdp1.....	37
1.2.2	DNA-bridging proteins: H-NS, StpA and Lsr2.....	41
1.2.3	Ligand-binding NAPs: Lrp and CRP.....	43
1.2.4	DNA-wrapping proteins: Fis.....	45
1.2.5	DNA-protecting proteins: Dps and SASPs.....	46
1.2.6	Less well-characterised NAPs.....	47
1.3	Global RNA chaperones.....	49
1.3.1	Small non-coding RNAs (sRNAs)	50
1.3.2	Global RNA chaperones.....	51
1.3.3	sRNA-mediated regulation in Actinobacteria and <i>S. coelicolor</i>	53
1.4	Post-translational modification of chromatin.....	54
1.4.1	Acetylation.....	56
1.4.2	Methylation.....	57
1.4.3	Phosphorylation.....	58
1.4.4	Chemical epigenetics.....	58
1.5	<i>Streptomyces coelicolor</i>	59
1.5.1	<i>Streptomyces</i> life cycle.....	60
1.5.2	Natural product diversity of the Actinomycetes.....	62

Chapter 2 : Materials and Methods

2.1 Bacterial strains and plasmids.....	64
2.2 Culture media and antibiotics.....	67
2.2.1 Antibiotics.....	67
2.2.2 Agar media.....	67
2.2.3 Liquid media.....	68
2.3 Growth and storage of bacteria.....	70
2.3.1 Growth and storage of <i>E. coli</i>	70
2.3.2 Growth and storage of <i>Salmonella</i>	70
2.3.3 Growth and storage of <i>S. coelicolor</i>	70
2.3.4 Synchronous germination of <i>S. coelicolor</i> spores.....	71
2.3.5 Preparation of <i>S. coelicolor</i> protoplasts.....	71
2.3.6 <i>S. coelicolor</i> growth curves.....	71
2.4 Buffers and solutions.....	72
2.5 General molecular biology methods.....	74
2.5.1 Isolation of plasmid DNA.....	74
2.5.2 Isolated of cosmid DNA.....	74
2.5.3 Isolation of HMW genomic DNA.....	74
2.5.4 Agarose gel electrophoresis.....	75
2.5.5 Extraction of DNA fragments from gel slices.....	75
2.5.6 Purification of DNA following PCR or restriction digest.....	75
2.5.7 Measurement of DNA concentration.....	76
2.5.8 Digestion of DNA using restriction endonucleases.....	76
2.5.9 Ligation.....	76
2.5.10 Transformation of chemically competent <i>E. coli</i>	76
2.5.11 Preparation and transformation of electro-competent <i>E. coli</i>	76
2.6 PCR.....	77
2.6.1 General PCR.....	77
2.6.2 Colony PCR in <i>E. coli</i>	78
2.6.3 Colony PCR in <i>Streptomyces</i>	78
2.6.4 REDIRECT PCR.....	78
2.6.5 Sanger sequencing using Big Dye v3.1.....	79
2.7 Quantitative PCR (qPCR)	80
2.7.1 Primer design.....	80

2.7.2 qPCR assay optimisation.....	80
2.7.3 qPCR assay conditions.....	81
2.7.4 Ct value calculation.....	82
2.7.5 Evaluation of data quality.....	83
2.7.6 Copy number calculation.....	84
2.8 PCR-targeted mutagenesis (REDIRECT™)	85
2.8.1 Cosmids, template and primers used.....	85
2.8.2 Introduction of <i>S. coelicolor</i> cosmids into <i>E. coli</i> BW25113/pIJ790.....	86
2.8.3 Replacement of the gene on the cosmid in <i>E. coli</i>	86
2.8.4 Transfer of the mutant cosmids into <i>S. coelicolor</i>	86
2.8.5 Verification of mutant <i>Streptomyces</i> strains.....	87
2.8.6 Complementation of the mutant phenotypes.....	88
2.9 General protein methods.....	89
2.9.1 SDS-PAGE	89
2.9.2 Dialysis.....	89
2.9.3 Methanol precipitation	89
2.9.4 2D gel electrophoresis.....	90
2.9.5 Purification of His-tagged recombinant proteins by nickel affinity chromatography.....	90
2.9.6 His-tag removal using thrombin.....	91
2.10 Mass spectrometry.....	91
2.10.1 Trypsin digestion.....	91
2.10.2 Matrix-assisted laser-desorption ionisation time of flight (MALDI-ToF).....	91
2.10.3 ESI-Orbitrap MS/MS.....	92
2.10.4 Data analysis.....	92
2.11 Next-generation sequencing.....	93
2.12 Microscopy.....	93
2.12.1 Light microscopy.....	93
2.12.2 Cryo-scanning electron microscopy.....	93
2.13 Nucleoid isolation.....	94
2.13.1 10—30% linear sucrose gradient preparation.....	94
2.13.2 Drlica method of nucleoid isolation.....	94
2.13.3 Sarfert method of nucleoid isolation.....	95
2.14 Oligonucleotides used in this study.....	96

Chapter 3 : Proteomic survey of the *S. coelicolor* nucleoid

3.1	Introduction.....	98
3.2	System and approach.....	99
3.3	Gradient controls and Drlica method for preparing <i>E. coli</i> nucleoids.....	100
3.4	Isolation of nucleoids from <i>S. coelicolor</i>	102
3.5	Determination of <i>S. coelicolor</i> nucleoid protein content by 1D and 2D electrophoresis.....	104
3.6	Experimental design and system for LC-MS/MS experiment.....	106
3.7	Nucleoid isolation for LC-MS/MS.....	107
3.8	Protein identification and quantification.....	108
3.9	Abundant proteins present in the <i>S. coelicolor</i> nucleoid.....	108
3.10	emPAI threshold for classifying annotated transcriptional regulators as NAPs.....	111
3.11	Prediction of DNA-binding ability.....	113
3.12	Candidate NAPs/global regulators.....	113
3.13	Detection of candidate NAPs in Actinomycetes and other lineages.....	118
3.14	Post-translational modifications.....	121
3.15	Discussion.....	122

Chapter 4 : HU mutants of *S. coelicolor*

4.1	Introduction.....	126
4.2	Strains.....	127
4.3	Phenotypes screened.....	129
4.4	Summary of phenotypes observed.....	130
4.5	Phenotype of $\Delta hupA$ on agar media.....	131
4.6	Phenotype of $\Delta hupA$ in liquid media.....	134
4.7	Germination deficiency of $\Delta hupA$	137
4.8	Phenotype of $\Delta hupS$ on agar media.....	139
4.9	Phenotype of $\Delta hupS$ in liquid media.....	142
4.10	Effects of R5+ additives on $\Delta hupS$ phenotype.....	144
4.11	Spore morphology of $\Delta hupS$	145
4.12	$\Delta hupA$ and $\Delta hupS$ mature colony morphology.....	147
4.13	Phylogenetic analysis of HU distribution among the Actinomycetes.....	148
4.14	Discussion.....	153

Chapter 5 : Lsr2 evolution and Lsr2 mutants of *S. coelicolor*

5.1	Introduction.....	157
5.2	Strains.....	158
5.3	Phenotypes screened.....	159
5.4	Summary of phenotypes observed.....	159
5.5	Phenotype of $\Delta 3375$ on agar media.....	160
5.6	Phenotype of the SCO3375 over-expression strain B1005 on agar media.....	162
5.7	Phenotype of $\Delta 4076$ on agar media.....	163
5.8	Phenotype of the SCO4076 over-expression strain B1007 on agar media.....	165
5.9	Liquid growth phenotypes of the $\Delta 3375$ and SCO3375 over-expression strains.....	167
5.10	Liquid growth phenotypes of the $\Delta 4076$ and SCO4076 over-expression strains.....	168
5.11	Spore heat tolerance of $\Delta 4076$	169
5.12	Spore morphology of $\Delta 3375$ and $\Delta 4076$	170
5.13	Phylogenetic analysis of Lsr2 distribution among the Actinomycetes.....	172
5.14	Discussion.....	176

Chapter 6 : SCO5592, a putative sRNA-binding protein

6.1	Introduction.....	179
6.2	Candidates arising from chapter 3.....	181
6.3	Microsynteny of the <i>hfq</i> superoperon.....	182
6.4	Screening of candidates for Hfq-like activity by attempting to complement an <i>S. enterica</i> Δhfq mutant with <i>S. coelicolor</i> genes.....	183
6.5	Microsynteny of <i>sco5592</i>	185
6.6	Sequence motif analysis of SCO5592.....	186
6.7	$\Delta 5592$ mutant phenotype.....	188
6.8	Purification and tag removal of recombinant His-5592.....	193
6.9	Oligomerisation state of His-5592.....	195
6.10	Discussion.....	198

Chapter 7 : DNase-seq as a tool to probe the structure of the *S. coelicolor* nucleoid

7.1	Introduction.....	200
7.2	System and approach.....	202
7.3	qPCR experimental design.....	204
7.4	Preparation of DNase I-digested mycelium.....	206
7.5	qPCR validation of DNase I-digested samples.....	208
7.6	<i>Sma</i> I digestion to reduce the average fragment size.....	209
7.7	Library construction and sequencing.....	211
7.8	DNase-seq read quality control and alignment.....	211
7.9	Analysis of DNase-seq cutting pattern.....	213
7.10	<i>Hae</i> III assay design.....	217
7.11	Preparation of <i>Hae</i> III-digested mycelium.....	218
7.12	Discussion.....	220

Chapter 8 : Overall discussion

8.1	Summary of findings.....	223
8.2	Relevance of this work to natural product discovery.....	225
8.3	Future work.....	225
8.4	Concluding remarks.....	227

References.....	228
------------------------	------------

List of tables and figures

Figure 1.1 The gross structure of the bacterial nucleoid.

Figure 1.2 Supercoiling may exist in an unconstrained form in plectonemic loops or in a constrained form in toroidal structures.

Figure 1.3 Macrodomain structure of *E. coli* and correlation of macrodomain layout with long-range patterns in gene expression level, gene essentiality and gene evolutionary retention.

Figure 1.4 Transcriptional foci in *E. coli* wild-type cells and HU mutant cells.

Figure 1.5 Control of gene expression by NAP-mediated modification of promoter architecture.

Figure 1.6 Chromosome layout of *S. coelicolor*.

Figure 1.7 Electron micrographs of *S. coelicolor* nucleoids.

Figure 1.8 Contrasting properties of NAPs and local transcription factors.

Figure 1.9 General classes of NAP.

Figure 1.10 Crystal structure of an IHF heterodimer in complex with DNA.

Figure 1.11 Stabilisation of DNA in a plectonemic form by extended oligomeric of H-NS.

Figure 1.12 Structures of three *E. coli* sRNAs from the Rfam database.

Figure 1.13 Association of chromatin modifications with changes in higher-order compaction and the multiplicity of modifications found on the eukaryotic non-histone chromatin protein HMGA1a.

Figure 1.14 Developmental cycle of *Streptomyces* and the phenotypic consequences of *bld* and *whi* mutations.

Figure 2.1 Positioning of qPCR quantitation threshold.

Figure 2.2 Calculation of qPCR reaction efficiency, repeatability and R^2 from gDNA standards.

Figure 2.3 Copy number calculation from logarithmic gDNA standard curves.

Figure 2.4 Sample primer locations for PCR confirmation of mutant construction.

Figure 2.5 Sample PCR results confirming PCR-targeted replacement of *sco4076* with the apramycin cassette.

Table 2.1 *E. coli* strains used in this study.

Table 2.2 *S. coelicolor* strains used or constructed in this study.

Table 2.3 Plasmids used or constructed in this study.

Table 2.4 Antibiotics used in this study.

Table 2.5 Agar media used in this study.

Table 2.6 Liquid media used in this study.

Table 2.7 Buffers and solutions used in this study.

Table 2.8 Cosmids used for PCR-targeted deletion in this study.

Table 2.9 Genomic co-ordinates of fragments used to complement $\Delta hupA$, $\Delta 3375$ and $\Delta 4076$.

Table 2.10 Oligonucleotides used in this study.

Figure 3.1 Distribution on a 10 - 30% linear sucrose gradient of naked DNA and soluble protein; distribution on a similar sucrose gradient of *E. coli* nucleoids isolated using the Drlica method; protein content of *E. coli* nucleoids as analysed by 1D SDS-PAGE.

Figure 3.2 Application of the Drlica method and the Sarfert method of nucleoid isolation to *S. coelicolor* mycelium.

Figure 3.3 Representative 2D gel of proteins in *S. coelicolor* nucleoid fractions.

Figure 3.4 Sucrose gradient profiles and 1D SDS-PAGE for nucleoids harvested at each time point.

Figure 3.6 Classes of protein represented in the nucleoid fractions at each time point, by summed emPAI.

Figure 3.7 emPAI scores for known *S. coelicolor* NAPs at each time point.

Figure 3.8 Protein abundances were not found to follow a normal distribution.

Figure 3.9 Annotated DNA-binding proteins ranked by abundance at each time point.

Figure 3.10 The 25 most promising candidate NAPs/global regulators found in *S. coelicolor*.

Figure 3.11 The uncharacterised protein SCO4199.

Figure 3.12 Phylogenetic distributions of 42 putative NAPs.

Figure 3.13 Coverage of siHF (SCO1480) seen in an LC-MS/MS experiments on a thin (0.5-1 kDa) gel slice cut from an SDS-PAGE gel of *S. coelicolor* nucleoids.

Figure 4.1 Growth of *E. coli* BL21(DE3) pLysS harbouring pET15b-sco5592 or pET15b-hupS, with and without IPTG induction.

Figure 4.2 Early and increased production of actinorhodin by $\Delta hupA$ grown on SFM plates.

Figure 4.3 Mild developmental delay and actinorhodin reduction of $\Delta hupA$ on R5+ plates.

Figure 4.4 Over-production of actinorhodin and greater extent of sporulation of $\Delta hupA$ on DNA plates.

Figure 4.5 Liquid growth curve of M145 and $\Delta hupA$ in TSB:YEME34%.

Figure 4.6 Production of undecylprodigiosin and actinorhodin in TSB:YEME34% of M145 and $\Delta hupA$ -*B*.

Figure 4.7 Morphology of M145 and $\Delta hupA$ mycelial pellets during early vegetative and late vegetative growth phases.

Figure 4.8 $\Delta hupA$ spore preps were found to have a low number of CFUs/ml compared to M145 and other mutant genotypes.

Figure 4.9 A loss-of-refractility assay showed that $\Delta hupA$ spores germinated with similar timing to M145 but with a smaller magnitude of change in optical density.

Figure 4.10 $\Delta hupS$ took longer than M145 to produce spore dark spore pigment, but finally produced darker single colonies than M145.

Figure 4.11 Less extensive aerial mycelium formation of $\Delta hupS$ on R5+ plates.

Figure 4.12 Reduced actinorhodin secretion of $\Delta hupS$ on DNA plates.

Figure 4.13 Liquid growth curve of M145 and $\Delta hupS$ in TSB:YEME34%.

Figure 4.14 Production of undecylprodigiosin and actinorhodin of M145 and $\Delta hupS$ in TSB:YEME34%.

Figure 4.15 Morphology of M145 and $\Delta hupS$ mycelial pellets during early vegetative and late vegetative growth phases.

Figure 4.16 Actinorhodin secretion of $\Delta hupS$ is more sensitive than M145 to the presence of the R5+ additives proline and NaOH.

Figure 4.17 Scanning electron micrograph of M145 and $\Delta hupS$ grown on SFM for 7 days.

Figure 4.18 Colony morphology of M145, $\Delta hupA$ and $\Delta hupS$ following two weeks of incubation.

Figure 4.19 Colony diameter of M145, $\Delta hupA$ and $\Delta hupS$ grown on SFM plates for 8 and 11 days.

Figure 4.20 Unrooted neighbour-joining tree of HU amino acid sequences from Actinomycetes and representatives from bacteriophages, Gram negative bacteria and Firmicutes.

Figure 4.21 Representative HU amino acid sequences showing variability in lysine-rich domain length.

Figure 5.1 Delayed aerial mycelium formation of $\Delta 3375$ on SFM plates.

Figure 5.2 Delayed aerial mycelium and actinorhodin secretion of $\Delta 3375$ on R5+ plates.

Figure 5.3 Delayed aerial mycelium and actinorhodin secretion of B1005 on R5+ plates.

Figure 5.4 Delayed aerial mycelium and reduced actinorhodin secretion of $\Delta 4076$ on R5+.

Figure 5.5 Over-production of actinorhodin from densely-plated regions of $\Delta 4076$ on DNA.

Figure 5.6 Delayed aerial mycelium and actinorhodin secretion of B1007 on R5+ plates.

Figure 5.7 Over-production of actinorhodin from less densely-plated regions of B1007 on DNA plates.

Figure 5.8 Growth, undecylprodigiosin production and actinorhodin production of $\Delta 3375$ and B1005 as compared to M145.

Figure 5.9 Growth, undecylprodigiosin production and actinorhodin production of $\Delta 4076$ and B1007 as compared to M145.

Figure 5.10 Heat sensitivity of spores of M145, $\Delta hupA$, $\Delta hupS$ and $\Delta 4076$.

Figure 5.11 Scanning electron micrograph of $\Delta 4076$ grown on SFM for 7 days.

Figure 5.12 Scanning electron micrograph of $\Delta 3375$ grown on SFM for 7 days.

Figure 5.13 Unrooted neighbour-joining tree of Lsr2 amino acid sequences from Actinomycetes and bacteriophages.

Figure 5.14 Lsr2 sequences containing lysine-/alanine-/proline-rich repeats.

Figure 6.1 NMR structure of the tellurium-resistance protein TerB and crystal structures the eukaryotic RNA-binding proteins Lsm and Ro60.

Figure 6.2 The genomic contexts of *hfq* genes in different bacterial species.

Figure 6.3 Regions surrounding *miaA* and *hflX* in *S. enterica* and *S. coelicolor*.

Figure 6.4 Growth of *Salmonella* strains in liquid culture over 16 hours.

Figure 6.5 The conserved operon containing the translation-related genes *rpsP*, *rimM*, *trmD* and *rpsL* is interrupted by *sco5592* in *Streptomyces* species.

Figure 6.6 Alignment of SCO5592 with KH domains from other proteins.

Figure 6.7 Graphical representation of the secondary structure prediction for SCO5592 by SCRATCH Protein Predictor.

Figure 6.8 Alignment of amino acid sequences of SCO5592 (*S. coelicolor*), CsrA (*E. coli*) and RsmA (*E. carotovora*).

Figure 6.9 Growth in liquid TSB:YEME34% of M145 and $\Delta 5592$.

Figure 6.10 Production of undecylprodigiosin and actinorhodin in liquid TSB:YEME34% of M145 and $\Delta 5592$.

Figure 6.11 $\Delta 5592$ produced browner spores on SFM plates and secreted less actinorhodin from the surface of the colonies.

Figure 6.12 $\Delta 5592$ grew more slowly and produced actinorhodin later on R5+ plates.

Figure 6.13 $\Delta 5592$ sporulated early and more fully on DNA plates.

Figure 6.14 Typical elution profile of His-5592 from a nickel affinity chromatography column and SDS-PAGE of fractions collected during purification.

Figure 6.15 Thrombin cleavage removed the His tag but also degraded SCO5592 by cleavage at an unknown internal site.

Figure 6.16 SDS-PAGE gels of purified recombinant His-5592 treated with formaldehyde (left) or glutaraldehyde (right).

Figure 6.17 Predicted and observed band masses (kDa) for oligomers of His-5592.

Figure 6.18 Gel filtration profile of His-5592 showing one peak corresponding to an oligomer of 10 - 12 subunits.

Figure 7.1 Manganese favours the creation of double-stranded cuts while magnesium favours single-stranded nicks which require extensive end-repair prior to incorporation into the sequencing library.

Figure 7.2 Replicate MnX showed greater fragmentation as the amount of DNase I increased.

Figure 7.3 qPCR primers for DNase-seq in the promoter regions of *afsK* and *sti1*.

Figure 7.4 qPCR primers for *HaeIII*-seq in the coding regions of *afsK*, *sti1*, *hrdB* and *actII- ORF 4*.

Figure 7.5 The copy number ratio of *afsK* to *sti1* increased with digestion level, as measured by qPCR.

Figure 7.6 *SmaI* was used to reduce the average size of fragments in the DNase I-digested samples.

Figure 7.7 DNA concentration and average size of fragments used to generate the sequencing library.

Figure 7.8 Number and quality of reads generated for each sample.

Figure 7.9 Number of non-*SmaI* reads per sample.

Figure 7.10 Cutting index (CI) for each gene along the entire genome for samples MnX0, MnX8, MnY0 and MnY3.

Figure 7.11 Scatter plot of CI against relative expression level for twelve genes.

Figure 7.12 Integrated Genome Browser view of individual aligned cut sites.

Figure 7.13 Cut sites across the genes *afsK*, *sti1* and the rRNA operon *rrnA* visualised in the Integrated Genome Browser.

Figure 7.14 Theoretical set of fragment sizes produced by digesting the *S. coelicolor* genomic DNA insert of cosmid Ste59 to completion with *HaeIII*, ranked by size.

Figure 7.15 DNA fragments produced by *HaeIII*.

Figure 7.16 qPCR was used to measure the copy numbers of four genes in mock-digested and *HaeIII*-digested samples.

Chapter 1: General introduction

The overall aim of this work was to investigate whether global regulators of gene expression could be useful tools in activating cryptic gene clusters in the Actinomycetes. While cluster-encoded local regulators are already manipulated to increase production, the effects of global transcriptional and post-transcriptional regulators have been relatively neglected because these proteins are often challenging to work with and because much of the core nucleoid proteome is poorly characterised in this lineage.

We were particularly keen to identify individual proteins whose loss or over-expression could create a more permissive chromatin regime as this knowledge could be used to improve model heterologous hosts or to facilitate additional or ectopic production in less well-characterised native hosts. Also of interest was the identification of small molecules which affect expression via post-translational modification of nucleoid-associated proteins and whether these could be used to induce production of novel compounds from Actinomycetes.

1.1 Levels of organisation in the nucleoid

While bacterial systems have historically provided most of the paradigms used to understand gene expression at the level of the promoter or operon, the effects of three-dimensional (3D) structure on control above the level of the individual promoter have long been neglected. It was previously believed that the bacterial nucleoid was a simple, featureless mass of almost naked DNA (Struhl, 1999) and its structure is still generally underplayed in textbooks in comparison with eukaryotic chromatin despite great advances in understanding of the action of architectural proteins and the *in vivo* spatial arrangement of the chromosome.

In eukaryotes, differences in chromatin structure allow genetically identical cells to respond to the same set of environmental cues in distinct ways. This is particularly important during differentiation or multicellular development and may be one of the mechanisms underpinning the phenomenon of cellular memory. While the protein occupancy of the bacterial chromosome is much lower than that of nucleosome-bound eukaryotic chromatin, bacterial chromatin and nucleoid structure are now recognised to play an important role in determining the transcriptional programme.

Control of gene expression is a complex multifactorial process. A wide variety of factors such as DNA-binding proteins and polyamines (spermine etc.) interact with DNA in a concentration-dependent manner against a backdrop of physical parameters such as macromolecular crowding, global supercoiling levels and charge shielding by counterions. The response of the transcriptional programme to this collection of globally-acting factors is encoded in the local structure of the DNA of genes themselves, for example different sequences may have different intrinsic curvature or be differently amenable to strand separation (both of these are a feature of AT-rich DNA) and so respond differently to changes in NAP concentration or supercoiling levels.

These levels of organisation are highly interconnected. For example, DNA gyrases alter the level of supercoiling within a topological domain (section 1.1.3), which affects the ability of the supercoiling-sensitive HU protein to bind the promoters of the genes within that domain (Kobryn *et al.*, 1999). DNA gyrase also regulates the amount of HU protein present by controlling the rate at which the *hup* gene is transcribed (Kohno *et al.*, 1994). Conversely HU is able to restrain negative supercoiling in a way that makes it inaccessible to DNA gyrase (Guo and Adhya, 2007; Tanaka *et al.*, 1995). The intracellular concentrations of the different NAPs are also interconnected as they generally control transcription of themselves and multiple other NAPs, making genetic dissection of their individual functions challenging. Additionally, structure and

function are not static: processes such as transcription, translation (see section 1.1.5) and replication will all affect the protein occupancy and 3D positioning of genes.

1.1.1 Compaction

In eukaryotes the genome is located inside the membrane-bound nucleus while the prokaryotic genome occupies the central portion of the cell interior but is not separated from the cytosol. The contour length (length at maximum extension) of an *E. coli* genome is around 1,600 μm while the cell itself is only around 1 μm by 3 – 5 μm , requiring compaction on the order of 100-fold relative to the volume it would occupy upon cell lysis. This is achieved through the synergistic effects of macromolecular crowding, supercoiling (section 1.1.3) and DNA-bound proteins (section 1.2). Compaction is all the more important when the nucleoid must be packaged to fit inside a spore which is many times smaller than a vegetative cell.

Macromolecular crowding is the force which induces mixtures of polymers of different physicochemical properties or diameters to spatially assort themselves by type. It occurs when globular macromolecules such as cytosolic proteins or polymers such as PEG/Dextran are excluded from narrow spaces between closely-apposed long polymers such as DNA. The proteins therefore exert greater osmotic pressure on the DNA from the outside, pushing duplexes closer together and inducing phase separation (de Vries, 2010). This effect is enhanced by the adherence of architectural proteins and nascent mRNA which effectively thicken the DNA filaments. However, the contribution of macromolecular crowding to compaction of a functional nucleoid *in vivo* remains largely theoretical.

Compaction of isolated nucleoids can be maintained for a short time *in vitro* using buffers containing non-physiological concentrations of salt and polyamines (to overcome the repulsive effects of negative charges on the DNA phosphate backbone) but is most likely regulated *in vivo* by a combination of macromolecular crowding, supercoiling and protein occupancy. While many studies measure compaction level of the nucleoid as a whole (i.e. the nucleoid diameter), it is important to note that compaction is not necessarily uniform over the nucleoid. Factors antagonising compaction in local areas may include the binding of certain NAPs (e.g. HU), cotranslational insertion and active transcription.

1.1.2 Gross nucleoid morphology and positioning

Due to the small size of bacteria relative to eukaryotic cells, the resolution of light microscopes is insufficient to generate detailed images of live or fixed nucleoids. There is currently somewhat of a disconnect between top-down studies (e.g. microscopy, FISH) and bottom-up studies (e.g. biochemical characterisation of NAPs). Combinations of approaches, including new high-throughput sequencing technologies and perturbations of structure with chemicals of known action, can be a way to bridge this gap (Dame and Dorman, 2010).

Light microscopy of nucleoids is technically challenging as preparation of these fragile and dynamic structures can generate artefacts during fixation or due to photolytic degradation (Zimmerman, 2006). Transmission electron microscopy of bacterial sections stained in osmium tetroxide (OsO_4) shows the nucleoid to be a weakly-staining, lobed mass occupying the centre of the cell (Figure 1.1, left). Phase contrast microscopy can be combined with fluorescence microscopy (e.g. against the general DNA stain DAPI) to acquire low-resolution information about nucleoid behaviour, e.g. gross abnormalities in positioning in the presence of antibiotics or defects in partitioning upon cell division. Smaller structural units of the nucleoid can be studied using atomic force microscopy (AFM), which uses a minuscule cantilevered tip to “feel” a surface onto which nucleoids or nucleoprotein complexes have been attached. This does not require staining and also circumvents the diffraction limit, resulting in nanometre-scale resolution. It can be used to visualise the thicknesses of chromatin fibres (Kim *et al.*, 2004) and even smaller structures such as the types of nucleoprotein complexes that different combinations of NAPs can form on DNA (Maurer *et al.*, 2009; van Noort *et al.*, 2004).

Cytogenetic techniques such as the use of fluorescently-labelled nucleotide or protein probes are useful tools in dissecting the *in vivo* structure of the nucleoid. The positions of particular loci can be revealed by fluorescent *in situ* hybridisation (FISH) or by inserting an array of *lacO* (or *tetO*) sequences into the desired location and detecting binding of a LacI-fluorophore (or TetR-fluorophore) fusion protein that binds to them (Umbarger *et al.*, 2011; Wang *et al.*, 2011). For example the region of the *E. coli* chromosome containing the origin of replication was found by FISH to be located at the mid-cell position during replication but moved to the cell poles following duplication, while the region containing the terminus of replication was found at the quarter points following duplication (Niki and Hiraga, 1998).

Techniques developed for working with eukaryotic chromatin which take advantage of the enormous recent improvements in sequencing have been applied to the bacterial nucleoid in a small number of cases with good results. Chromatin conformation capture (3C) involves shearing

and religating cross-linked chromatin, such that distant DNA sequences held in physical proximity by chromatin structure will preferentially be ligated together. The rates at which different pairs of sequences are joined is then measured by qPCR. Recently a detailed view of the *Caulobacter crescentus* nucleoid was mapped using 5C (an adaptation of 3C incorporating high-throughput sequencing instead of qPCR), revealing it to be in the shape of a twisted ellipse (Figure 1.1 , right). The orientation of the nucleoid relative to the cell poles could be altered by transplanting the *parS* sites (which are anchored via ParB to the cell pole) from the origin of replication to a location around 400 kb away. Surprisingly, this occurred without strongly affecting gene expression (Umbarger *et al.*, 2011).

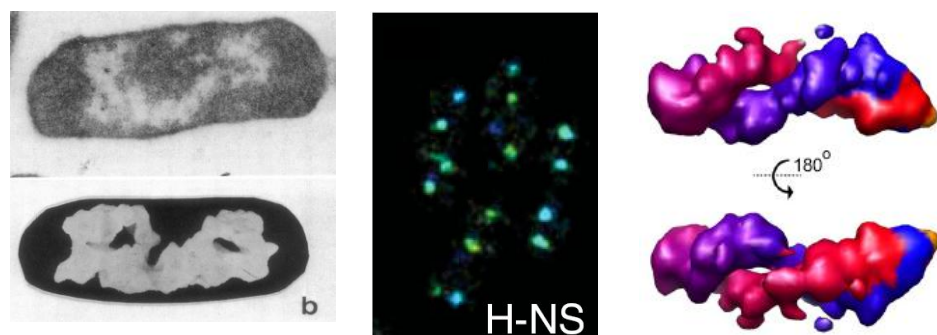


Figure 1.1 The gross structure of the bacterial nucleoid as visualised by transmission electron microscopy (*E. coli*, left), by imaging a fluorescently-tagged architectural protein H-NS (*E. coli*, centre) and by chromatin conformation capture (*Caulobacter crescentus*, right).

Another technique adapted from eukaryotic chromatin studies is the use of non-specific nucleases (e.g. DNase I) to partially digest the chromosome. This can differentiate between transcriptionally active, open regions (more cuts) and transcriptionally inactive, closed regions (fewer cuts). Chapter 7 describes an attempt to use this approach to probe the structure of the *S. coelicolor* chromosome.

A number of small-molecule tools are available for perturbing nucleoid structure. Repeatable changes in nucleoid function can be induced using certain antibiotics: quinolones block the action of DNA topoisomerases, rifampicin halts transcription by binding the β subunit of RNA polymerase and treatment with chloramphenicol halts protein synthesis and therefore also transcription, with the result that nucleoids are quickly compacted. The terpenoid camphor has been shown to cause decondensation of *E. coli* nucleoids, although the mechanism by which this happens is unclear (Harrington and Trun, 1997; Hu *et al.*, 1996).

Some of the most illuminating results have been generated by combining techniques to study a well-defined hypothesis. Wang *et al.* (2011) used super-resolution fluorescence microscopy to

observe the spatial arrangements of five NAPs in the *E. coli* nucleoid (HU, Fis, IHF, StpA and H-NS). While most were evenly distributed, H-NS was seen to form a pair of distinct clusters per nucleoid which is consistent with its assumed role of sequestering silent genes in the centre of the nucleoid. They then used a *tetO*/TetR-fluorophore system to show that known H-NS target genes were indeed associated with these clusters in wild-type cells and shifted location in an *hns* null strain, while genes not targeted by H-NS did not shift. Finally they used 3C to show that H-NS target genes were preferentially associated with one another in live cells. This approach is particularly useful for proteins where there is ample information on target genes, mutant phenotypes and biochemical modes of action.

1.1.3 Supercoiling, topoisomerases and topological domains

Supercoiling is the over- or under-twisting of the DNA helix relative to its relaxed B-form state. It is greatly affected by many NAPs, which are discussed individually in more detail in section 1.2. Negative supercoiling (under-twisting) is associated with exponential growth phase and transcriptional competence and is also generated by the passage of replication forks, where it is removed by topoisomerase I (Bates and Maxwell, 2005). Relaxed DNA is associated with stationary phase and its concomitant transcriptional quiescence. Positive supercoiling (over-twisting) is found ahead of replication forks (requiring relaxation by DNA gyrase) and has also been observed in some hyperthermophilic Archaeal species as a putative means of protecting DNA from damage and destabilisation during extremes of temperature (Perugino *et al.*, 2009).

The global level of supercoiling is thought to form a link between growth phase (or environmental conditions) and the transcriptional programme (Hatfield and Benham, 2002; Travers and Muskhelishvili, 2005). Negative supercoiling causes torsional stress in the DNA duplex which favours localised melting, especially of AT-rich tracts (Benham, 1979), in a process known as stress-induced duplex destabilisation (SIDDD). This reduces the amount of energy required to melt the labile AT-rich -10 promoter element which occurs during conversion of the RNA polymerase closed complex to the open complex which is necessary for initiation of transcription (Drew *et al.*, 1985). However not all promoters respond to supercoiling in the same way. Expression of around 7% of *E. coli* genes were found to change their expression levels in response to artificially-induced DNA relaxation during exponential growth phase (Peter *et al.*, 2004). These genes were dispersed throughout the genome and represent many different functional classes, including *fis*, tRNA synthetases and several topoisomerases.

The level of supercoiling is homeostatically controlled by the topoisomerases: enzymes that break one (type I) or both (type II) strands of the duplex then rotate the ends relative to one another before covalently rejoining them (see Champoux (2001) or Bates and Maxwell (2005) for a review of their mechanisms). As well as controlling the global level of superhelicity they function during chromosome replication to alleviate excess positive supercoiling ahead of and negative supercoiling behind the replication fork, and then to decatenate the resulting linked circular DNA molecules. DNA gyrase is a prokaryotic-specific type II topoisomerase which uses ATP to introduce negative supercoiling or remove positive supercoiling. It links the nutritional status of the cell to transcriptionally-relevant supercoiling level because its activity is highly responsive to the cellular energy charge (simplistically represented by the $[ATP] / [ADP]$ ratio). Gyrase interacts with DNA at specific locations via intergenic elements which contain multiple imperfect palindromes, known as bacterial interspersed mosaic elements (BIMEs). These sites have been shown to undergo double-stranded breakage *in vivo* by gyrase when cells are treated with oxolinic acid (which prevents gyrase from re-ligating the two broken ends it has generated) but each different BIME sequence is cut at a different rate (Espeli and Boccard, 1997). These sites are distributed unequally across the genome, with more being present near the origin of replication (Jeong *et al.*, 2004; Sobetzko *et al.*, 2012). It has been suggested that these elements may be equivalent to the eukaryotic matrix attachment regions (MARs) which anchor the chromosome to the nuclear matrix and also interact with topoisomerases (Hsu *et al.*, 2006). Some gyrase binding sites (particularly those at the end of transcriptional units) are also bound by IHF, which may regulate how efficiently the gyrase is loaded onto the DNA (Clarkson and Bates, 1996). HU is also thought to modulate topoisomerase activity by stimulating gyrase activity, perhaps by creating bends which facilitate the wrapping of DNA around the gyrase (Knight and Samuels, 1999; Malik *et al.*, 1996) or by inhibiting the activity of topoisomerase I (Bensaid *et al.*, 1996).

Around half of the supercoiling in a nucleoid is unconstrained (free), i.e. the supercoiling tension it contains is released by nicking the DNA. The other half of the supercoiling in a nucleoid is constrained by proteins which alter the path of the DNA upon binding such as HU (Hsieh *et al.*, 1991; Tanaka *et al.*, 1995), Fis (Rochman *et al.*, 2002), Lrp (Beloin *et al.*, 2003) and H-NS (Arold *et al.*, 2010; Mojica and Higgins, 1997). The superhelical tension of constrained supercoiling is not released by nicking and constrained supercoils are less accessible to topoisomerases. Constrained DNA may form toroidal structures (e.g. HU) or plectonemic (interwound) loops which are stabilised by proteins such as H-NS (Arold *et al.*, 2010). The relative concentrations of NAPs bound to the DNA determines the proportions of DNA present in either form, as was elegantly visualised by Maurer *et al.* (2009) using AFM (Figure 1.2).

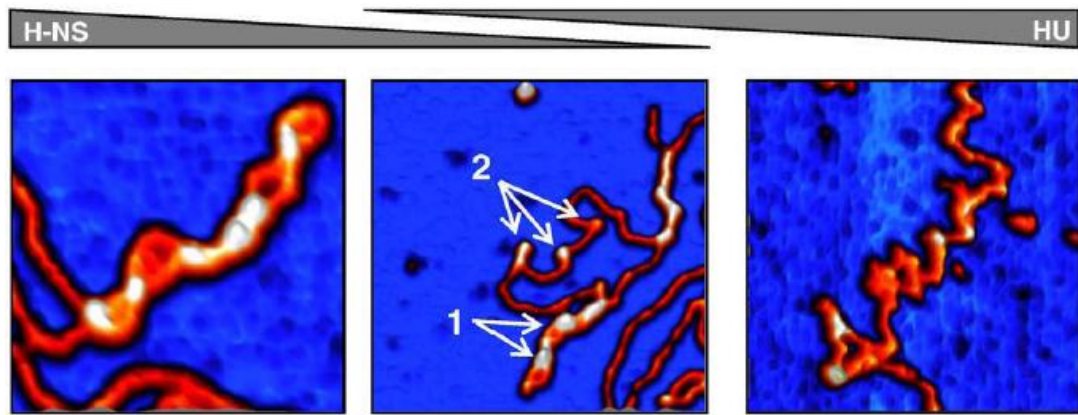


Figure 1.2 Atomic force microscopy of nucleoprotein complexes showing supercoiling in an unconstrained form in plectonemic loops stabilised by H-NS (left) and as toroidal structures stabilised by HU (right). Where both proteins are present (middle) both the plectonemic loops (1) and toroidal structures (2) may co-exist. Figure from Maurer *et al.* (2009).

Supercoiling is not distributed evenly throughout the nucleoid: the chromosome is divided into numerous discrete topological domains (sometimes referred to as loops), with the level of supercoiling within each domain being independent of the levels in the two adjoining domains. The apparent size of these domains depends in part on the method used to measure them. Early estimates of *E. coli* domain sizes were large (approximately 100 kb) based on electron microscopy of loops visible in isolated nucleoids (Kavenoff and Bowen, 1976; Kavenoff and Ryder, 1976) and on experiments measuring the number of nicks needed to fully relax the chromosome (Sinden and Pettijohn, 1981). Estimates based on gyrase-induced fragmentation of the entire genome using oxolinic acid (or the related antibiotic norfloxacin) estimated the domains to be approximately 50 - 100 kb (Brunetti *et al.*, 2001; Hsu *et al.*, 2006). However the estimates generated from gyrase-induced fragmentation rely on several assumptions: there must be no more than one break per topological domain and there must be a break in every topological domain. A more recent estimate gave a smaller domain size (approximately 10 kb) based on an assay using supercoiling-sensitive reporter genes to measure the rate of spread from a nicked site to the nearest domain boundary in combination with electron microscopy of isolated nucleoids (Postow *et al.*, 2004).

Topological domains have domain barriers to prevent the diffusion of free supercoils. In eukaryotes, topological loops are tethered at both ends to the nuclear scaffold via DNA-encoded (*cis*) elements such as scaffold/matrix-attachment regions (S/MARs) which effectively form boundary elements, but it is not known whether similar structures exist in bacteria. It has been suggested that the loops may simply be tethered by DNA-bridging proteins such as H-NS, as any physical bridge between two sites imposes a topological restraint on the intervening region.

Noom *et al.* (2007) used ChIP-on-chip to locate H-NS binding sites across the genomes of *Salmonella typhimurium* and *E. coli*, revealing approximately 350 patches of H-NS per genome of approximately 2 kb each. The average inter-patch distance (corresponding to topological domain size) was approximately 11 kb in both species, corresponding to the estimate of Postow *et al.* (2004), although there was a large degree of variation in domain size. The authors estimate this could account for around 85% of the topological domain barriers.

All nucleoids are thought to contain supercoiled DNA but not all of the knowledge generated in *E. coli* will be directly applicable to all species. For example *Salmonella* is closely related to *E. coli*, and has the same complement of topoisomerases but its global supercoiling levels showed much less variation in response to changing conditions, to which its promoters have apparently adapted (Cameron *et al.*, 2011). The contribution of supercoiling to the transcriptional programme in *Streptomyces* is a neglected topic but there is reason to expect that it is an important factor. Isolated *Streptomyces hygroscopicus* nucleoids were probed with the intercalating dye ethidium bromide and were found to contain similar global levels of supercoiling (1 negative turn per 200 bp) as the *E. coli* chromosome (Sarfert *et al.*, 1983) but the topological domain structure has never been investigated and fewer of its supercoiling-sensitive promoters have been characterised. Induction of expression from the *dpsA* promoter during osmotic stress (but not heat stress) was found to be dependent on an increase in supercoiling mediated by gyrase B (Facey *et al.*, 2011).

1.1.4 Macrodomains and genome organisation

On a larger organizational scale than the topological domains are the ~1 Mb chromosomal macrodomains (Figure 1.3, left). In *E. coli* the Ori and Ter macrodomains (containing the origin and terminus of replication respectively) had previously been described on the basis of their characteristic cellular positions throughout the cell cycle (Niki and Hiraga, 1998). Two further structured macrodomains (Left and Right) and two less-structured regions were defined using the λ Int recombination system to measure the frequency of collisions between *att* sites inserted at different positions in the chromosome (Valens *et al.*, 2004). Recombination was found to occur preferentially within macrodomains, rather than between them, suggesting that they are somehow physically insulated from one another. Exponentially-growing cells contain multiple copies of the genome, however interactions between these sister chromatids were found to be infrequent. The functional differences between the four structured macrodomains and the two less-structured regions are not clear but the cellular positions of the less-structured regions were found to be 2 - 4 times more spatially mobile (Espeli *et al.*, 2008).

Allen *et al.* (2006) used wavelet analysis of DNA sequences to detect periodic patterns in the arrangements of many microbial genomes, looking at characteristics such as GC content, gene density and codon usage over multiple period lengths (70 – 1,500 kb) across the genomes. Gene density did not show as strong a pattern as has been seen for eukaryotes, perhaps because bacteria devote less of their genome to untranslated regions and repetitive elements. Most strikingly, at a 600 - 650 kb scale the macrodomain structure of *E. coli* correlated with patterns of gene expression level, gene essentiality and evolutionary retention (Figure 1.3, right). Genes in the middle of macrodomains or less-structured regions were more likely to be highly expressed, essential and conserved. The level of patterning seen for each characteristic was variable between species. No *Streptomyces* species were analysed but patterns were seen in several Actinobacteria including *Mycobacterium*, *Nocardia* and *Leifsonia*, particularly for codon usage.

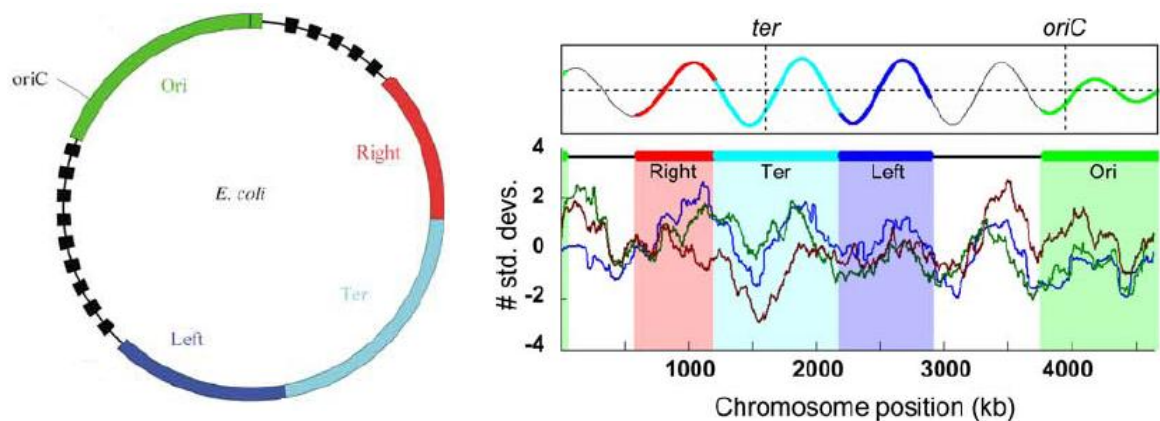


Figure 1.3 Left: macrodomain structure of the *E. coli* chromosome as defined by Valens *et al.* (2004). Ori (green), Right (red), Left (dark blue) and Ter (light blue) are structured macrodomains while the two less-structured regions are represented by black squares. Right: correlation of macrodomain layout with long-range patterns in gene expression level (blue), gene essentiality (green) and gene evolutionary retention (red) from Allen *et al.* (2006).

The determinants which define and maintain insulation of the macrodomains are not fully known. The major NAPs bind throughout all macrodomains and are not known to contribute to the differences between these structures. RNAP preferentially associates with the Ori macrodomain but this is most likely due to the presence of the highly-transcribed rRNA operons in that region. However some proteins have been identified which govern the behaviour of the Ori and Ter macrodomains (see Dame *et al.* (2011) for a review). Macrodomain-specific proteins include MatP (enriched in the Ter macrodomain) and SeqA and SlmA (excluded from the Ter macrodomain). These help to organise the chromosome but are distinct from NAPs in that they bind with high specificity and do not affect gene expression but instead have roles in chromosome replication and segregation. SeqA limits initiation of replication and may help orient the chromosome during

cell division (von Freiesleben *et al.*, 1994). SlmA is TetR-like nucleoid-occlusion factor which prevents Z-ring formation (and therefore cytokinesis) across an unsegregated chromosome. It is sequence-specific and has binding sites in the Ori, Right and Left (but not Ter) macrodomains in *E. coli*, *Salmonella* and *Klebsiella* (Tonthat *et al.*, 2011), suggesting that this mechanism is conserved. MatP prevents premature chromosome segregation by keeping Ter regions together.

Other patterns of gene organisation are seen at the whole-genome scale which link the transcriptional programme to growth phase. For example the genes nearest the origin are replicated first and are therefore present in greater copy number in growing cells. In rapidly-growing species such as *E. coli* the region around the origin is enriched in ribosomal RNA operons and RNA polymerase (Couturier and Rocha, 2006) while slow-growing species such as *Mycobacterium* do not show this effect (Rocha, 2004). There are also other gradients of transcription-related elements (Sobetzko *et al.*, 2012) such as gyrase binding sites (which are enriched at the origin) and of σ^S -regulated promoters (which are enriched at the terminus).

1.1.5 Gene localisation during transcription

In bacteria the processes of transertion (coupling of transcription and translation to membrane insertion of the nascent protein) and the formation of transcriptional foci both require genes to move to new cellular locations during expression, requiring reorganisation of nucleoid structures.

Transertion is a process unique to prokaryotes whereby mRNA transcripts encoding membrane proteins are simultaneously transcribed and translated, with the transmembrane domain of the nascent peptide being inserted before transcription has terminated. Recently Libby *et al.* (2012) measured this movement directly for two membrane proteins and observed a shift from mid-cell (the centre of the nucleoid) to a position near the membrane upon transcription, which did not occur when the membrane protein gene was replaced with a gene of similar expression level which encoded a non-membrane protein. Surrounding regions were also perturbed, with a shift in position seen for genes up to 90 kb away. The loss of these membrane-tethering proteins is thought to explain the collapse of the nucleoid into the centre of the cell seen when translation-inhibiting antibiotics are added to the cell (Binenbaum *et al.*, 1999). The tethering of these loci to fixed points may also act as barrier to supercoiling diffusion (Deng *et al.*, 2004; Lynch and Wang, 1993), perhaps accounting for some of the missing 15% of topological domain barriers discussed in section 1.1.3 (Noom *et al.*, 2007).

Fluorescent labelling of *E. coli* or *B. subtilis* RNA polymerase revealed that transcription does not occur uniformly throughout the nucleoid. In exponentially growing cells the polymerase is concentrated in a small number of foci (approximately 3 per cell) where highly-expressed genes such as those encoding rRNA are transcribed (Berger *et al.*, 2010; Lewis *et al.*, 2008). Following activation of the stringent response there is a sharp decline in stable RNA production and the foci disperse. Formation of these foci is dependent on the NAP HU (Figure 1.4), perhaps because HU stabilises and thereby maintains the negative supercoiling generated behind the transcription bubble, making it more likely that a new round of transcription will begin and facilitating a high rate of stable RNA expression.

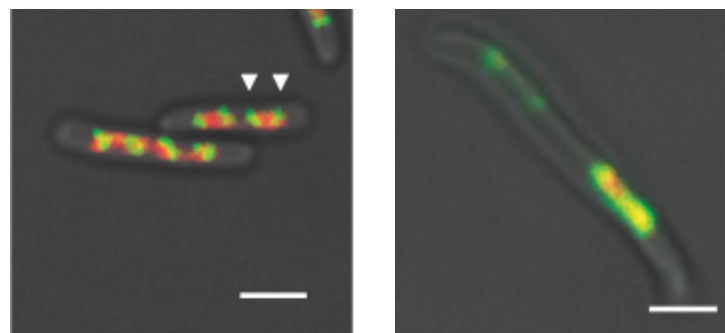


Figure 1.4 Transcriptional foci in exponentially-growing *E. coli* wild-type cells (left) and *hupA/B* mutant cells (right). Both are merged images showing phase-contrast image (greyscale), an RpoC-YFP translational fusion protein (green) and DNA stained with DRAQ5 (red). White arrows indicate transcriptional foci. Scale bars represent 2 μ m. Figure from Berger *et al.* (2010).

1.1.6 Local structural considerations

In addition to roles in organising the nucleoid on a large scale, NAPs also affect the function of individual genes or operons by altering the 3D shapes of their promoters. A NAP will often bind to a particular element within the promoter, either causing a kink in the DNA which affects protein-protein interactions (e.g. of transcription factors with polymerase), or by occluding promoter elements, or by causing a change in the local supercoiling level which affects transcriptional initiation. For reviews see McLeod and Johnson (2001) or Dillon and Dorman (2010).

NAPs can block transcription from a promoter by binding across the -35 and/or -10 elements and thereby preventing access to RNA polymerase. Fis (section 1.2.4) regulates many genes in this way, most famously the *E. coli gyrA* and *gyrB* promoters (Schneider *et al.*, 1999). H-NS (section 1.2.2) can also suppress transcription by preventing polymerase from binding, or by binding across the -35 element of the *E. coli* rRNA P1 promoter in such a way that polymerase can bind and form an open complex but not elongate transcripts beyond the first three nucleotides (Figure

1.5, middle). A dimer of *E. coli* HU (section 1.2.1) binds close to the transcriptional start site in the *gal* promoter, creating a sharp kink in the DNA which allows two dimers of the GalR repressor to interact (Figure 1.5., left) and results in loss of access of the promoter to the transcriptional machinery (Lewis *et al.*, 1999).

Conversely NAPs can facilitate transcription. IHF (section 1.2.1) activates the *ilvP_G* promoter using SIDD (see section 1.1.3) to displace torsional strain to the -10 element from an upstream site, causing melting which favours transcription (Sheridan *et al.*, 1998). IHF can also induce a sharp bend in the DNA which can facilitate interactions between transcription factors, such as in the *Citrobacter tpi* promoter where three dimers of TyrR are brought together to activate transcription (Figure 1.5, right) (Bai and Somerville, 1998), in a very similar but oppositely-acting mechanism to the HU-mediated repression of the *gal* operon described previously. In a manner similar to standard transcription factors, some NAPs can contact subunits of the polymerase holoenzyme to stimulate transcription. For example Fis contacts the C-terminal domain (CTD) of the RNA polymerase α -subunit at the *rrnB* P1 promoter (McLeod *et al.*, 1999) and at the *proP* promoter (McLeod *et al.*, 2000). Many, if not most, promoters use a combination of transcription factors and NAPs which work in concert or in opposition to regulate transcription. NAPs sharing a binding site can act in opposition, for example Fis binding to upstream sites can relieve H-NS-mediated repression of the *E. coli rrnB* P1 promoter (Dorman, 2004).

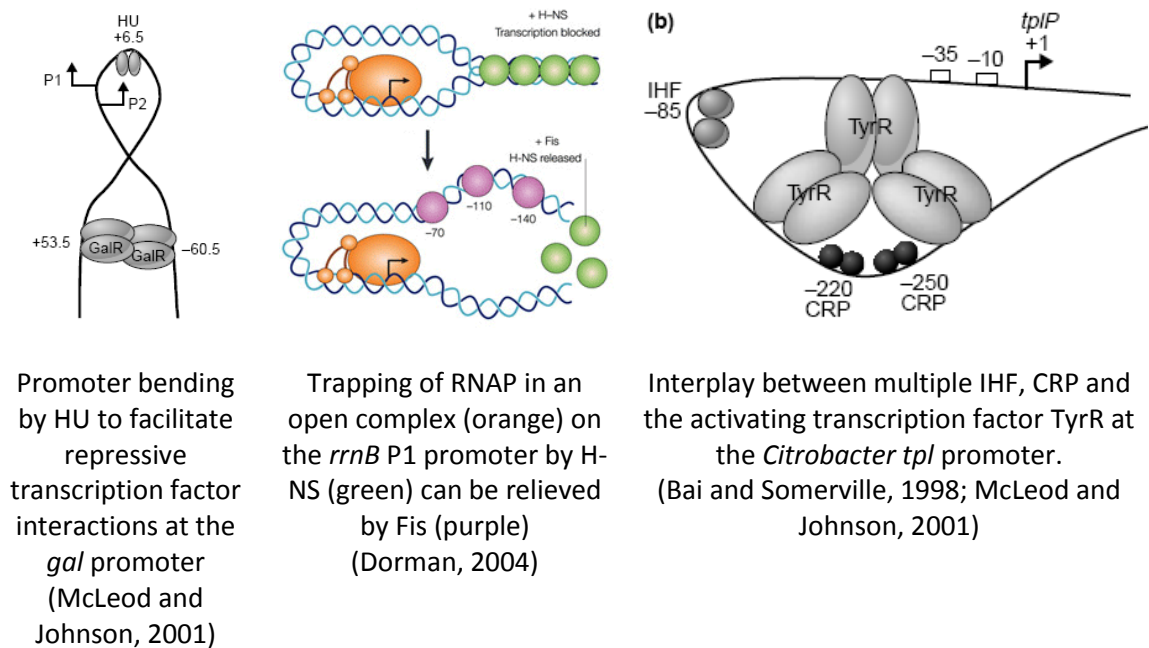


Figure 1.5 Control of gene expression by NAP-mediated modification of promoter architecture.

1.1.7 *Streptomyces coelicolor* chromosome organisation

There are a number of significant differences between the *E. coli* and *Streptomyces* nucleoids (Figure 1.6). The *S. coelicolor* genome is far larger (8.7 Mb) than the *E. coli* genome (4.6 Mb). *E. coli* has a circular chromosome divided into six macrodomains while *Streptomyces* has a linear chromosome whose macrodomain structure is not yet known. In *E. coli* essential and highly-expressed genes are preferentially located around the conserved centre of each macrodomain while *Streptomyces* essential genes (and all six rDNA operons) are found preferentially in the core region of the genome with non-essential and secondary metabolic genes enriched in the 1.5 Mb (left) and 2.5 Mb (right) arm regions (Karoonthaisiri *et al.*, 2005).

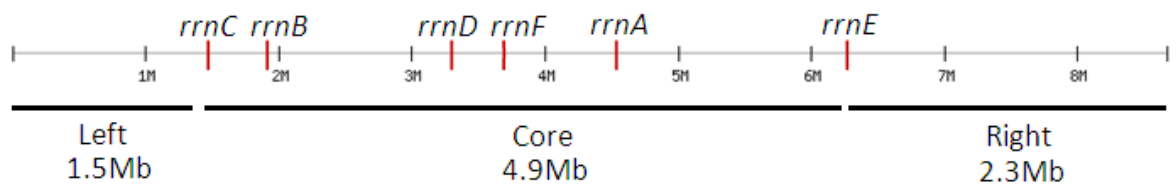


Figure 1.6 Chromosome layout of *S. coelicolor*. All six rRNA operons are marked in red and the conserved core domain and two variable arm domains are underlined in black.

While the sequence of the genome is available (Bentley *et al.*, 2002), therefore patterns of distribution of features such as the rRNA operons can be studied, very little is known about the nucleoid as a living entity: its major protein composition, its 3D arrangement and its topological domain structures are currently a mystery. While the chromosome is linear, it behaves as if it were circular when probed by recombination or microscopy due to physical joining of the ends by TIR proteins (Tsai *et al.*, 2011). In addition to the chromosome, wild-type *S. coelicolor* harbours two large plasmids: a 31 kb circular plasmid called SCP2 and a very large 350 kb linear plasmid called SCP1 which also carries TIR sequences at its termini (Kinashi *et al.*, 1992). McArthur and Bibb (2006) demonstrated a relationship between structure and function of the genome by showing a correlation between susceptibility to digestion by DNase I (a non-specific nuclease) and transcriptional activity. This is discussed in greater depth in chapter 7.

It is perhaps incorrect to refer to “the” *Streptomyces* nucleoid because it shows such large differences in gross structure depending on developmental stage: the nucleoids are large and heterogeneous in the vegetative mycelium, with around a dozen chromosomes per multinucleate compartment (Figure 1.7, left). Later in the aerial mycelium large, rod-like nucleoids containing multiple copies of the chromosome subdivide into compact, spherical spore nucleoids (Figure 1.7, below right) containing a single copy of the chromosome (Hopwood and Glauert, 1960b).

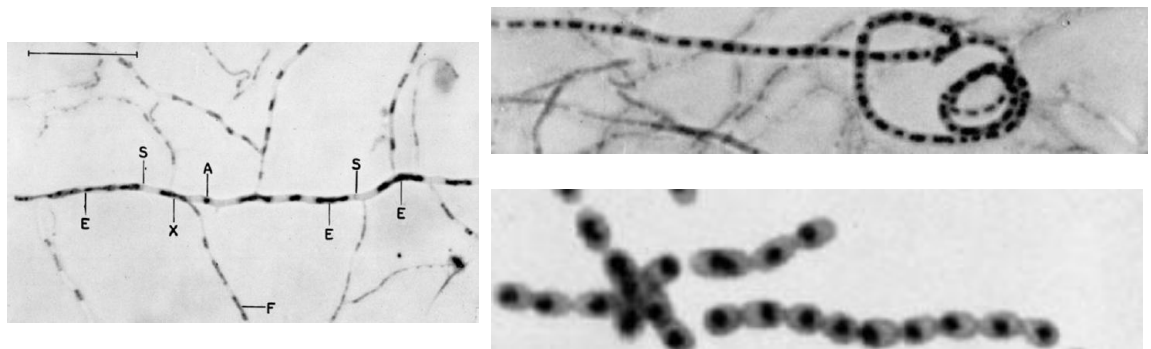


Figure 1.7 Electron micrographs showing *S. coelicolor* nucleoids at different stages of development. Left: substrate mycelium hypha with side branches. Above right: aerial hypha in which spores are forming. Below right: spore chains. From Hopwood and Glauert (1960b).

1.2 Nucleoid-associated proteins (NAPs)

Nucleoid-associated proteins (NAPs) are small, highly abundant DNA-binding proteins of a variety of protein folds which both regulate gene expression and have architectural roles in many nucleoid processes. Unfortunately NAPs are like obscenities: it is difficult to define them precisely, but you know them when you see them. It is not possible to give a definitive list of criteria for categorising a protein as a NAP because they are so diverse but there are a number of ways in which, on the whole, they differ from conventional sequence-specific transcription factors (Figure 1.8). NAPs are highly abundant, with 10,000s of monomers per cell (Ohniwa *et al.*, 2011; Osuna *et al.*, 1995; Schroder and Wagner, 2000), and bind with low or no sequence specificity. In contrast, transcription factors are much less abundant (dozens of monomers) and bind with higher sequence specificity, often to a definable consensus sequence. Peterson *et al.* (2007) were able to quantify the specificity of Lrp (section 1.2.3), finding a 20-400-fold difference in affinity between specific and non-specific sites whereas transcription factors showed a 10,000 – 10,000,000-fold difference. Transcription factors commonly bind as monomers or dimers while NAPs may bind in variable numbers with a high degree of co-operativity. Transcription factors are limited to roles in transcription while NAPs are also used in DNA processes as diverse as replication (Atlung and Hansen, 2002), segregation, recombination, DNA repair (Kamashev and Rouviere-Yaniv, 2000), phage integration (Pedulla *et al.*, 1996) and transposition.

NAPs regulate many genes, both directly and indirectly, and knockout mutants generally show changes in expression of around 5 - 10% of their genomes (Bradley *et al.*, 2007; Mangan *et al.*, 2006; Muller *et al.*, 2006; Prieto *et al.*, 2012). To be classified as a NAP the protein must also alter the structure of the DNA at a local or global level instead of, or sometimes as well as, affecting transcription by making direct contact with the polymerase. Their biological roles are often difficult to delineate because mutants are highly pleiotropic, with defects in processes including cell morphology, metabolism, growth rate, stress response and pathogenicity. In addition some NAPs have non-DNA-binding roles in the cell wall (e.g. Mdp1, section 1.2.1) or in combating oxidative damage under stressful conditions (e.g. non-DNA-binding versions of Dps, section 1.2.5). The “core” set of NAPs of *E. coli* usually includes HU, IHF, H-NS, Fis, Lrp and Dps (Dorman and Deighan, 2003; Maurer *et al.*, 2009) but sometimes the definition is extended to include chromosome-organising proteins such as DnaA, CbpA, SMC (Dillon and Dorman, 2010; Luijsterburg *et al.*, 2006) and IciA or proteins which are primarily RNA chaperones such as StpA and Hfq (Azam and Ishihama, 1999). Proteins such as CRP which have exceptionally large regulons are classified as global regulators when there is little or uncertain evidence that they alter chromatin structure, or if this property has never been tested.

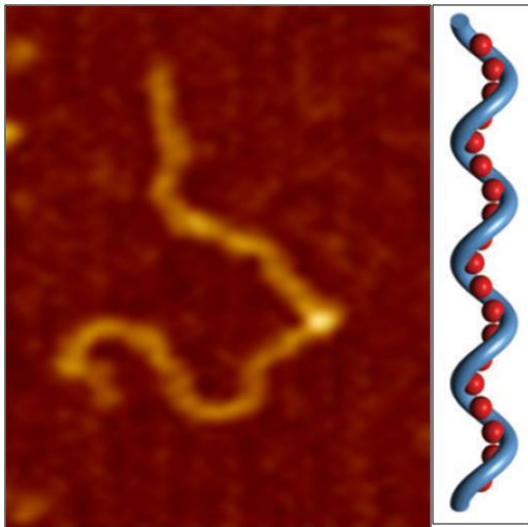
	Transcription factors	Nucleoid-associated proteins
Abundance	10s of monomers	10,000s of monomers
Specificity	Small number of specific sites 10^3 - 10^7 -fold greater for specific than non-specific sites	Very large number of non-specific sites 20 – 400-fold greater for specific than non-specific sites (Peterson <i>et al.</i> , 2007)
Consensus sequence	Usually definable (but individual sites may fit the consensus better or worse)	Not usually definable but some (e.g. IHF) have some specificity by indirect readout or recognise structures (e.g. curved DNA) instead of sequences
Mode of binding	Usually one mode Often homo- or hetero-dimeric	Multiple modes Often highly co-operative
Mode of regulation	Bind to promoter and contact the transcriptional apparatus directly	Bind within promoters and coding regions in heterogeneous nucleoprotein complexes, alter promoter architecture or occlude TF binding sites
Role/s	Limited to transcription	Many DNA transactions

Figure 1.8 Contrasting properties of nucleoid-associated proteins and local transcription factors.

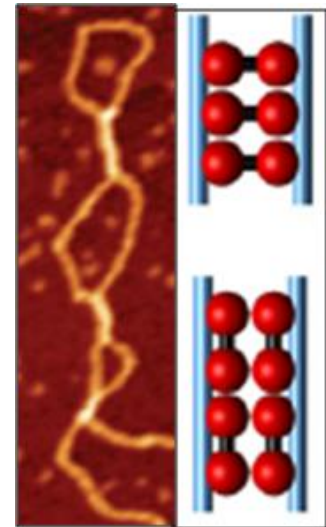
There are multiple challenges associated with working with NAPs as compared to conventional transcription factors. NAPs are difficult to study genetically in isolation because they form a dense cross-regulatory network with each other, so that a knockout of one NAP will often have altered levels of the other proteins. They also work in concert with the topoisomerases and NAP deletion may cause compensating secondary mutations in those, for example an *E. coli* HU mutant was found to be spontaneously suppressed by spontaneous mutations in the gyrase B subunit (Malik *et al.*, 1996). While discrete binding sites may be located by techniques such as ChIP, these generally lack a consensus binding sequence and the fold difference between a weak binding site and background may be lower than between a conventional transcription factor's binding site and non-binding site. They are also more challenging to investigate at the biochemical level as some of the factors which affect binding are different between *in vivo* nucleoids and *in vitro* binding experiments, for example the degree of supercoiling in a promoter or the formation of large-scale nucleoprotein complexes. There is also a high occurrence of co-operativity between dimers which form larger nucleoprotein complexes of variable sizes, for example H-NS dimers bind at a nucleation site spanning a pair of DNA duplexes which are then “zipped up” in long train-track-like structures by additional protein units (Arold *et al.*, 2010; Bouffartigues *et al.*, 2007; Dame *et al.*, 2006). It is not possible to state “the” affinity of a NAP for DNA because there are generally many

different preferred binding sites of a range of affinities and because the affinity of an oligomer (e.g. four Lrp dimers) for a DNA segment will be a factor of the individual dimer binding sites and the degree of co-operativity possible at that site. To complicate matters even further, many species have several homologues of a given NAP within one genome which are partially redundant and may form heterodimers with one another.

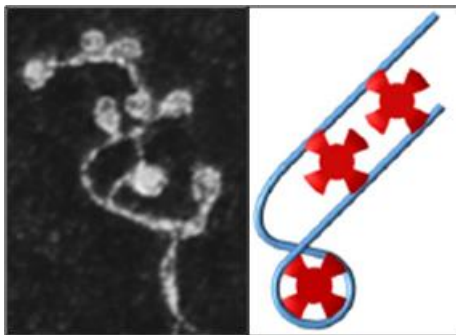
Most of the research discussed here was carried out in *E. coli* and *Salmonella* because these are the species for which the most is known. Where possible, their counterparts in *Streptomyces* and also in *Mycobacterium* (another Actinomycete) are also discussed. It is not always possible to make useful comparisons between lineages because the roles of NAPs on phenotype are very varied, but the biochemical modes of action of individual NAPs and also their general functions are often retained. NAPs are organised here into classes based on biochemical mode of action (Figure 1.9): DNA-bending (1.2.1), DNA-bridging (1.2.2), ligand-binding (1.2.3), DNA-wrapping (1.2.4) and DNA-protecting (1.2.5). Global regulators of gene expression at the post-transcriptional level are discussed separately (section 1.3) although there is considerable similarity and overlap between these areas.



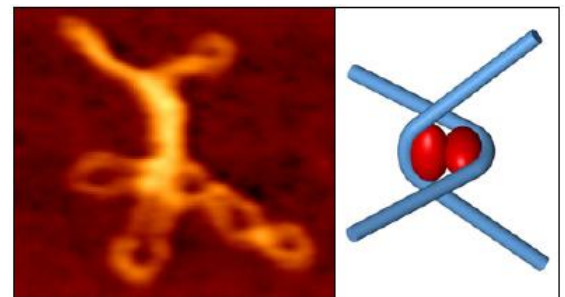
DNA-bending NAPs: HU, IHF, Mdp1 (page 37)



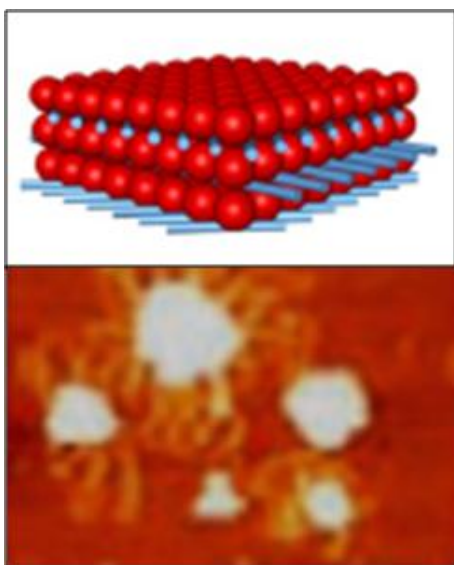
DNA-bridging NAPs: H-NS, StpA, Lsr2 (page 41)



Ligand-binding NAPs: Lrp (page 43)



DNA-wrapping NAPs: Fis (page 45)



DNA-protecting NAPs: Dps, SASPs (page 46)

Figure 1.9 General classes of NAP. Images of nucleoprotein complexes were generated by scanning force microscopy (DNA-bending; DNA-bridging, DNA-wrapping; DNA-protecting) or electron microscopy (ligand-binding). Accompanying diagrams show DNA in blue and proteins in red. Figures from Luijsterburg *et al.* (2006).

1.2.1 DNA-bending proteins: HU, IHF, Mdp1.

HU (heat unstable) and IHF (integration host factor) are the model DNA-bending proteins from *E. coli*. They are small and basic (around 9.5 kDa; pI = around 9.5). *E. coli* and *Salmonella* each possess two paralogues of HU (HU α and HU β) which are developmentally specialised: during vegetative growth the HU α / α homodimer predominates while in stationary phase (Oberto *et al.*, 2009) or cold shock (Giangrossi *et al.*, 2002) the HU α / β heterodimer is upregulated. They also possess two paralogues of IHF (IHF α and IHF β) which form a heterodimer whose properties are distinct from those of HU despite belonging to the same protein family. A larger and more basic homologue, Mdp1 (22.3 kDa ; pI = 11.2), is found in *Mycobacterium* which comprises an N-terminal HU domain coupled to a long lysine-rich tail. *Streptomyces* has one short HU (SCO2950; HupA; 9.9 kDa; pI = 9.5) and one long Mdp1-like HU (SCO5556; HupS; 22.3 kDa; pI = 11.2) which are primarily used during vegetative growth and spore formation respectively.



Figure 1.10 Crystal structure of an IHF heterodimer in complex with DNA. Protein subunits are shown in green and red; DNA backbone is shown in orange; DNA bases are shown in blue. Figure from Rice *et al.* (1996).

HU and IHF both form spacehopper-shaped dimers which comprise a compact alpha-helical body with a pair of flexible beta-sheet arms that lie in the minor groove (Figure 1.10); at the tip of each arm is a proline residue that partially intercalates into the DNA to induce and/or stabilise bending by disrupting base stacking (Rice *et al.*, 1996). Additional contacts may be made by the phosphate backbone against basic patches on the sides of the dimer (Swinger and Rice, 2004). Despite being paralogues, HU and IHF have distinct modes of action: IHF makes a planar 160° bend in DNA while HU makes a non-planar bend with a more variable angle, which explains why HU can constrain supercoiling while IHF cannot. HU is not sequence specific but does have a preference for unusual

DNA structures such as nicks (Balandina *et al.*, 2002) or four-way junctions (Kamashev and Rouviere-Yaniv, 2000) and often occupies defined sites within promoters which can be identified by footprinting experiments. IHF shows a degree of specificity, in that it recognises a degenerate consensus sequence (Goodrich *et al.*, 1990), but it recognises the conformation of the DNA rather than forming specific contacts with the base pairs (Travers, 1997); this is known as indirect readout. For further explanation of the contribution of the different amino acid residues to specificity see Travers (1997) and Lee *et al.* (1992).

HU is generally associated with transcriptionally active DNA (Swinger and Rice, 2004). At supercoiling-sensitive promoters this is likely due to its ability to increase levels of negative supercoiling by stimulating DNA gyrase (Malik *et al.*, 1996) and by constraining negative superhelicity (Tanaka *et al.*, 1995). HU contributes to formation of nucleoprotein complexes in several different ways. Only a single HU dimer is required for Mu transpososome assembly (Lavoie and Chaconas, 1993) or for GalR loop formation (Lewis *et al.*, 1999), where it locally reduces the stiffness of the DNA to allow loop formation. However when multiple dimers bind in close proximity the effect on DNA is more variable. These changes can be studied by adding recombinant HU protein to plasmid DNA *in vitro* and examining the effects by AFM or electron microscopy, however the relative importance of these structures *in vivo* are difficult to assess. At a low protein to DNA ratio the overall effect is to compact the DNA while at a saturating protein to DNA ratio the plasmids become rigidified and appear as wide open circles under AFM (Dame and Goosen, 2002). The dimers can also form octamers which may form higher-orders structures such as spiral filaments around which DNA is wound (Guo and Adhya, 2007).

E. coli HU mutants grow slowly, form a high proportion of anucleate daughter cells (Huisman *et al.*, 1989), are sensitive to UV (Li and Waters, 1998) and γ -irradiation (Boubrik and Rouviere-Yaniv, 1995), are immotile (Nishida *et al.*, 1997) and produce increased amounts of the porin OmpF which leads to hypersensitivity towards detergents and certain antibiotics (Painbeni *et al.*, 1997). HU is essential in *Bacillus* (Micka and Marahiel, 1992). These pleiotropic effects are partly due to widespread defects in fundamental genetic processes such as DNA repair (Miyabe *et al.*, 2000) and a loss of the ability to translate mRNA encoding the sigma factor RpoS (Balandina *et al.*, 2001), which is required to activate specific stress response pathways.

IHF was originally described as a host factor required for bacteriophage- λ integration and it is also involved in gene expression and site-specific recombination (Haniford, 2006). It is able to directly recruit σ^{54} -containing RNA polymerase in the same way as a conventional transcription factor (Macchi *et al.*, 2003) or it can alter promoter architecture to facilitate interactions between

transcription factors and the polymerase (Santero *et al.*, 1992). At the *ilvP_G* promoter, IHF binding displaces torsional strain from a supercoiling-induced duplex destabilization (SIDD) site to the -10 region, facilitating open complex formation and transcription (Sheridan *et al.*, 1998). *ihf* mutants are more sensitive to heat stress and are significantly worse at surviving in prolonged culture (Bykowski and Sirko, 1998). They also produce greater amounts of OmpF (Tsui *et al.*, 1988) and OmpC (Huang *et al.*, 1990). IHF is upregulated during stationary phase and an *ihf* mutant shows reduced expression from many stationary-phase-specific promoters such as those controlling virulence genes, leading to a reduction in ability to invade epithelial cells (Mangan *et al.*, 2006). As IHFa and IHFb function as a heterodimer very little difference is seen between single knockouts of the two IHF subunits, apart from that an *ihfb* mutant was able to support an increased copy number of plasmid p15A during stationary phase while an *ihfa* mutant was not (Hiszczynska-Sawicka and Kur, 1997).

HU/IHF homologues have been identified in almost all bacterial lineages as well as the *Bacillus* phage SPO1 (Geiduschek *et al.*, 1990), red algal chloroplasts (Kobayashi *et al.*, 2002), the *Toxoplasma* / *Plasmodium* apicoplast organelles (Reiff *et al.*, 2012) and even the African swine fever virus (Borca *et al.*, 1996). In many species only a single copy is present but in some cases multiple paralogues are present and the differences in their roles are often unknown. *E. coli* has four paralogues (HU α , HU β , IHF α , IHF β) and *Streptomyces* has two (HupA, HupS). *Bacillus subtilis* has one standard version (HBSu) and one prophage-encoded version (YonN) which cannot compensate for the loss of HBSu despite the two proteins sharing high sequence similarity (Kohler and Marahiel, 1998). Several non-HU proteins are able to complement an HU mutant, for example the DNA-bending eukaryotic HMG1 proteins are able to complement HU (but not IHF) despite sharing no sequence or structural similarity (Bianchi, 1994; Sasaki *et al.*, 2003). Also *Borrelia burgdorferi* produces a unique small protein from the C-terminal part of its DNA gyrase gene that is able to complement an HU mutant (Knight and Samuels, 1999).

The roles of a number of residues on HU function have been studied by mutational analysis, demonstrating that even small changes in the amino acid sequence of HU can give severe pleiotropic effects. Within the DNA-binding arm of *B. subtilis* HBSu, replacement of R58 or R51 with leucine or replacement of K80 or K86 with alanine caused a dramatic loss of affinity *in vitro*, with the R58L mutation causing slowed growth and diminished sporulation *in vivo* (Kohler and Marahiel, 1998). Several residues in the DNA-binding arm of *E. coli* IHF were also found to be important for specificity (Lee *et al.*, 1992). In *E. coli* one mutant version containing just two point mutations (one of which was conservative) were sufficient to change the transcriptional programme, metabolism, cell shape and even colony morphology (Kar *et al.*, 2005).

The single *Mycobacterium* HU, Mdp1, has a long lysine-rich C-terminal tail which increases its affinity for DNA (Kumar *et al.*, 2010), promotes DNA end-joining in the presence of T4 ligase (Mukherjee *et al.*, 2008) and protects the DNA it is bound to against degradation by free radicals generated by the Fenton reaction in a similar fashion to Dps/ferritin (Takatsuka *et al.*, 2011). Expression of Mdp1 is associated with dormancy (Lee *et al.*, 1998) and slower, clumped growth in liquid culture (Lewin *et al.*, 2008). Bizarrely, as well as binding to DNA in a non-specific fashion (Matsumoto *et al.*, 2000) it is also found on the outside of the cells where it regulates cell wall composition (Katsube *et al.*, 2007) and attachment to host cells (Portugal *et al.*, 2008; Shimoji *et al.*, 1999). When expressed in *Mycobacterium*, Mdp1 carries a heterogeneous pattern of mono- and di-methylation marks on the lysine residues in its tail (Petthe *et al.*, 2002), suggesting that its function is regulated by post-translational modification, but the identity of the methyltransferase creating these is unknown. Mdp1 has shown promise as the main component of some commercial applications: the DNA-bound form was found to produce a strong immune response when trialled as a vaccine in mice (Matsumoto *et al.*, 2005) and a PCR-based assay based on the C-terminal tail was found to discriminate well between the closely-related species *M. tuberculosis* and *Mycobacterium bovis* (Prabhakar *et al.*, 2004).

The two HU proteins of *Streptomyces* (HupA and HupS) have been partially characterised. Translational fusion of these proteins to the fluorophore EGFP (enhanced green fluorescent protein) showed that HupA is primarily found in the nucleoids of the vegetative mycelium while HupS has an important role in later stages of development: a $\Delta hupS$ strain was found to produce pale, heat-sensitive spores with decondensed nucleoids (Salerno *et al.*, 2009). Previously a $\Delta hupA$ mutant generated in *S. lividans* was found to grow more slowly during vegetative phase and it was suggested that HupS was upregulated to compensate (Yokoyama *et al.*, 2001) so there may be some degree of redundancy between the two paralogues. It is not known whether the *Streptomyces* HupS lysine-rich tail is methylated in the same way as Mdp1.

No orthologue of IHF from the HU/IHF family of proteins has been identified in Gram positive bacteria. However a putative equivalent from a different family (mIHF/sIHF) has been recognized in Actinomycetes including *Mycobacterium* and *Streptomyces* which is discussed in section 1.2.6.

1.2.2 DNA-bridging proteins: H-NS, StpA and Lsr2

H-NS (histone-like nucleoid-structuring protein) from *E. coli* is the model DNA-bridging NAP. It is slightly larger than HU (15.4 kDa) and is acidic (pI = 5.4). While the requirement for a DNA-bridging function is conserved between species, the role is often fulfilled by functional equivalents rather than homologues. A single copy of the H-NS equivalent *lsr2* is found in *Mycobacterium* alongside one called *hns*. *Streptomyces* has two paralogues of *lsr2* (*sco3375* and *sco4076*) about which nothing is known, but no copies of *hns*.

E. coli H-NS has a C-terminal winged helix-turn-helix DNA-binding domain which binds in the minor groove via an “RGR” motif (Gordon *et al.*, 2011). This is connected by a flexible linker to an N-terminal domain which interacts in an antiparallel fashion with the N-terminal of a second subunit to form a dimer (Esposito *et al.*, 2002). H-NS can either form homodimers or it can form heterodimers with its paralogue StpA (Dorman *et al.*, 1999; Johansson and Uhlin, 1999) or with horizontally-acquired homologues (Dorman, 2010) such as Hfp (in uropathogenic *E. coli*) or Sfh (in *Shigella flexneri*). Interaction of recombinant H-NS with DNA *in vitro* shows that the protein bridges DNA duplexes in an extended nucleoprotein filament, with the DNA held in a plectonemic (interwound) formation (Dame *et al.*, 2005; Maurer *et al.*, 2009). These filaments are thought to form when H-NS dimers bind at high-affinity “nucleation” sites in AT-rich sequences and then spread along the DNA by co-operative binding (Amit *et al.*, 2003) in a “zipping-up” action.

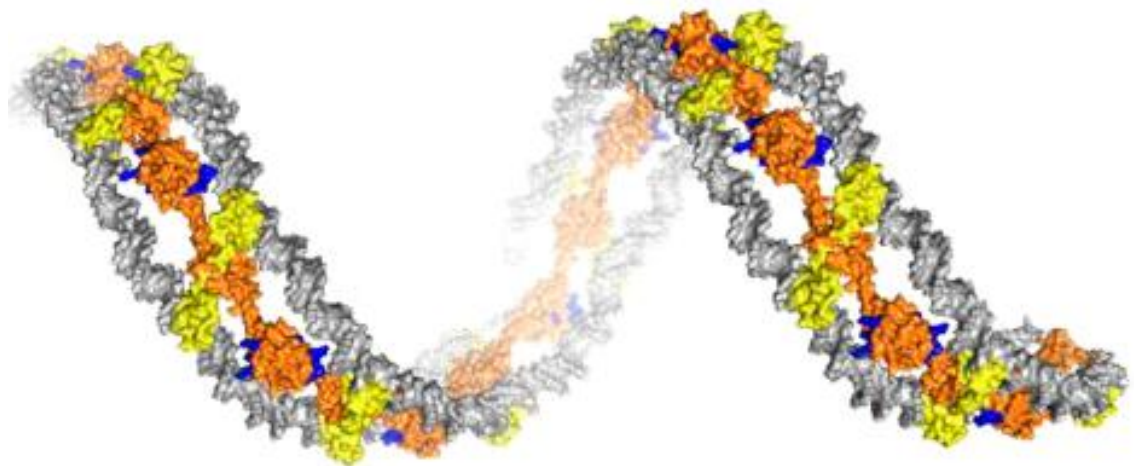


Figure 1.11 Stabilisation of DNA in a plectonemic (interwound) form by extended oligomers of H-NS. Orange: N-terminal oligomerisation domains; yellow: C-terminal DNA-binding domains; blue: stabilising basic residues; grey: DNA duplexes. Figure from Arold *et al.* (2010)

H-NS acts primarily as a repressor of gene expression (Lucchini *et al.*, 2006), with up to 5% of the *E. coli* genome (Muller *et al.*, 2006) and 10% of the *Salmonella* genome (Navarre *et al.*, 2006) being either up- or down-regulated in an *hns* mutant. H-NS has been called “the genome sentinel”

for its role in regulating genes acquired by horizontal gene transfer such as pathogenicity islands, particularly when those genes have a lower GC content than the rest of the genome (Lucchini *et al.*, 2006). The H-NS to DNA ratio can be deleteriously reduced by the addition of an overly large amount of foreign DNA into the genome, creating a mild version of the *hns* phenotype. The AT-rich plasmid pSf-R27 encodes its own H-NS homologue, Sfh, which compensates for this titration and aids transmission (Doyle *et al.*, 2007). An additional copy of an H-NS-like protein is also encoded in the major pathogenicity island of enteropathogenic *E. coli*, the locus of enterocyte effacement, which may perform the same function. Paradoxically, while H-NS is a repressor of gene expression its binding sites show a positive correlation with the binding sites of RNA polymerase (Grainger *et al.*, 2006). This is probably due to the “trapping” of polymerase on inactive promoters (section 1.1.6). H-NS regulates expression of many genes in response to environmental conditions (Atlung and Ingmer, 1997) and *hns* mutants exhibit diverse phenotypes including ectopic expression of virulence genes (Lucchini *et al.*, 2006), a higher proportion of anucleate daughter cells (Kaidow *et al.*, 1995), reduced cold-adaptation (Dersch *et al.*, 1994) and fewer copies of the chromosome per cell during exponential growth (Atlung and Hansen, 2002).

In addition to having enormous effects on the transcriptional programme, H-NS also has particularly strong roles in organising the nucleoid at the larger scale (see section 1.1), for example it may be a major component of topological domain boundaries (Noom *et al.*, 2007) and sequester inactive genes into two clusters per nucleoid (Wang *et al.*, 2011).

H-NS is able to form heterodimers with its paralogue StpA, which has a similar size but is far more basic (pI = 8.6). StpA has a major role in post-transcriptional regulation (discussed in section 1.3) and a *stpA* strain has a largely normal transcriptional programme (Sonden and Uhlin, 1996), suggesting that it is not a key NAP. ChIP-on-chip experiments showed that StpA is associated with the same regions of the chromosome as H-NS and that this association is dependent on the presence of H-NS (Uyar *et al.*, 2009). The function of H-NS is also known to be directly modulated by the two small (approximately 8 kDa), moderately basic co-regulators Hha and YmoA which structurally resemble and interact with the H-NS N-terminal dimerization domain (Garcia *et al.*, 2005). It has been suggested that H-NS by itself regulates expression of the core genome while H-NS in concert with Hha/YmoA regulates expression of the accessory genome, such as genes acquired by horizontal gene transfer (Banos *et al.*, 2009).

Many bacteria contain highly abundant DNA-bridging proteins which are functional equivalents, but not homologues, of H-NS (Dame *et al.*, 2005). These are often able to complement an *E. coli* or *Salmonella hns* mutant despite a lack of sequence similarity. The *Pseudomonas* MvaT proteins

(Tendeng *et al.*, 2003) and the *Bacillus* Rok protein (Smits and Grossman, 2010) are not similar to H-NS by sequence comparison but carry out the same functions and share a similar fold with *E. coli* H-NS. The DNA-bridging protein Lsr2 has been identified in *M. tuberculosis* and appears to represent a new class of H-NS-like NAPs. It forms large oligomeric complexes by non-sequence-specific association with AT-rich DNA (Colangeli *et al.*, 2007). It shields DNA against reactive oxygen species (Colangeli *et al.*, 2009) and is involved in regulating genes involved in response to antibiotics, including the activity of the promoter governing the drug-resistance operon *iniBAC* (Colangeli *et al.*, 2007). *Mycobacterium* also has a protein named HNS which has similar *in vitro* properties to *E. coli* H-NS and which can be complemented by *E. coli* H-NS despite lacking sequence similarity (Sharadamma *et al.*, 2010). *Streptomyces* species encode two paralogues of *lsr2* about which nothing has been published. SCO4076 (pI = 9.6) is much more basic than SCO3375 (pI = 6.5) but their respective roles are unknown.

1.2.3 Ligand-binding NAPs: Lrp and CRP

Some NAPs contain a DNA-binding domain (e.g. a helix-turn-helix) coupled to a ligand-binding domain which modulates their activity to reflect nutrient conditions. For example in *E. coli* Lrp (leucine-responsive regulatory protein) is responsive to amino acid supply while CRP (cAMP-responsive regulatory protein) is responsive to carbon supply. These NAPs are generally larger than HU or H-NS and resemble standard transcription factors which have been co-opted into a global role, making them difficult to predict from sequence data alone.

Lrp is a globally-acting member of the Lrp/AsnC family, colourfully known as the feast-or-famine regulatory proteins (FFRPs; for a review of this family see (Yokoyama *et al.*, 2006)). It is basic (pI = 9.1) and is one of the larger NAPs (19 kDa). It binds DNA via an N-terminal helix-turn-helix motif and forms dimers via its C-terminal domain (Leonard *et al.*, 2001). These dimers form more than one higher-order structure depending on ligand binding: in the absence of leucine Lrp self-assembles to form hexadecamers but in the presence of leucine these dissociate to form octamers (Chen and Calvo, 2002). Lrp is also responsive to a lesser degree to other amino acids such as alanine and methionine (Hart and Blumenthal, 2011). The Lrp octamer wraps an approximately 100 bp section of DNA with some degree of specificity: a consensus site of the sequence “(C/T)AG(A/C/T)A(A/T)ATT(A/T)T(A/G/T)CT(A/G)” is bound with high affinity while non-consensus sites are bound with a 20 - 70-fold lower affinity (Peterson *et al.*, 2007). Lrp is also known to competitively inhibit the action of Dam methyltransferase at promoters such as that of the *E. coli* *pap* operon where these proteins act on overlapping binding sites to regulate phase variation (Nou *et al.*, 1995). Lrp has been reported to affect the expression of around 10% of the

E. coli genome, either as a positive or a negative regulator (Newman and Lin, 1995; Tani *et al.*, 2002). Operons in its regulon include those encoding amino acid biosynthesis, nutrient transport, pili formation and many of the genes required upon entry to stationary phase (Ernsting *et al.*, 1992; Lin *et al.*, 1992). At some promoters its activity is enhanced by leucine binding (*fan*, *ilvIH* and *sdaC* operons) but it may also act independently of leucine binding (*fae* and *pap* operons).

Lrp/AsnC-family proteins are widely distributed amongst the eubacteria and homologues have also been found in the Archaea (Yokoyama *et al.*, 2006) but they are not all globally-acting. In addition to the NAP Lrp, *E. coli* has an asparagine-dependent paralogue of Lrp (AsnC) which is a local regulator and a third paralogue YbaO whose ligand and function are unknown (Yokoyama *et al.*, 2006) although it may be orthologous to the *Zymomonas* Grp protein which responds to glutamate (Peekhaus *et al.*, 1995). LrpA from *M. tuberculosis* is strongly upregulated in response to starvation (Betts *et al.*, 2002) and may have a role in persistence. The *S. coelicolor* genome contains around 15 genes encoding annotated as members of the Lrp/AsnC family based on sequence similarity but it is not known how many of these genuinely belong to the Lrp/AsnC family and which, if any, are global regulators.

CRP (cAMP receptor protein, also known as CAP) from *E. coli* is a basic protein (pI = 8.4) of larger size (23.7 kDa). It is sometimes described as a NAP due to its function as a global regulator, its ability to act as either an activator or a repressor and its formation of nucleoprotein complexes in combination with other NAPs. However its mode of action is more like a conventional transcription factor (i.e. by contacting one of the RNA polymerase subunits directly, rather than by altering promoter architecture) and it has a defined binding site sequence (Cameron and Redfield, 2008). CRP is activated by the allosteric effector cAMP, whereupon it binds DNA as a dimer and induces a 90° bend in the duplex (Gaston *et al.*, 1992). Its regulon includes around 200 operons in *E. coli*, including those encoding carbon transport and metabolism, amino acid biosynthesis, membrane protein and other transcriptional regulators (Zheng *et al.*, 2004). Its promoter is under stringent control and negatively autoregulates its own expression. The CRP/FNR major family of regulators has many members in almost every bacterial lineage (Korner *et al.*, 2003), with members capable of responding to such diverse environmental conditions as glucose levels, nitrogen availability and RNS stress. *S. coelicolor crp* (*sco3571*) is orthologous to *E. coli crp* and has been found to bind cAMP. Mutants are unable to respond to the rise in cAMP during early germination and are consequently defective in germination, as well as showing changes in growth, antibiotic production and sporulation (Derouaux *et al.*, 2004).

1.2.4 DNA-wrapping proteins: Fis

Fis (factor for inversion stimulation) is a small (11 kDa), basic (pI = 9.3) DNA-wrapping protein which was first characterised in *E. coli* and contains a helix-turn-helix domain but is not known to bind ligands. It binds as a homodimer in the minor groove of A- or AT-rich tracts and bends the DNA by 50 - 90°. Where several of these tracts are present in close proximity this results in co-operative binding with DNA being bent into a “microloop” which can bring RNA polymerase into contact with transcriptional activators or facilitate open complex formation and promoter clearance (Travers and Muskhelishvili, 1998). Alternatively Fis can stimulate transcription by directly contacting the σ and α subunits of RNAP (McLeod *et al.*, 1999) or occlude binding sites for proteins such as RNAP, for example at the P_{mom} promoter of phage Mu (Karambelkar *et al.*, 2012). Fis has an exceptionally large regulon. ChIP-chip showed it bound to almost 900 regions of the *E. coli* genome during exponential growth which was consistent with the finding that around 21% of *E. coli* genes were up- or down-regulated in a Δ *fis* mutant (Cho *et al.*, 2008). Fis targets include many genes involved in primary metabolism, translation and transport (Bradley *et al.*, 2007). Fis is the most abundant NAP during early exponential growth, with around 50,000 monomers per cell, however it is under control of the stringent response and its abundance falls sharply down to only a few hundred monomers during stationary phase (Ali Azam *et al.*, 1999). Mutants grow more slowly during exponential phase and often have species-specific phenotypes: Fis regulates the expression of virulence genes in the plant pathogen *Erwinia* (Lautier and Nasser, 2007), in enteropathogenic *E. coli* (Goldberg *et al.*, 2001) and in *Salmonella* (Wilson *et al.*, 2001) and it is involved in quorum sensing in *Vibrio cholerae* (Lenz and Bassler, 2007). It is involved in facilitating growth-phase-dependent changes in supercoiling of the chromosome, and particularly of the plasmids, by regulating expression of topoisomerases (Schneider *et al.*, 1999) and also by inhibiting their action (Schneider *et al.*, 1997; Travers *et al.*, 2001). Fis is highly conserved within the Enterobacteriaceae to the extent that the *E. coli* and *S. enterica* proteins are identical at the amino acid level and large parts of the promoter and upstream regions are identical at the nucleotide level (Osuna *et al.*, 1995). However species outside the Enterobacteriaceae (e.g. *B. subtilis*, *S. aureus*) lack the gene (Ohniwa *et al.*, 2011) and no homologues have been detected within the Actinomycetes.

1.2.5 DNA-protecting proteins: Dps and SASPs

Upon entry to stationary phase, during conditions of stress or during spore formation, the nucleoids of many bacteria become highly condensed by association with specialized NAPs as a way of physically protecting the genome from damage.

Dps (DNA-binding protein from starved cells) is an approximately 19 kDa doughnut-shaped protein in the same superfamily as ferritin (but lacking a heme group) which is expressed during stress or stationary phase in many bacterial lineages. In *E. coli* it is the most abundant protein during stationary phase (Azam and Ishihama, 1999). Dps protects DNA physically by forming dense nucleoprotein lattices and protects it from chemical damage by sequestering up to 500 insoluble Fe^{3+} ions in a cage-like cavity inside the dodecamer to prevent the formation of ROS and RNS that would be generated by the Fenton reaction if soluble Fe^{2+} was present in the cytoplasm. Dps proteins have a highly-conserved ferritin-like core which typically have basic N- and C-terminal extensions that bind DNA non-specifically to effect condensation (Ceci *et al.*, 2004). Dodecamers of Dps form ordered arrays (Frenkiel-Krispin *et al.*, 2004) with the N-terminal DNA-binding tails residing in the channels between dodecamers although in some cases smaller oligomers such as dimers (*Deinococcus*; (Grove and Wilkinson, 2005)) and trimers (*Mycobacterium*; (Gupta and Chatterji, 2003)) are formed which may bind DNA without causing hyper-condensation. While *E. coli* Dps is thought to simultaneously perform physical latticing, iron-binding and nucleoid compaction, other Dps homologues carry out only a subset of these functions. Those found in species such as *Agrobacterium tumefaciens*, *Helicobacter pylori* and *Bacillus anthracis* can protect DNA from chemical damage but are unable to bind it directly (Ceci *et al.*, 2003). *S. coelicolor* has three Dps homologues which are developmentally controlled and are required for efficient nucleoid compaction and hence segregation into spores (Facey *et al.*, 2009).

SASPs (small acid-soluble spore proteins) are small, slightly acidic (5 - 7 kDa, pI = around 4.5) proteins which coat the DNA of the endospores of *Bacillus* and *Clostridium* species, making up 6 - 15 % of the spore total protein. They are expressed in sufficient quantities to saturate the DNA-binding sites of the spore genome and protect it from heat, desiccation, oxidising agents and UV irradiation (Fairhead *et al.*, 1993). These proteins are relatively unstructured in solution but upon binding DNA form a structure dominated by α -helices. Binding also renders spore DNA more resistant to UV-induced damage by changing the conformation of DNA from the B form to the more UV-resistant A form (Mohr *et al.*, 1991). Binding is highly co-operative and the nucleoprotein filaments readily form side-by-side aggregates (Griffith *et al.*, 1994). Surprisingly, the HU homologue Mdp1 has also been described as having iron-detoxifying properties. Mdp1

was found to bind Fe^{3+} (but not Fe^{2+}) and a mutant strain was far more sensitive to H_2O_2 treatment in the presence of iron (Takatsuka *et al.*, 2011). It may compensate for the lack of Dps in *Mycobacterium*, which must have mechanisms to resist oxidative stress during host macrophage colonisation.

1.2.6 Less well-characterised NAPs

There are other DNA-binding proteins whose properties suggest they are NAPs, for example a large regulon and a pleiotropic mutant phenotype, but which have as yet been incompletely characterised. Many of these would have been difficult to identify by sequence searching because they are either of a novel class (e.g. mIHF/sIHF), because they are of a class which does not generally have NAP activity (e.g. *Mycobacterium* GroEL1), or because they belong to a known class of transcription factors and are only highly abundant under certain conditions (e.g. Dan/YgiP). Full validation will require experiments to measure their abundance, regulon size and effects on chromatin or nucleoid structure.

Dan (DNA-binding protein under anaerobic conditions, (also known as YgiP) is a LysR-family transcription factor from *E. coli* with a preference for AT-rich DNA. It is relatively scarce under standard lab conditions but is upregulated to a similar abundance as HU under anaerobic conditions. Genomic systematic enrichment of ligands by exponential enrichment (SELEX) identified 688 putative binding sites (with the short consensus sequence GTTNATT), predominantly in intergenic regions, and EMSA/DNase I footprinting revealed that these were often found in groups of closely-spaced sites which suggests a co-operative binding mode. DNA-Dan complexes visualised under AFM showed foci of binding at lower protein to DNA ratios and rigid filaments at saturating protein to DNA ratios (Teramoto *et al.*, 2010).

mIHF (mycobacterial integration host factor) is an essential gene which is required for site-specific integration of mycobacteriophage L5 and, despite the name, is not homologous to *E. coli* IHF. Like *E. coli* IHF, mIHF is most abundant at the end of exponential growth (Pedulla and Hatfull, 1998) and the *S. coelicolor* orthologue, sIHF, was found to bind the promoters controlling production of the two antibiotics undecylprodigiosin and actinorhodin (Park *et al.*, 2009). Null mutant strains were defective in sporulation, had reduced viability and showed upregulation of undecylprodigiosin and actinorhodin (Yang *et al.*, 2012).

Isolated *M. tuberculosis* nucleoids were found to contain large amounts of the protein chaperone GroEL1. Surprisingly, the protein was found to bind multiple types of ssDNA and the minor groove

of dsDNA with nanomolar affinity. *In vitro* experiments showed that it was able to protect DNA from degradation by DNase I or hydroxyl radicals and that it was also able to compact DNA into globular structures. Other GroEL homologues did not share these properties so this may represent a lineage-specific co-option of an existing protein to a nucleoid-associated role (Basu *et al.*, 2009).

EspR is a helix-turn-helix regulator of the ESX-1 secretion system in *Mycobacterium*. ChIP-seq experiments estimated that it had at least 165 binding sites throughout the genome, with the majority of targets being involved in cell wall synthesis and maintenance in addition to regulating its own transcription. The estimated concentration was high throughout the growth cycle and rose from approximately 20,000 monomers during early vegetative growth to approximately 100,000 monomers during stationary phase (Blasco *et al.*, 2012).

1.3 Global RNA chaperones

In addition to the DNA-binding NAPs, which regulate gene expression at the level of transcription, there are also globally-acting proteins which control gene expression at the post-transcriptional level. These operate by regulating the stability and translational capacity of mRNAs, often by mediating interactions between mRNAs and their cognate regulatory small non-coding RNAs (sRNAs). There is a degree of overlap between DNA-binding and RNA-binding global regulators, for example StpA and Hfq are both primarily RNA chaperones but are also able to interact with specific sequences in genomic DNA (Muller *et al.*, 2006; Updegrove *et al.*, 2010), although the consequences of this are not understood. Global RNA chaperones have several important characteristics in common with NAPs: very high abundance, generally low sequence specificity (preference for AT content or structural features), large regulon, modulation of substrate structure, co-operative binding and small size. They govern the global transcriptional programme by interacting with myriad transcripts, with the determinants of binding being encoded in the RNAs themselves.

As a flexible single-stranded nucleic acid, RNA is far more able than DNA to form complex secondary structures via intramolecular base-pairing. These structures are crucial for their function and interactions with other RNA species or proteins. Three examples of sRNA secondary structure are shown in Figure 1.12, illustrating the assembly of sRNAs into multiple stem-loop structures. Chaperones are required by some RNAs to reach their native structure in a process akin to chaperone-assisted protein folding, whereby the molecule is guided to its correct, functional conformation by disruption of the incorrect intermediates which represent kinetic traps (Cristofari and Darlix, 2002). This type of chaperone activity is common to many proteins, including around half the ribosomal proteins. Other types of RNA chaperone activity include strand displacement, where a base-paired region of RNA is destabilised so that an alternate conformation can be adopted, or strand annealing (“matchmaking”), where the chaperone facilitates base pairing between two separate RNAs. Strand annealing is achieved by “RNA crowding”, whereby a protein binds an mRNA and its corresponding regulatory sRNA simultaneously, increasing their effective local concentrations and making it more likely that they will interact by base pairing than if they were free in solution (Rajkowitsch and Schroeder, 2007).

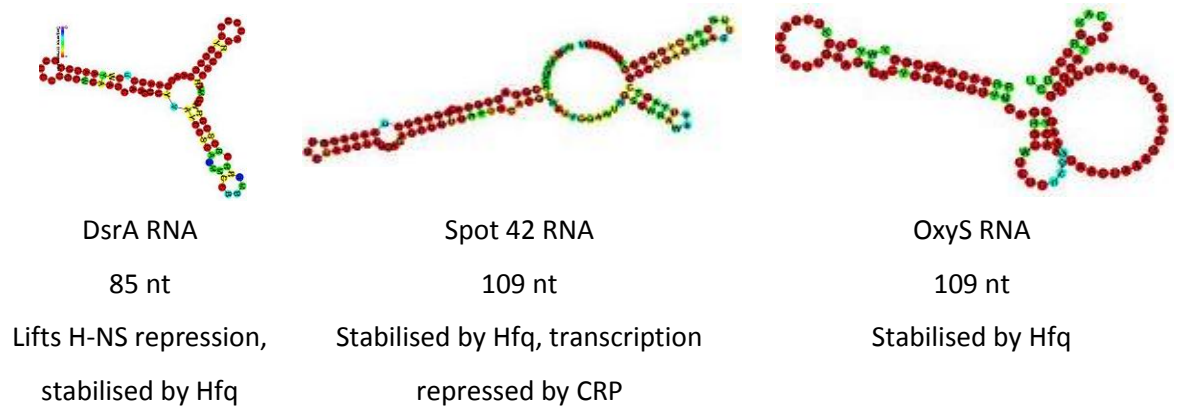


Figure 1.12 Structures of three representative *E. coli* sRNAs. Colour indicates degree of sequence conservation between species: red indicates strong conservation; yellow and green indicate moderate conservation; blue indicates weak conservation. Figures from the Rfam database (<http://rfam.sanger.ac.uk/>).

1.3.1 Small non-coding RNAs (sRNAs)

Many sRNAs have been identified in bacteria. More than 80 have been found in *E. coli* (Raghavan *et al.*, 2011; Vogel *et al.*, 2003) and in various species they regulate processes such as primary and secondary metabolism (Sonnleitner and Haas, 2011) and stress responses (Battesti *et al.*, 2011; Gottesman *et al.*, 2006). They are described as acting *in cis* when they are transcribed from their target gene and contain a large region of sequence antisense to the cognate mRNA, or *in trans* when they are transcribed from an intergenic region and have a short (10 – 25 nt) region of incomplete complementarity with the mRNA (Storz *et al.*, 2005). The latter group require an adaptor protein to support the weak interaction. Mechanisms of sRNA regulation include stabilisation of mRNA, destabilisation of mRNA (targeting them for cleavage by RNases such as RNaseE) and ribosomal binding site occlusion. A given sRNA may bind multiple targets and a given mRNA may be targeted by multiple sRNAs, allowing the formation of complex regulatory networks. Novel sRNAs can be discovered in a number of ways, for example as signals in transcriptomics data (Vockenhuber *et al.*, 2011), by screening cDNA libraries prepared from LMW RNA (Arnvig and Young, 2009), or by co-IP using antibodies against a global RNA chaperone such as Hfq (Zhang *et al.*, 2003), with the RNA species being identified by hybridising them to microarrays or by sequencing cDNA libraries prepared from them. Computational methods can detect canonical sRNAs simply by locating conserved intergenic regions containing a predicted Rho-independent terminator which is a feature characteristic of sRNAs (Panek *et al.*, 2008). Many studies employ several discovery approaches and then confirm the presence of putative sRNAs by Northern blotting of total RNA.

RNA-binding proteins are more difficult to investigate because RNA is less stable than DNA, however some methods for identifying the targets of RNA-binding proteins are available. Co-IP has been used to identify novel sRNAs which interact with Hfq in *E. coli* (Sittka *et al.*, 2007; Zhang *et al.*, 2003) and *Listeria* (Christiansen *et al.*, 2004). An extension of this method, individual-nucleotide resolution cross-linking and immunoprecipitation (iCLIP), uses UV light to induce covalent bonds between RNAs and their bound proteins. Following immunoprecipitation and proteinase treatment, one or several amino acids remains attached to the RNA so that the cDNAs produced from it are truncated at that site, revealing the binding site to high resolution (reviewed in König *et al.*, 2012).

1.3.2 Global RNA chaperones

Hfq is the most well-known global RNA chaperone and was originally discovered in *E. coli* as a host-derived component of the phage Q β replicase. It forms a cyclical homo-hexameric ring with two distinct RNA-binding faces whose structures in complex with DNA have been determined by X-ray crystallography. The distal face has a preference for triplet repeats of the sequence A-R-N (Link *et al.*, 2009) while the proximal face has a preference for U-rich sequences (Schumacher *et al.*, 2002) such as the 6 - 8 nt poly(U) tract resulting from Rho-independent termination which is often seen on sRNAs (Otaka *et al.*, 2011; Sauer and Weichenrieder, 2011). In both cases the RNA is in contact with each of the six monomers in the ring and the interaction is stabilized by co-operative binding. The C-terminal portion of the protein is not resolved in crystal structures and is likely to be disordered. It may act as a “fly-casting” net to pull in RNA (Wright and Dyson, 2009).

Hfq can stabilise sRNAs (e.g. DsrA, Spot42, RhyB) and/or mediate their annealing to cognate mRNAs, for example Hfq is required to mediate the inhibitory interactions between the sRNA OxyS and its targets *rpoS* and *fhlA* (Zhang *et al.*, 1998). This interaction may mask or unmask the mRNA’s ribosomal binding site, or it may make the mRNA a substrate for degradation by RNaseE, possibly by a direct Hfq-RNaseE interaction (Morita *et al.*, 2005). For a recent review of Hfq targets see Vogel and Luisi (2011). Hfq has additional roles in mRNA turnover which do not involve *trans*-acting sRNAs, for example in the general polyadenylation pathway where it is required for full activity of *E. coli* poly(A) polymerase I, which adds poly(A) tails to the 3’ ends of transcripts containing a Rho-independent terminator (Mohanty *et al.*, 2004). It may also have a role in transcriptional termination (Rabhi *et al.*, 2011). There is also some evidence that Hfq also binds DNA: double-stranded genomic DNA fragments found to be associated with Hfq during

protein purification were sequenced and found to represent curved DNA. EMSA experiments showed that Hfq bound several of these sequences with a $K(d)$ of around 400 nM via its C-terminal domain, although the significance of this is unknown (Updegrave *et al.*, 2010). Hfq is extremely important in the response to several stress conditions in Enterobacteria and is a regulator of antibiotic production in *Pseudomonas* (Wang *et al.*, 2012). However in other species the effect of deleting Hfq may have a lesser effect (e.g. *Burkholderia*; (Sousa *et al.*, 2010)) or no discernible effect (*S. aureus*; (Bohn *et al.*, 2007)), or Hfq may be absent entirely (e.g. ϵ -proteobacteria, high GC Gram positives, *Deinococcus*, *Streptococcus* etc.; (Sun *et al.*, 2002)). Homologues have recently been identified with extremely low sequence conservation in lineages where Hfq had previously been considered absent. Cyanobacterial Hfq has an unusual Sm2 motif but is still able to form the Hfq fold (Boggild *et al.*, 2009) and in the Archaeal species *Methanococcus jannaschii* a genuine Hfq (rather than the expected eukaryotic-like Sm protein) was identified that was able to complement an *E. coli* Δhfq mutant (Nielsen *et al.*, 2007).

StpA is a paralogue of H-NS found within the Enterobacteria. It is able to bind DNA and form heterodimers with H-NS but also associated with RNA. It can bind two RNA molecules simultaneously and possesses both RNA annealing and strand displacement activities (Rajkowitsch and Schroeder, 2007). Mayer *et al.* (2007) investigated the sequence preferences of StpA using a genomic selection (SELEX) experiment and detected a preference for unstructured RNA, but specific targets could not be identified so its specificity could be very low. The *E. coli* $\Delta stpA$ mutant has a very mild phenotype, with no discernible effect on global patterns of gene expression (Muller *et al.*, 2006).

CsrA (carbon storage regulator) is a small and basic KH domain protein (6.9 kDa, pI = 8.1) which forms symmetrical homodimers that bind in the 5'UTR of mRNA transcripts to occlude the ribosomal binding site. Its action can be counteracted by a large sRNA (CsrB; 350 nt) which sequesters multiple dimers of CsrA. It is not as truly global as Hfq, but acts alongside it to pleiotropically regulate important functions such as carbon storage, cell surface properties and virulence (Romeo, 1998). A highly conserved homologue from *Erwinia* species, RsmA (regulator of secondary metabolism), acts in a similar fashion and is known to regulate antibiotic production, as does Hfq in *Pseudomonas* (Wang *et al.*, 2012).

1.3.3 sRNA-mediated regulation in Actinobacteria and *S. coelicolor*

Trans-encoded (i.e. adaptor-protein-requiring) sRNAs are found in all lineages, including the Actinobacteria. At least 34 sRNAs are known in *Mycobacterium*, of which several are associated with virulence (DiChiara *et al.*, 2010). Swiercz *et al.* (2008) identified nine sRNAs in *S. coelicolor*, of which eight were *trans*-encoded and one was *cis*-encoded. Vockenhuber *et al.* (2011) identified another 63 by deep sequencing, of which 11 were confirmed by Northern blotting. Many of these were medium- or growth-phase-dependent, although their functions have not yet been determined.

The *hfq* gene is conspicuously absent from Actinobacterial genomes, based on searches using amino acid sequences (Sun *et al.*, 2002) and predicted secondary structures (Swiercz *et al.*, 2008). However, biochemical techniques such as co-sedimentation with nucleoids and pull-down using RNA aptamers have not yet been attempted in this group.

Hfq homologues have been identified in some Gram positive lineages though it is possible these have a less vital role than in the Enterobacteria: some strains of *S. aureus* was not found to have a Δhfq mutant phenotype distinguishable from wild-type (Bohn *et al.*, 2007; Liu *et al.*, 2010b), implying that other machinery may be responsible for handling *trans*-acting sRNAs. It is not known whether another protein has taken over this function or whether *trans*-acting sRNAs are simply not used in the same fashion. Given the discovery of Hfq proteins with highly divergent sequences (Boggild *et al.*, 2009; Nielsen *et al.*, 2007; Ramos *et al.*, 2011) and the independent evolution of non-homologous H-NS-like proteins (Dorman *et al.*, 1999), it is tempting to speculate that at least a functional equivalent of Hfq remains to be found. One of the expected side-benefits of the proteomic survey described in chapter 3 was that it would either generate candidate Hfq equivalents or that it would support the hypothesis that this function truly is absent in this species, as Hfq is a major component of isolated *E. coli* nucleoids.

1.4 Post-translational modification of DNA-binding proteins

The properties of bacterial chromatin are determined in part by the relative proportions of the different nucleoid proteins which it contains. Changing these proportions gives a diversity of chromatin types suited to different purposes, for example in *E. coli* there is more HU during vegetative growth and more IHF and Dps in stationary phase (section 1.2). Another device commonly used to introduce variety to the chromatin of eukaryotes and Archaea is post-translational modification: the covalent addition of chemical groups to the side chains of amino

acids. Acetylation, methylation and phosphorylation are the most widely-studied modifications but many other chemical groups such as ADP-ribosyl or short peptide tags (ubiquitin or SUMO) can be added to eukaryotic chromatin proteins (Figure 1.13, right), often in combination (Jenuwein and Allis, 2001). These may alter the function of a protein by changing its stability or conformation, which in turn modulates the protein's interactions with DNA or other proteins.

Post-translational modifications (PTMs) are a common method of protein regulation in many processes in every kingdom of life (Khouri *et al.*, 2011). For example the activity of bacterial enzymes such as acetyl-CoA synthetase can be regulated by reversible acetylation (Starai *et al.*, 2002) while the action of response regulators (transcriptional regulators which are part of two-component signalling systems) are often regulated by reversible phosphorylation (Stock *et al.*, 1989). The *Yersinia* virulence factor YopJ is even able to acetylate the proteins of its host (Mukherjee *et al.*, 2007). The role of PTMs in controlling chromatin function has been intensively studied in eukaryotes, notably in the context of the basic N-terminal and C-terminal tails found on histones which protrude from the nucleosome. Modifications of these are strongly correlated with changes in transcriptional activity, for example acetylation is associated with active, open chromatin while methylation is associated with silent, tightly-packaged chromatin (Jenuwein and Allis, 2001). The modifications can work directly by affecting formation of higher-order structures (Figure 1.13, left) and also indirectly by controlling interactions with proteins such as chromatin remodelling complexes which recognise modified histones such as bromodomains and chromodomains (de la Cruz *et al.*, 2005).

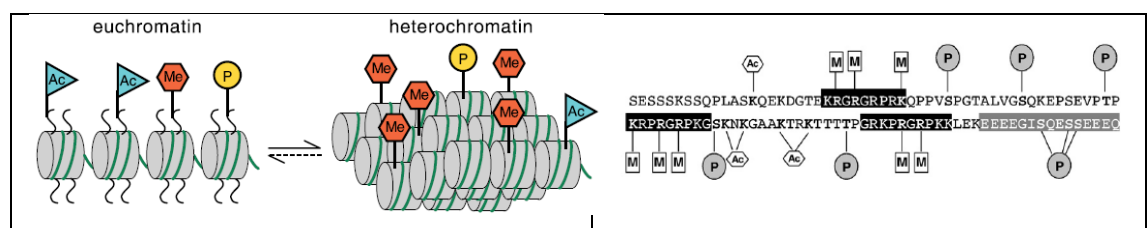


Figure 1.13 Post-translational modification of eukaryotic chromatin proteins. Ac = acetylation; M = methylation; P = phosphorylation. Left: association of chromatin modifications on histone tails with changes in higher-order compaction. Grey cylinders represent nucleosomes, DNA is shown in green. Figure from Jenuwein and Allis (2001). Right: the multiplicity of modifications found on the eukaryotic non-histone chromatin protein HMGA1a. Figure from Zhang and Wang (2010).

While research on histone tails has provided a conceptual and methodological toolkit with which to study PTMs, it is unclear to what degree these observations will apply to bacterial gene regulation as these structures (and the higher-order complexes they form) are absent from the Archaea and eubacteria. It is therefore beneficial to also consider non-histone chromatin proteins

which may be more relevant to the function of bacterial chromatin. The HMG proteins of eukaryotes are functionally equivalent to the HU proteins of bacteria in that they bend DNA sharply and are involved in DNA transactions such as transcription and repair following UV irradiation (Oberto *et al.*, 1994). The HMG-box region of the mitochondrial protein Glom is even able to complement an HU mutant in *E. coli* despite being evolutionarily unrelated and having a different protein fold (Sasaki *et al.*, 2003). Like histones, HMG proteins undergo acetylation, phosphorylation and methylation on multiple residues (Figure 1.13, right) which influence the strength of interactions with DNA and other proteins as well as influencing their subcellular localisation (Zhang and Wang, 2010). While Archaeal histone homologues lack modifiable tails, two other unrelated chromatin proteins in this kingdom are known to be reversibly modified: Marsh *et al.* (2005) showed that the chromatin protein Alba (Sso10b) from *Sulfolobus solfataricus* is reversibly acetylated on a DNA-contacting lysine residue, which lowers its affinity for DNA or dsRNA by 2-fold; Sul7d chromatin proteins from several species in the order Sulfolobales are heterogeneously methylated on several lysine residues in response to heat shock, although this does not affect their affinity for DNA (Baumann *et al.*, 1994).

Modifications are added and removed in a regulated fashion by enzymes. For example response regulators are phosphorylated by their cognate sensor kinases (Stock *et al.*, 1989) while Alba is acetylated by the histone acetyltransferase (HAT) ssPat and deacetylated by the histone deacetylase (HDAC) ssSir2 (Marsh *et al.*, 2005), which are homologues of the HAT/HDAC pair that reversibly acetylate acetyl-CoA synthetase in *Salmonella* (Starai *et al.*, 2002) and are similar to the enzymes that act on eukaryotic histones and HMG proteins.

1.4.1 Acetylation and ADP-ribosylation

Reversible (ϵ -amino) acetylation is a moderately labile mark which occurs on lysine or arginine residues, neutralizing the terminal positive charge. It is transferred to the protein from a donor molecule such as acetyl-CoA by an acetyltransferase and removed by a deacetylase. This is distinct from N-terminal acetylation which is seen on some proteins involved in translation (Ramagopal and Subramanian, 1974; Yoshikawa *et al.*, 1987).

Yu *et al.* (2008) showed that at least 125 residues on 85 proteins from many different functional classes were acetylated in *E. coli*, of which around 70% were specific to stationary phase. More than half of the acetylated proteins were involved in protein synthesis and carbohydrate

metabolism (e.g. tricarboxylic acid cycle) and several global regulators of transcription such as RNA polymerase, sigma factor, Rho factor and topoisomerase I were also found to be acetylated. Within 2 hours of transfer to fresh medium all but two of the stationary-phase-related acetylation marks disappeared, implying that this method of regulation allows a rapid response to changing environmental conditions. The two most well-characterised examples of protein acetylation in bacteria are the chemotaxis regulator CheY in *E. coli* (Barak and Eisenbach, 2001) and acetyl-CoA synthetase in *Salmonella* (Starai *et al.*, 2002). These are both reversibly acetylated on a lysine residue via the HAT/HDAC enzyme pair Pat/CobB. In *E. coli* the Pat/CobB enzyme pair also regulates reversible acetylation of four transcription factors including the helix-turn-helix domain of the global regulator RcsB which may alter its DNA-binding ability (Thao *et al.*, 2010). While little is currently known about the global levels of acetylation in *Streptomyces*, these species are known to possess several highly-conserved HDACs which are likely to act on proteins: a single class I HDAC (Gregorette *et al.*, 2004) and two class III HDACs/sirtuins (Frye, 2000), one of which was shown *in vitro* to deacetylate both BSA and *Streptomyces*-encoded Acetyl-CoA synthetase in a manner that could be inhibited by nicotinamide, which is a known inhibitor of eukaryotic sirtuins (Mikulik *et al.*, 2012). No candidate for the HAT enzyme has been identified in *Streptomyces* as HATs are not well conserved and their substrates cannot be predicted from sequence information alone. In addition to their deacetylase activities, some sirtuins are also able to ADP-ribosylate proteins on arginine, glutamic acid or aspartic acid residues. This modification was observed on two proteins of unknown function which appeared at the start of stationary phase in *Bacillus subtilis*. Prevention of ADP-ribosylation using the inhibitor 3-methoxybenzamide strongly inhibited sporulation and antibiotic production (Huh *et al.*, 1996). In a parallel mechanism, ADP-ribosylation of four proteins was shown to be important for differentiation in *S. coelicolor* (Shima *et al.*, 1996). Also the α subunit of *E. coli* RNA polymerase can be ADP-ribosylated by two genes products of the T4 phage, although this is not essential for infection (Goff and Setzer, 1980).

1.4.2 Methylation

Lysine and arginine residues can be methylated by the transfer of a methyl from a donor such as S-adenosyl-L-methionine (AdoMet) by a methyltransferase. Each hydrogen on the amine group of a lysine side chain can be replaced with a methyl group, i.e. each residue can be in an unmethylated, monomethylated, dimethylated or trimethylated form. Arginine can be methylated either once or twice on its terminal guanidyl group.

Fewer examples of protein methylation are known than protein acetylation or phosphorylation in prokaryotes, however this could be because methylation has been studied less intensively in this kingdom. Several of the 50S ribosomal proteins are methylated on lysine residues (Chang and Chang, 1975; Chang *et al.*, 1974). The positions of these marks are well conserved between the distantly-related species *E. coli* and *B. subtilis* but the same pattern is not seen in the Archaeal species *Halobacterium cutirubrum* (Amaro and Jerez, 1984). The outer membrane protein B (OmpB) from virulent strains of *Rickettsia prowazekii* is heavily mono-, di- and tri-methylated on multiple lysine residues while OmpB from attenuated strains is hypomethylated (Chao *et al.*, 2004), suggesting that methylation plays a role in progression of infection.

Methylation of residues other than lysine and arginine also occurs: the N-terminal methionine residues of the chemotaxis protein CheZ and ribosomal protein L16 of *E. coli* were found to be monomethylated (Stock *et al.*, 1987) and the outer membrane protein L32 (OmpL32) of *Leptospira interrogans* is methylated on several glutamic acid residues, which may contribute to the ability of the bacteria to evade the host immune system (Eshghi *et al.*, 2012).

Some but not all of the enzymes involved are known, for example ribosomal protein L11 is modified encoded by the methyltransferase PrmA in *E. coli* (Vanet *et al.*, 1994) and *Thermus thermophilus* (Cameron *et al.*, 2004), although this function is dispensable for growth.

The *Mycobacterium* HU homologue Mdp1 is heterogeneously mono- and di-methylated on around a dozen lysine residues by an unknown methyltransferase (Petthe *et al.*, 2002; Soares de Lima *et al.*, 2005). This modification confers resistance to proteolysis (Petthe *et al.*, 2002) but it is not known whether this modification is reversible or whether it confers altered DNA-binding properties. In contrast to the well-defined modification sites of eukaryotic histones, this protein contained a highly mixed population of different combinations of methylation that seem unlikely to form a coherent “code”. Protein methylation has not yet been reported in *Streptomyces* species.

1.4.3 Phosphorylation

Serine, threonine, histidine or tyrosine residues can be phosphorylated by the transfer of a phosphate group from donor molecules such as ATP or from another phosphorylated protein by a protein kinase. This is used extensively in two-component systems (TCSs) which allow bacteria to respond rapidly to environmental changes. In these simplest of these systems a membrane-bound histidine kinase (HK) undergoes histidine autophosphorylation in response to a stimulus, then the phosphate is transferred to an aspartic acid on the N-terminal receiver domain of a cognate response regulator (RR). The RR then undergoes a conformational change which changes the action of the C-terminal effector domain, resulting in a change in transcription. The phosphate is eventually removed by the intrinsic autophosphatase activity of the RR (Gao *et al.*, 2007). Several response regulators govern key steps in *Streptomyces* development and/or antibiotic production, such as BldM (Molle and Buttner, 2000), AfsR (Sawai *et al.*, 2004), AbsA2 (McKenzie and Nodwell, 2007) and RamR (Nguyen *et al.*, 2002). Phosphorylation is also used to control the action of sigma factors and anti-sigma factors (Osterberg *et al.*, 2011; Park *et al.*, 2008). A systematic study of phosphorylation sites in *S. coelicolor* revealed Lsr2 (SCO3375) to be phosphorylated on one threonine but the function of this is unknown (Parker *et al.*, 2010).

1.4.4 Chemical epigenetics

Chemical epigenetics is the use of small-molecule inhibitors to negate the activity of enzymes removing or adding PTMs, resulting in an increase or decrease respectively in the level of that modification. For example acetylation levels can be increased by inhibition of HDACs with chemicals such as sodium butyrate (Davie, 2003) or decreased by inhibition of HATs with anacardic acid (Balasubramanyam *et al.*, 2003; Sun *et al.*, 2006). As well as having several therapeutic uses, particularly in treating cancer (Blumenschein *et al.*, 2008; Lane and Chabner, 2009), this process can also be used to control gene expression without the need for genetic engineering which could be useful when screening strain collections whose members have not been sequenced and which are not genetically tractable. Williams *et al.* (2008) found that the secondary metabolic repertoire of several filamentous fungi could be de-repressed by treating cells with HDAC inhibitors. If this process could be applied to the transcriptionally-silent “cryptic” biosynthetic clusters encoded by Actinomycetes (section 1.5) then it would facilitate natural product discovery. Several HDAC inhibitors have already been shown to increase production of known metabolites and increase transcription from cryptic pathways in *S. coelicolor* and other Actinomycetes (Moore *et al.*, 2013), although the targets of the HDACs are not yet known.

1.5 *Streptomyces coelicolor* A3(2) as a model organism for nucleoid biology

The organisation (section 1.1.7) and protein composition (section 1.2) of the *Streptomyces* nucleoid has been studied far less extensively than that of *E. coli* or *Salmonella*. There is a modest body of information about the biochemical and phenotypic effects of some NAPs in *Streptomyces* (and the related species *Mycobacterium*) but these have not been linked to structural effects on the nucleoid, for example the bend angle of HupS/Mdp1 is unknown and the numbers of macrodomains or topological domains have never been estimated. It would be beneficial to have a well-developed model of chromatin structure and function within the Actinobacteria as this group contains a great number of commercially and clinically important species to which the knowledge could be applied. Many members of the subgroup Actinomycetales produce vast numbers of natural products, from which 2/3 of current clinically-used antibiotics are derived (Kieser *et al.*, 2000) and from which we expect to harvest many more novel antibiotics, antifungals, antiparasitics and anticancer agents. The phylum Actinobacteria also contains the pathogens *Mycobacterium* and *Corynebacterium* as well as to the human gut endosymbionts *Bifidobacterium*. The species selected for this study, *Streptomyces coelicolor*, is commonly used as a model for the Actinomycetes. It is easily culturable, non-pathogenic, genetically tractable and has a well-annotated genome sequence complemented with abundant transcriptomic and proteomic datasets. Much of the existing nucleoid-related research in *Streptomyces* has been in this species, such as work on HupS (Salerno *et al.*, 2009) and nucleoid segregation during development (Hopwood and Glauert, 1960a, b). Also it produces two pigmented antibiotics, actinorhodin and undecylprodigiosin, which are an easily-assayable barometer of secondary metabolism.

1.5.1 *Streptomyces* life cycle

Streptomyces species have a complex life cycle, undergoing physiological and morphological differentiation into multiple cell types (Figure 1.14, above). For a review see Claessen *et al.* (2006). On agar media, germ tubes emerge from the spores and then grow by tip extension with occasional branching to form the vegetative mycelium which undergoes assimilative growth (Jyothikumar *et al.*, 2008). Vegetative hyphae are made up of multinucleate compartments with occasional septa. After around 24 hours aerial hyphae are erected which grow upwards in an unbranched coil, requiring the action of hydrophobic agents such as SapA (Tillotson *et al.*, 1998) and chaplins (Claessen *et al.*, 2003) on their surfaces to overcome the surface tension at the interface between the aqueous colony surface and the air. There is a round of programmed cell death in the vegetative mycelium, possibly to provide the aerial mycelium with nourishment, followed by an upregulation of secondary metabolic genes. The aerial hyphae then undergo a round of synchronous septation to form small compartments containing single genomes which subsequently become ovoid spores (Hopwood and Glauert, 1960b). Finally, the spore walls thicken and turn dark grey. Mature spores contain a single compacted copy of the genome and are resistant to stresses such as heat, UV radiation and detergent. A number of mutations have been discovered that arrest the developmental program at particular stages: the *bld* mutants are unable to form aerial hyphae, so appear flat and bald, while the *whi* mutants are unable to produce spore pigment, so appear white (Figure 1.14, below). These phenotypes may be medium-specific, for example the morphological phenotypes of most of the *bld* mutants are carbon-source dependent (Pope *et al.*, 1996).

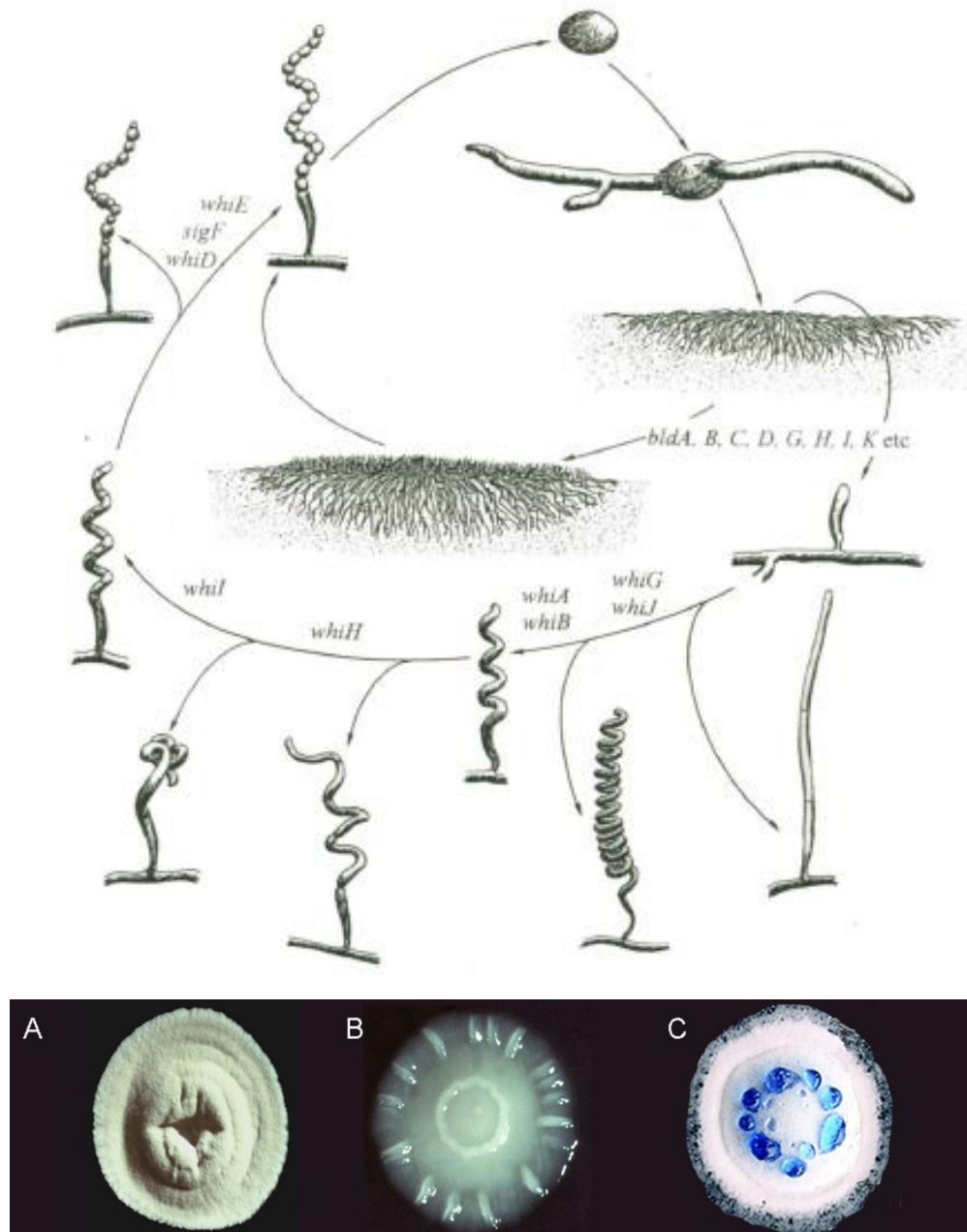


Figure 1.14 Above: developmental cycle of *Streptomyces* (Kieser *et al.*, 2000). Below: wild-type *S. coelicolor* (A) and the phenotypic consequences of blocking aerial mycelium formation (B; *bld* mutant) and mature spore formation (C; *whi* mutant). Picture by Matt Hutchings.

In rich liquid media *S. coelicolor* does not sporulate but it does produce proteins and secondary metabolites typical of later stages of development (Manteca *et al.*, 2010), most notably the pigmented compounds undecylprodigiosin and actinorhodin. Undecylprodigiosin (sometimes referred to simply as “red”) is a tripyrrole antibiotic which acts as a shunt for excess proline during transition and stationary phases (White and Bibb, 1997). Actinorhodin is an intracellular quinone antibiotic which is a pH indicator, appearing blue following ammonia fuming but often red-purple at lower pH (Kieser *et al.*, 2000), while the related compound γ -actinorhodin is secreted into the

medium (Bystrykh *et al.*, 1996). Both forms of actinorhodin are produced during stationary phase in response to nutrient limitation (nitrogen/phosphate) and are subject to repression by glucose (Kim *et al.*, 2001; Melzoch *et al.*, 1997). These compounds are a useful system for studying the effects of chromatin on secondary metabolism because they are easily assayed, because they appear sequentially (undecylprodigiosin is produced before actinorhodin) and because their production is dependent on growth conditions, so that increased production may be detected clearly. However actinorhodin is often produced as part of the general stress response so care must be taken when interpreting results.

1.5.2 Natural product diversity of the Actinomycetes

The Actinomycetes are a particularly valuable source of natural products, with >5% of an average genome being devoted to secondary metabolism (Udwary *et al.*, 2007). There is great deal of useful chemical diversity, with the genus *Streptomyces* alone yielding (among others) the compounds chloramphenicol, neomycin, daptomycin, erythromycin, tetracycline and vancomycin. They also produce antimycotics (e.g. Nystatin from *Streptomyces noursei*), antiparasitics (e.g. Ivermectin from *Streptomyces avermitilis*), immunosuppressants (e.g. FK-506 from *Streptomyces tsukubaensis*), herbicides (e.g. bialaphos from *Streptomyces viridochromogenes*) and antineoplastics (e.g. Trichostatin A from *Streptomyces sp.*). It is thought that many of these compounds are used to increase access to nutrients by inhibiting other soil-dwelling bacteria or as signalling molecules to co-ordinate behaviour (Yim *et al.*, 2007). They are also a source of novel cellulases (Ghangas and Wilson, 1987; Li, 1997), which break down cellulose by hydrolysis and are widely used in the food, paper, ethanol (biofuel) and textile industries.

While the increasingly high rediscovery rate of antibiotics and the rise of combinatorial chemistry (Michels *et al.*, 1998) have curtailed the efforts put into screening libraries of isolates by large pharmaceutical companies, recent evidence suggests that there is still a large amount of useful chemical diversity waiting to be explored. Genome sequencing has revealed that even the well-studied laboratory model strains of *Streptomyces* encode many more products than expected: *S. coelicolor* alone has 26 predicted natural product clusters (Bentley *et al.*, 2002) while *S. avermitilis* has 25 clusters (Omura *et al.*, 2001). It has been estimated that the genus *Streptomyces* alone has around 150,000 antimicrobials, of which around 3% have so far been identified (Watve *et al.*, 2001). The more recently sequenced marine species *Salinispora tropica* has 17 novel clusters in addition to several known clusters, representing 10% of its genome (Udwary *et al.*, 2007). The majority of these are concentrated at the opposite end of the chromosome to the origin of

replication, in genomic islands which are likely to have arisen by horizontal gene transfer (Penn *et al.*, 2009). The existence of many secondary metabolic clusters within a single organism is known as “OSMAC”: one strain, many compounds. Products have not been observed for a large proportion of these clusters under standard laboratory conditions, leading them to be known as “cryptic” clusters. Activation of some of these can be achieved by labour-intensive methods such as extensive medium optimisation (Kim *et al.*, 2001) or genetic deletion of a cluster-encoded transcriptional repressor (Horinouchi *et al.*, 1989). However these methods are not necessarily suitable for large-scale screening programmes. Two other ways of increasing the yield of commercially-relevant drugs are to use libraries of isolates from exotic environments or to widen the scope of screening to encompass more applications - the discovery of immunosuppressants and anticancer agents has been relatively recent, suggesting that there may be many more medical, industrial or nutraceutical compounds left to discover.

Chapter 2 Materials and methods

2.1 Bacterial strains and plasmids

Table 2.1 *E. coli* strains used in this study.

Strain	Antibiotic resistance	Genotype	Reference / Source
DH5α	None	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac</i>	(Sambrook et al., 2001)
BL21(DE3)	Cm	F- <i>ompT gal dcm lon hsdSB (rB- mB-) lacIq</i> lambda <i>lacUV5-T7 gene 1 ind1 sam7 nin5</i>	(Studier and Moffatt, 1986)
ET12567	Cm, Tet	<i>dam13::Tn9 dcm6 hsdM hsdR recF143 zjj201::Tn10 galK galT22 ara14 lacY1 xyl5 leuB6 thil tonA31 rpl136 hisG4 tsx78 mtli glnV44 F-</i>	(MacNeil et al., 1992)
BW25113	None	(<i>araD-araB</i>)567 <i>lacZ4787(::rrnB-4) lacIp-4000(lacI_q) -rpoS369(Am) rph-1 (rhaD-rhaB)</i> 568 <i>hsdR514</i>	(Datsenko and Wanner, 2000)

Table 2.2 *S. coelicolor* strains used or constructed in this study.

Strain no.	Genotype	Reference / Source
<i>S. coelicolor</i> A3(2) M145 used to construct derivatives		
M145	SCP1 ⁻ SCP2 ⁻	(Chakraborty and Bibb, 1997)
B1000	$\Delta hupS::(oriT-apra)$	Dr. Michael McArthur
B1001	$\Delta hupA::(oriT-apra)$	This work
B1002	B1001 harbouring pMS82-HupA	This work
B1003	$\Delta SCO3375::(oriT-apra)$	This work
B1004	B1003 harbouring pMS82-3375	This work
B1005	M145 harbouring pIJ10257-3375	This work
B1006	$\Delta SCO4076::(oriT-apra)$	This work
B1007	M145 harbouring pIJ10257-4076	This work
B1008	B1006 harbouring pMS82-4076	This work

Strain no.	Genotype	Reference / Source
B1009	Δ SCO5592::(<i>oriT-apra</i>)	This work
B1010	pMS82 integrated at ϕ BT1 <i>attP</i> site	This work
B1011	pIJ10257 integrated at ϕ BT1 <i>attP</i> site	This work

Table 2.3 Plasmids used or constructed in this study.

Plasmid	Vector	Insert	Use	Antibiotic resistance	Reference / Source
pSupercos1	Supercos1	None	Cosmid backbone pUC ori cos	Carb, Kan	Stratagene
pIJ790	pSC101	pKD20 with <i>bla</i> replaced by <i>cat</i>	[<i>oriR101</i>] [<i>repA101(ts)</i>] <i>araBp-gam-be-exo</i> Arabinose-inducible λ red recombinase on TS replicon	Cm	(Gust <i>et al.</i> , 2003)
pIJ773	pBluescript	<i>aac(3)/IV</i> (Apra ^R) + <i>oriT</i>	Apramycin resistance gene flanked by FLP recognition sites	Apra, Carb	(Gust <i>et al.</i> , 2003)
pUZ8002	RK2	None	RK2 derivative with defective <i>oriT</i>	Kan	(Paget <i>et al.</i> , 1999)
pGEM	pGEM	None	TA-cloning vector	Carb	Promega
pLysS	pLysS	None	Expression of T7 lysozyme	Cm	(Moffatt and Studier, 1987)
pET15b	pET15b	None	Expression vector with N-terminal 6xHis tag	Carb	Novagen
pHis5592	pET15b	SCO5592	pET15b derivative used to express N-terminally His-tagged fusion of SCO5592	Carb	This work
pHis5556	pET15b	SCO5556	pET15b derivative used to express N-terminally His-tagged fusion of SCO5556	Carb	This work
pWKS30	pSC101	2.3 kb fragment containing Amp ^R and <i>lacZα</i>	Low-copy-number expression vector for use in <i>S. enterica</i>	Carb	(Wang and Kushner, 1991)
pMS82	pMS82	None	Integrative <i>Streptomyces</i> expression vector for complementation	Hyg	(Gregory <i>et al.</i> , 2003)

Plasmid	Vector	Insert	Use	Antibiotic resistance	Reference / Source
pIJ10257	pMS82	330 bp <i>ermEp*</i> (<i>KpnI-PstI</i>) with RBS and MCS from pIJ8723	Integrative <i>Streptomyces</i> expression vector for strong, constitutive expression	Hyg	(Hong <i>et al.</i> , 2005)
Gene deletion constructs					
pIJ11400	Supercos1	StE59	Construction of $\Delta hupA$	Carb, Kan, Apra	This work
pIJ11401	Supercos1	StE126	Construction of $\Delta 3375$	Carb, Kan, Apra	This work
pIJ11402	Supercos1	StD25	Construction of $\Delta 4076$	Carb, Kan, Apra	This work
pIJ11403	Supercos1	St7A1	Construction of $\Delta 5592$	Carb, Kan, Apra	This work
Complementation constructs					
pIJ11404	pMS82	P ₂₉₅₀ -SCO2950	pMS82 derivative used to complement $\Delta hupA$	Hyg	This work
pIJ11405	pMS82	P ₃₃₇₅ -SCO3375	pMS82 derivative used to complement $\Delta 3375$	Hyg	This work
pIJ11406	pMS82	P ₄₀₇₆ -SCO4076	pMS82 derivative used to complement $\Delta 4076$	Hyg	This work
Over-expression constructs					
pIJ11407	pIJ10257	SCO3375	pIJ10257 derivative used to over-express SCO3375 under the control of <i>ermEp*</i>	Hyg	This work
pIJ11408	pIJ10257	SCO4076	pIJ10257 derivative used to over-express SCO4076 under the control of <i>ermEp*</i>	Hyg	This work

2.2 Culture media and antibiotics

2.2.1 Antibiotics

Antibiotics were stored at -20°C as 10x concentrated stocks.

Table 2.4 Antibiotics used in this study.

Antibiotic	Concentration in media (µg/ml)	Solvent	Supplier
Apramycin (Apra)	50	Water	Sigma-Aldrich
Carbenicillin (Carb)	100	Water	Formedium
Chloramphenicol (Cm)	25	Ethanol	Sigma-Aldrich
Hygromycin B (Hyg)	40	Water	Invitrogen
Kanamycin (Kan)	50 (<i>E. coli</i>) 10 (<i>S. coelicolor</i>)	Water	Formedium
Nalidixic acid (Nal)	20	Water	Sigma-Aldrich

2.2.2 Agar media

Unless otherwise stated, the media used for the culturing of *S. coelicolor* and *E. coli* were prepared as previously described (Kieser *et al.*, 2000; Sambrook *et al.*, 2001).

Table 2.5 Agar media used in this study.

Medium	Composition	Instructions for preparation
L-agar	Agar (10 g) Difco Bacto tryptone (10 g) NaCl (5 g) Glucose (1 g) Distilled water (to 1,000 ml)	Tryptone, NaCl and glucose were dissolved in the distilled water and 200 ml aliquots were dispensed into 250 ml Erlenmeyer flasks containing 2 g agar. The flasks were then closed and autoclaved.
Difco nutrient agar (DNA)	Difco nutrient agar (4.6 g) Distilled water (to 200 ml)	Difco Nutrient Agar was placed in a 250 ml Erlenmeyer flask and the distilled water was added to it. The flasks were then closed and autoclaved.
Soya flour and mannitol medium (SFM)	Agar (20 g) Mannitol (20 g) Soya flour (20 g) Tap water (to 1,000 ml)	Mannitol was dissolved in the water, then 200 ml aliquots were dispensed into 250 ml Erlenmeyer flasks which each contained 2 g agar and 2 g soya flour. The flasks were autoclaved twice at 115°C for 15 min, with gentle shaking between runs.

Medium	Composition	Instructions for preparation
R5 +	Sucrose (103 g) K ₂ SO ₄ (0.25 g) MgCl ₂ ·6H ₂ O (10.12 g) Glucose (10 g) Difco Casaminoacids (0.1 g) Trace elements solution (2 ml) Difco yeast extract (5 g) TES buffer (5.73 g) Distilled water (to 1,000 ml) <hr/> Added at the time of use: 0.5 % KH ₂ PO ₄ (1 ml) 5 M CaCl ₂ ·2H ₂ O (0.4 ml) 20 % L-proline (1.5 ml) 1 M NaOH (0.7 ml)	The ingredients were dissolved in the distilled water, then 100 ml aliquots were dispensed into 250 ml Erlenmeyer flasks which each contained 2.2 g Difco Bacto agar. The flasks were closed and autoclaved. At the time of use, the medium was re-melted and the remaining autoclaved solutions were added to each flask in the order listed.

2.2.3 Liquid media

Unless otherwise stated, the media used for the culturing of *S. coelicolor* and *E. coli* were prepared as previously described (Kieser *et al.*, 2000; Sambrook *et al.*, 2001).

Table 2.6 Liquid media used in this study.

Medium	Composition	Instructions for preparation
L broth	Difco Bacto tryptone (10 g) Difco yeast extract (5 g) NaCl (5 g) Glucose (1 g) Distilled water (to 1,000 ml)	Ingredients were dissolved in the distilled water, then aliquots were dispensed into universals or flasks and autoclaved.
2 x YT	Difco Bacto tryptone (16 g) Difco yeast extract (10 g) NaCl (5 g) Distilled water (to 1,000 ml)	Ingredients were dissolved in the distilled water, then 10 ml aliquots were dispensed into universals or flasks and autoclaved.
DSGM	Difco yeast extract (1 g) Difco Casaminoacids (1 g) Distilled water (to 100 ml) <hr/> Added at the time of use: 5 M CaCl ₂ (200 µl)	Ingredients were dissolved in the distilled water, then 10 ml aliquots were dispensed into universals and autoclaved.
TSB	Difco nutrient broth powder (8 g) Distilled water (to 1,000 ml)	Difco nutrient broth powder was placed in a 250 ml Erlenmeyer flask and the distilled water was added to it. The flasks were closed and autoclaved.

Medium	Composition	Instructions for preparation
YEME34%	Difco yeast extract (3 g) Difco Bacto-peptone (5 g) Oxoid malt extract (3 g) Glucose (10 g) Sucrose (340 g) Distilled water (to 1,000 ml)	The ingredients were combined in the distilled water, then 200 ml aliquots were dispensed into 250 ml Erlenmeyer flasks. The flasks were closed and autoclaved.
50/50 TSB:YEME34%	TSB (25 ml) YEME34% (25 ml)	TSB and YEME34% were autoclaved separately then mixed together in equal volumes upon use.
SOB (minus Mg)	Tryptone (20 g) Yeast extract (5 g) NaCl (0.5 g) Distilled water (to 950 ml)	Ingredients were dissolved in the distilled water, then 10 ml of 250 mM KCl was added and the pH was adjusted to pH 7.0 with 5 M NaOH. The volume was made up to 1,000 ml with deionised water then aliquots were dispensed into universals and autoclaved.
SOC	SOB (1,000 ml) Added at the time of use: 1 M glucose (20 ml) 2 M MgCl ₂ (5 ml)	SOB medium was prepared, then glucose and MgCl ₂ were added following autoclaving.
P buffer	Sucrose (103 g) K ₂ SO ₄ (0.25 g) MgCl ₂ .6H ₂ O (2.02 g) Trace element solution (2 ml) Distilled water (to 800 ml) Added at the time of use: KH ₂ PO ₄ (1 ml) CaCl ₂ .2H ₂ O (10 ml) TES buffer (10 ml)	Ingredients were dissolved in the distilled water then dispensed into 80 ml aliquots and autoclaved. KH ₂ PO ₄ , CaCl ₂ .2H ₂ O and TES buffer were added following autoclaving.

2.3 Growth and storage of bacteria

2.3.1 Growth and storage of *E. coli*

E. coli strains were typically grown in L broth with shaking at 250 rpm or on L agar, with antibiotic selection as appropriate. Most strains were grown at 37°C but BW25113/pIJ790 was grown at 30°C to maintain the plasmid which contains a temperature-sensitive origin of replication. For long-term storage, an equal volume of culture grown in L broth and 50% sterile glycerol (f.c. 25%) were mixed by pipetting then stored at -80°C.

2.3.2 Growth and storage of *Salmonella*

Salmonella strains were typically grown in L broth with shaking at 250 rpm or on L agar, with antibiotic selection as appropriate. For long-term storage, an equal volume of culture grown in L broth and 50% sterile glycerol (f.c. 25%) were mixed by pipetting then stored at -80°C. For growth curves, strains were grown overnight with antibiotics as appropriate. These cultures were used to inoculate 10 ml aliquots of L containing the same antibiotics which were grown for 2 – 3 hours until OD₆₀₀ had reached approximately 0.4. Cells were washed three times in 10 ml L to remove traces of antibiotics then resuspended to an OD₆₀₀ of 0.10 - 0.15 in L with 0.1 mM IPTG. 200 µl culture was added to each well of a sterile 96-well plate (Corning Costar), with six wells per strain and six blank wells of L only. OD₆₀₀ was measured every 30 minutes using an Eon microplate reader (BioTek), with shaking between readings, over the course of 16 hours until the cultures had reached stationary phase.

2.3.3 Storage of *Streptomyces*

Streptomyces strains were grown and manipulated as described in (Kieser *et al.* 2000). Strains were stored as spores in 10% glycerol at -20°C (short-term) or at -80°C (long-term). Spores were prepared by growing strains on SFM at 30°C for up to 7 days until the surface turned dark grey. Spores were collected in around 5 ml ddH₂O per plate using a sterile cotton wool pad through which spores were filtered by suction with a 1 ml pipette tip. They were then centrifuged at 13,000 rpm in a microcentrifuge and the pellet was resuspended in approximately 10 volumes sterile 10% glycerol (e.g. 100 µl per 10 mg pelleted spores). Spore stocks were titred by making serial dilutions in water and plating on SFM, then counting the number of colonies seen after 3 days. Typically three plates each of four dilutions were used. Most strains gave a highly consistent titre of around 1×10^{10} CFUs per ml.

2.3.4 Synchronous germination of *S. coelicolor* spores

25 µl dense spore suspension was centrifuged at 13,000 rpm in a micro centrifuge and the pellet was resuspended in 5 ml TES buffer, to remove storage glycerol. Spores were heat shocked at 50°C for 10 minutes in a water bath then cooled under cold running water. 5 ml DSGM was added and the spores were incubated at 37°C with shaking for 2 – 3 hours. Spores were centrifuged again at 13,000 rpm then resuspended in 0.5 ml TES and vortexed briefly. They were used to inoculate typically a 50 ml volume of medium to an OD₄₅₀ of 0.05.

2.3.5 *S. coelicolor* growth curves

The growth rate and antibiotic production of *S. coelicolor* was assessed in the rich liquid medium 50/50 TSB:YEME34%. Generally 25 µl of a dense spore suspension was used to inoculate 50 ml 50/50 TSB:YEME34% in a 250 ml flask containing a baffle, resulting in a starting OD₄₅₀ of 0.05. Four flasks were used per genotype or condition. The flasks were incubated at 30°C with shaking at 225 rpm. Growth was measured by reading the OD₄₅₀ of well-mixed culture on a Thermo Spectronic BioMate3 spectrophotometer, using 50/50 TSB:YEME34% as a blank.

Undecylprodigiosin concentration was measured by resuspending 500 µl culture in an equal volume of acidified methanol and centrifuging again briefly (to pellet the debris), then measuring the absorbance of the supernatant at 530 nm against a blank of acidified methanol. Total actinorhodin concentration was measured by mixing 660 µl culture with 330 µl 2 M KOH, vortexing for 10 seconds, then centrifuging briefly to pellet the debris. The absorbance of the resulting supernatant was measured at 640 nm against a blank of 660 µl TSB:YEME34% mixed with 330 µl 2 M KOH. For all three measurements, an appropriate dilution factor was used when OD₄₅₀>1.0 as the spectrophotometer is inaccurate outside this range.

2.3.6 Preparation of *S. coelicolor* protoplasts

25 ml YEME medium supplemented with 0.5% glycine and 5 mM MgCl₂ was inoculated with spores to an OD₄₅₀ of 0.05 and grown at 30°C for 36 – 40 hours. Mycelium was harvested by centrifugation at 4,000 rpm then washed twice in 15 ml sterile 10.3% sucrose to remove traces of media. The pellet could be frozen at -20°C at this stage. Mycelium was resuspended in 4 ml sterile P buffer containing 1 mg/ml lysozyme and incubated at 30°C for 15 – 60 minutes. Cells were drawn in and out of a 5 ml pipette three times then incubated for 15 minutes. 5 ml more P buffer was added and again the cells were drawn in and out of a 5 ml pipette three times then incubated for 15 minutes. Protoplasts were filtered through sterile cotton well then centrifuged at 1,000 rpm in a micro centrifuge for 7 minutes and finally resuspended in 1 ml P buffer.

2.4 Buffers and solutions

All solutions used for protein purification or as additives to sterile media were filtered through a membrane of pore size 0.22 µm immediately prior to use. The pH was adjusted using either 5 M HCl or 5 M NaOH at the temperature to be used.

Table 2.7 Buffers and solutions used in this study.

Solution / buffer	Composition	
Cosmid isolation solution 1	50 mM Tris/HCl, pH 8	50 mM
	EDTA	10 mM
Cosmid isolation solution 2	NaOH	200 mM
	SDS	1 %
Cosmid isolation solution 3	Potassium acetate, pH 5.5	3 M
EB (from Qiagen kit)	Tris-HCl, pH 8.5	10 mM
SET buffer (for gDNA isolation)	NaCl	75 mM
	Tris-HCl, pH 8	20 mM
	EDTA	25 mM
TBE buffer	Tris base	89 mM
	Boric acid	89 mM
	EDTA	2 mM
TES buffer	Tris-HCl, pH 8	10 mM
	EDTA	1 mM
	NaCl	1 M
Laemmli SDS loading buffer (2X)	Tris-HCl, pH 8	63 mM
	Glycerol	10 %
	SDS	2 %
	β-mercaptoethanol	2.5 % (v/v)
	Bromophenol blue	0.0025 % (w/v)
DNA loading buffer	Glycerol	10 %
	Bromophenol blue	0.125% (w/v)
	Xylene cyanol	0.125% (w/v)
TX buffer	Tris-HCl, pH 8	50 mM
	Triton X-100	0.001% (w/v)
His buffer A (low-imidazole)	Tris-HCl, pH 8	25 mM
	NaCl	500 mM
	Glycerol	2.5 %
	Imidazole	10 mM

His buffer B (high imidazole)	Tris-HCl, pH 8	25 mM
	NaCl	500 mM
	Glycerol	2.5%
	Imidazole	500 mM
Glutaraldehyde cross-linking buffer	HEPES, pH 7.5	20 mM
	NaCl	150 mM
Recombinant protein storage buffer	Tris-HCl, pH 8	10 mM
	NaCl	50 mM
	Glycerol	10 %
	EDTA	2 mM
Magnesium DNase I digestion buffer (recipe provided with Roche DNase I)	Tris-HCl, pH 8	40 mM
	NaCl	10 mM
	MgCl ₂	6 mM
	CaCl ₂	1 mM
Manganese DNase I digestion buffer N.B. This buffer must be kept in the dark to prevent precipitation of manganese.	Tris-HCl, pH 8	40 mM
	NaCl	10 mM
	MnCl ₂	6 mM
	CaCl ₂	1 mM
10x TBE buffer	Tris-borate (pH 8.0)	450 mM
	EDTA disodium salt	10 mM
Nucleoid prep digestion buffer	Tris –HCl, pH 8	10 mM
	Sucrose	20 %
	NaCl	100 mM
Drlica lysis buffer	NaCl	2 M
	EDTA	10 mM
	Sodium deoxycholate	0.4 %
	Brij-58	1 %
Sarfert lysis buffer	EDTA	10 mM
	Sodium deoxycholate	0.4 %
	Brij-58	1 %
	Spermidine	10 mM
Novagen 1x thrombin cleavage buffer	Tris –HCl, pH 8.4	20 mM
	NaCl	150 mM
	CaCl ₂	2.5 mM

2.5 General molecular biology methods

2.5.1 Isolation of plasmid DNA

The Qiagen miniprep kit, which works according to the principle of alkaline lysis, was used according to the manufacturer's instructions. Typically 10 ml of an overnight LB culture of cells harbouring the plasmid of interest was centrifuged at 4,000 rpm for 10 min. The pellet was resuspended in buffer P1 which contained 50 mM Tris-Cl pH 8.0, 10 mM EDTA and 100 µg/ml RNase A, then lysed by addition of buffer P2 which contained 1% SDS and 200 mM NaOH. After 5 min the lysate was neutralised with buffer N3, which contained 3 M potassium acetate pH 5.5, and centrifuged at 13,000 rpm in a microcentrifuge to remove debris. The supernatant was passed down a column containing a silica membrane by centrifugation. Plasmid DNA bound to the membrane and was washed with buffer PE containing ethanol and a high concentration of salt before being eluted using 30 - 50 µl of elution buffer and stored at -20°C.

2.5.2 Isolation of cosmid DNA

Cosmid isolation from *E. coli* was carried out by alkaline lysis as described by Sambrook *et al.* (2001). The cell pellet from 1.5 ml of culture was resuspended by vortexing in 100 µl solution I (50 mM Tris/HCl, pH 8; 10 mM EDTA). 200 µl solution II (200 mM NaOH; 1 % SDS) was added and the tube was inverted ten times. 150 µl solution III (3 M potassium acetate, pH 5.5) was then added and mixed in by inverting the tube five times. The tube was then centrifuged at 13,000 rpm in a microcentrifuge for 5 min at room temperature. The supernatant was mixed with 400 µl phenol-chloroform (buffered to pH 8.0), vortexed briefly to mix and then centrifuged at 13,000 rpm for 5 min. The upper phase was then transferred to a clean 1.5 ml tube. 600 µl of ice cold isopropanol was added and DNA precipitation was allowed to occur at room temperature for 10 min followed by centrifuging at 13,000 rpm for 5 min. The pellet was washed with 200 µl of 70 % ethanol and centrifuged again at 13,000 rpm. The supernatant was removed and the pellet was air-dried for 5 min at room temperature prior to resuspension in 50 µl of 10 mM Tris/HCl (pH 8).

2.5.3 Isolation of HMW genomic DNA

25 µl of *S. coelicolor* spore stock was used to inoculate 50 ml TSB:YEME34% and grown for three days with shaking at 30°C. The mycelium was washed twice with 25 ml of 10% sucrose to remove traces of media. 0.3 g wet cells were resuspended in 10 ml SET buffer (75 mM NaCl; 25 mM EDTA; 20 mM Tris/HCl, pH 7.5). 200 µl of SET buffer containing 50 mg/ml lysozyme was added and the mixture was incubated at 37°C for 30 - 60 min. 600 µl of 20% SDS and 560 µl of SET buffer containing 10 mg/ml proteinase K were added and the mixture was incubated at 55°C for at least

2 hours, with occasional inversion. 4 ml 5 M NaCl was added and the tube was gently inverted three times. Once the mixture had cooled, 10 ml phenol-chloroform (buffered to pH 8.0) was added and the tube was gently rocked for 30 min. The tube was centrifuged at 4,000 rpm for 15 min and the supernatant was removed to a clean tube using a cut-off pipette tip. A second wash using 10 ml of chloroform alone was performed to remove residual phenol. DNA was then precipitated by adding 0.6 volumes ice-cold isopropanol and leaving to stand at room temperature for 10 min. The tube was centrifuged at 4,000 rpm for 15 min and the pellet was washed in 70% ethanol before being resuspended in an appropriate volume of ddH₂O. RNA was removed by dissolving the HMW DNA in Promega RNase ONE™ reaction buffer (1x) and digesting with 2 µl Promega RNase ONE™ for 2 hours at 37°C. The phenol-chloroform wash, chloroform-only wash, isopropanol precipitation and 70% ethanol wash were then repeated as above. HMW DNA was finally resuspended in Qiagen elution buffer at a concentration of approximately 5 mg/ml and stored at -20°C. Before use, the DNA was diluted in Qiagen elution buffer to a lower concentration (typically 50-500 ng/µl) and allowed to redissolve fully at 37°C for 4 hours.

2.5.4 Agarose gel electrophoresis

Agarose gels were prepared and run in 1% TBE buffer at 100 V unless otherwise specified. 0.7% gels were used for HMW DNA, 1% gels were for general use and 2% gels were used for fragments below 500 bp. Size was indicated using 100 bp DNA ladder (0.1–1.2 kb), 1 kb DNA ladder (0.5-10 kb) or 2-log DNA ladder (0.1-10.0 kb), all from New England Biolabs.

2.5.5 Extraction of DNA fragments from gel slices

Bands were excised from agarose gels using a clean razor blade, taking as little excess gel as possible. DNA was extracted using the Qiaquick™ gel extraction kit (Qiagen) according to the manufacturer's instructions. Briefly, the gel slice was dissolved fully in buffer QG at neutral pH under high salt conditions, then bound to a silica membrane mounted in a column. This was washed with buffer QC (to remove traces of agarose) then washed with buffer PE containing ethanol and eluted in 30-50 µl of Qiagen elution buffer or ddH₂O.

2.5.6 Purification of DNA following PCR or restriction digest

Contaminants such as restriction enzymes or unincorporated dNTPs were removed from DNA using the Qiaquick™ gel extraction kit (Qiagen) according to the manufacturer's instructions. The sample was diluted in buffer QG at neutral pH under high salt conditions, then bound to a silica membrane mounted in a column. This was washed with buffer PE containing ethanol and eluted in 30-50 µl of Qiagen elution buffer or ddH₂O.

2.5.7 Measurement of DNA concentration

DNA concentration and purity were estimated using a Nanodrop ND-1000 on the nucleic acids setting, however this was only used as a rough guide as the Nanodrop is known to be sensitive to the presence of contaminants such as residual phenol or proteins. The concentrations of vectors and inserts for cloning were confirmed by running an aliquot on an agarose gel alongside a known quantity of ladder DNA and comparing the bands' intensities.

2.5.8 Digestion of DNA using restriction endonucleases

Enzymatic digests were carried out according to the manufacturer's instructions. Generally, 5-10 µg plasmid or cosmid DNA were digested with 1.5 µl enzyme in a 50 µl reaction for two hours at the appropriate temperature. Where a double digest was performed, a buffer was selected for which the activity of both enzymes was not less than 75% and a trial digest of each enzyme alone was carried out in 10 µl reactions in the same buffer.

2.5.9 Ligation

Following restriction digest, inserts and vectors were gel extracted and eluted in 30 µl ddH₂O in order to maximise concentration. 200 ng total DNA was dissolved in 9 µl T4 DNA ligase buffer (Promega), to which was added 1 µl T4 DNA ligase (Promega). Ligations were carried out overnight at 4°C then 3 µl was used to transform cells and the remaining 7 µl was resolved on a 1% agarose gel to confirm that ligation had occurred.

2.5.10 Transformation of chemically competent *E. coli*

For cloning, chemically competent DH5α cells were purchased from Life Technologies Ltd (Invitrogen Division). For transformation, 50-200 ng DNA and 100 µl competent cells were mixed in a pre-chilled 2 ml microcentrifuge tube then incubated on ice for 20 min. Cells were heat shocked at 42°C for 45 seconds then rescued by adding 750 µl SOC and incubating in a 2ml microcentrifuge tube with shaking at 37°C for one hour. Two different volumes (typically 20 µl and 730 µl) were then spread onto agar plates containing the appropriate antibiotics.

2.5.11 Preparation and transformation of electro-competent *E. coli*

BW25113/pIJ790 and ET12567/pUZ8002 cells were grown and made electro-competent on site. A single colony was used to inoculate 10 ml LB and grown overnight with shaking at 37°C. 100 µl of this preculture was inoculated into 10 ml SOB and grown at 30°C and 37°C respectively for 3-4 h with shaking at 200 rpm to an OD₆₀₀ of approximately 0.6. The cells were recovered by centrifugation at 3,000 rpm for 5 min at 4°C. The pellet was washed twice in ice-cold 10% glycerol

and finally resuspended in a small volume of 10% glycerol (typically 100-200 μ l) to give a viscous suspension. For transformation, 50-200 ng DNA was mixed with 50 μ l competent cells in a pre-chilled 0.2 cm electroporation cuvette (Cell Projects). Electroporation was performed using a GenePulser II (Bio-Rad) set to 200 Ω , 25 μ F and 2.5 kV. The time constant was typically 4.6-4.8 ms. Immediately following electroporation, 750 μ l ice-cold SOC was added and the cells were transferred to a 2 ml microcentrifuge tube. They were then incubated with shaking at 37°C for one hour. Two different volumes (typically 20 μ l and 730 μ l) were then spread onto agar plates containing the appropriate antibiotics.

2.6 PCR

Polymerase chain reaction (PCR) was typically carried out using a Bio-Rad Engine Thermal Cycler with a heated lid. Strips of 8 thin-walled 0.2 ml PCR tubes (Thermo Scientific) were used to ensure correct heat transfer from the block. 5% DMSO was included in all reactions to discourage the formation of secondary structures which can occur with DNA of high GC-content. Polymerase was added last to the reaction mix. For general PCR and colony PCR, Taq DNA polymerase from New England Biolabs was used. This is not a proofreading Taq but genes encoding NAPs are generally very small (<500 bp) so point mutations were rare and all constructs were verified by Sanger sequencing. Extension times were generally 60 seconds per kb of expected product, plus 10-60 seconds. The same reaction recipe was used for general and colony PCR. Typically a 50 μ l reaction volume was used for general PCR, a 25 μ l reaction volume was used for colony PCR and a 10 μ l reaction volume was used for Sanger sequencing PCR.

2.6.1 General PCR

Component	Stock concentration	Volume added (μ l)	Final concentration
Taq polymerase	5 U/ μ l	0.25	1.25 U
Taq buffer	10x	5	1x
DMSO	100 %	2.5	5 %
dNTPs	40 mM (10 mM each)	1.25	
Primer mix	50 μ M (25 μ M each)	1	1 μ M
Template DNA	25 ng (plasmid or cosmid) 50 ng (genomic DNA)	Variable	0.5 ng/ μ l 1 ng/ μ l
ddH₂O		to 50 μ l	

95°C	1 min	}	x 25 cycles
95°C	45 s		
58-60°C	30 s		
72°C	45 – 150 s		
72°C	10 min		
10°C	indefinitely		

2.6.2 Colony PCR in *E. coli*

Instead of using DNA in solution as a template, the template was provided by scraping a portion of a single colony to the PCR mix using a sterile pipette tip, which was then used to streak the remaining cells onto a fresh L agar plate containing the appropriate antibiotics.

95°C	10 min	}	x 30 cycles
95°C	45 s		
56-58°C	30 s		
72°C	45 – 150 s		
72°C	10 min		
10°C	indefinitely		

2.6.3 Colony PCR in *Streptomyces*

Streptomyces single colonies were patched on to DNA media containing 1% glucose and the relevant antibiotics, then incubated at 30°C for 24-48 h until dense mycelium had grown. Around 100 mg of this was mixed with 75 µl DMSO in a 1.5 ml microcentrifuge tube. Tubes were shaken vigorously at room temperature for 1 hour, then frozen at -80°C and thawed at room temperature to encourage lysis. 2.5 µl of the supernatant was used as template, replacing the 2.5 µl DMSO included in the recipe.

2.6.4 REDIRECT® PCR

All gene replacement cassettes used for PCR-targeted deletion were amplified from pIJ773 using the relevant REDIRECT primers listed in section 2.14. PCR was carried out using the same recipe as in section 2.6.1 but with the following modified program:

95°C	2 min		
95°C	45 s	}	x 10 cycles
50°C	45 s		
72°C	90 s		
95°C	45 s	}	x 15 cycles
55°C	45 s		
72°C	90 s		
72°C	10 min		
10°C	indefinitely		

Cassettes were separated from template plasmid by extraction from a 1% TBE agarose gel.

2.6.5 Sanger sequencing using Big Dye v3.1

All plasmid constructs were confirmed by Sanger sequencing before use. Primers were designed to anneal to the vector backbone either 50 bp upstream (forwards) or 50 bp downstream (reverse) of the MCS so that the region sequenced spanned the entire insert and at least 20 nt of vector sequence on either side. BigDye-labelled PCR product was amplified from typically 50 ng of purified plasmid using the ABI BigDye® 3.1 dye-terminator reaction mix (Applied Biosystems) according to the manufacturer's instructions. Reactions were subsequently submitted to the John Innes Genome Centre Sequencing Service for ABI Sanger Sequencing. The resulting sequence chromatogram files were analysed using BioEdit (available from <http://www.mbio.ncsu.edu/bioedit/bioedit.html>).

Component	Volume (µl)
ABI BigDye® 3.1 dye-terminator reaction mix	1
5x BigDye® 3.1 Reaction Buffer	1.5
Primer (3.2 µM)	1
Template	Variable
ddH ₂ O	to 10

96°C	1 min	}	x 25 cycles
96°C	10 s		
50°C	5 s		
60°C	4 min		

2.7 Quantitative PCR (qPCR)

The qPCR experiments in this thesis were designed to conform to MIQE guidelines. The thermocycler used was an MJ Mini Personal Thermocycler (Bio-Rad, CA, USA) and the software used was Opticon 3 (Bio-Rad). SYBRgreen assays were designed for use with the SYBRGreenER supermix kit (Molecular Probes, Inc) in a reaction volume of 25 µl in an MJ White 96-well plate (Bio-Rad).

2.7.1 Primer design

DNase-seq primers were designed against the promoter regions of two genes (*afsk* and *sti1*) while *HaeIII*-seq primers were designed against the coding regions of four genes (*afsk*, *hrdB*, *actII-ORF4* and *sti1*), with each amplicon spanning two *HaeIII* restriction sites (which were > 15 bp from each other and >3 bp from the ends of the primer annealing sequence). Primers were designed according to the following parameters:

- Amplicon sizes were 126 – 147 bp for *HaeIII* and 118 – 125 bp for DNase.
- Predicted melting temperature of 58-62°C, with each pair having a difference of < 2°C.
- Primers were 17-21 nt in length
- Some primers had a higher GC content than generally recommended (>65%), but this was unavoidable due to the GC-rich nature of the genome.
- BLAST against the *S. coelicolor* database showed no significant matches against other sequences in the genome.
- Expected low degree of secondary structure
- Primers were purchased from Invitrogen and were purified by desalting.

2.7.2 qPCR assay optimisation

In order to find the most favourable reaction conditions the assays were optimised by running single wells at different primer concentrations and temperatures, then selecting the conditions under which amplification occurred the fastest. Primer concentration was not found to make a significant difference to reaction efficiency so 250 nM was used in all experiments. While there were minor differences in optimal temperature between primer pairs, it was found that 60°C worked well for all of them.

2.7.3 qPCR assay conditions

A four-point standard curve was generated for each 48-well plate using the relevant primer pair and 10-fold serial dilutions of genomic DNA which had been diluted to 50 ng/μl and fully dissolved prior to use in order to increase the accuracy of serial dilutions. Based on a genome size of 8.7 Mb, the concentrations of standards used were 10^3 , 10^4 , 10^5 and 10^6 genomes per μl while sample DNA was freshly diluted to approximately 10^5 genomes per μl. All standard and sample wells were carried out in triplicate.

To set up each reaction, 2.5 μl of DNA and 22.5 μl master mix were pipetted into the bottom of each well of a white qPCR plate (Bio-Rad), taking care not to leave any liquid on the sides of the well. Each run included a four-point standard curve and a set of three water-only blank wells. Once set up, the plate was sealed with an adhesive lid (Bio-Rad) then either run immediately or stored in the dark at 4°C for up to four hours.

Reagent	Volume (μl)	Final concentration
SYBRgreen	12.5 μl	25 %
ROX	0.2 μl	250 nM
40% DMSO	3 μl	5 %
Primers	0.63 μl	250 nM
Template DNA	2.5	$10^2 - 10^5$ genomes per reaction
ddH ₂ O	to 25 μl	

50°C	2 min	
95°C	10 min	
95°C	15 s	} x 40 cycles
60°C	60 s	
74°C	1 s	
Plate read		
Melting curve analysis from 50°C - 95°C		

To confirm that each well contained only the expected product a meltcurve was run covering the temperature range 50°C - 95°C, which showed a single clean peak in the rate of change of fluorescence when a single product was present. Aliquots were also run on a 2% TBE gel to confirm that the product was of the anticipated size.

2.7.4 Ct value calculation

The quantitation graphs for each well were visualised using the Opticon 3 software (Bio-Rad), then quantity calculations were generated using the same software. Quantitation is given in the form of a Ct value which corresponds to the cycle in which sample fluorescence rises above a threshold representing the baseline (the “take-off point”), after which amplification is exponential until reagents are exhausted. These thresholds were set manually with all curves displayed in logarithmic view (Figure 2.1, left), using the display setting “local trend”. Any wells generating curves which did not show a suitable exponential shape were removed from the analysis at this stage. Thresholds were positioned at the mid-point of the linear range (Figure 2.1, left). Ct values were exported to Excel for manual analysis.

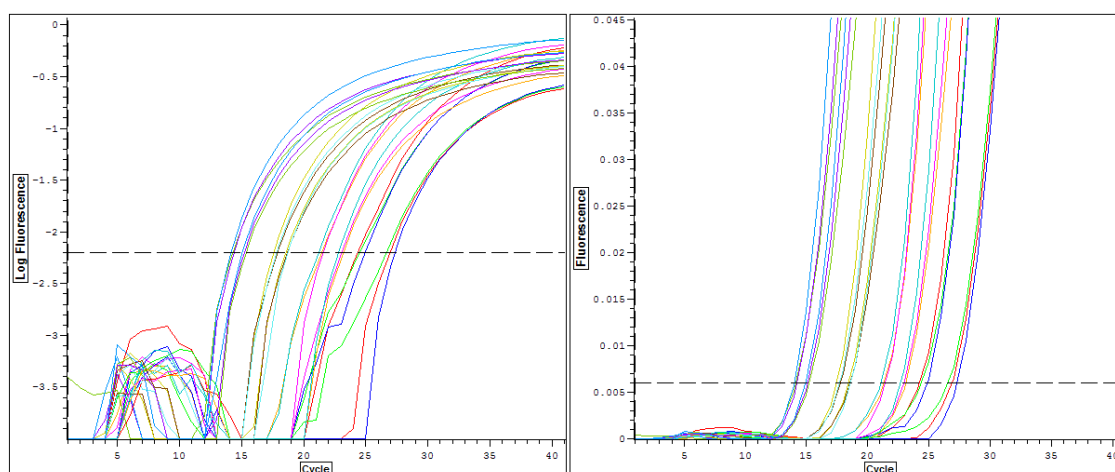


Figure 2.1 Positioning of qPCR quantitation threshold (take-off point) using Opticon 3 (Bio-Rad), in logarithmic view (left) and standard view (right).

2.7.5 Evaluation of data quality

As each sample was run in triplicate, the mean Ct of the three wells was calculated. The SEM of each triplicate was generally around 0.1 - 0.4 Cts, showing adequate repeatability. To generate the standard curve, the mean Ct of each dilution was plotted against the log of the number of genome equivalents in that standard and a regression line was drawn (Figure 2.2). An R^2 of >0.99 was considered acceptable, in accordance with MIQE guidelines. The slope of this line was used to calculate efficiency (E) using the formula: $E = (10^{(-1/\text{slope})}-1)*100$, where an efficiency of between 90% and 105% was considered acceptable in accordance with MIQE guidelines.

Genomes	Primer pair	Mean Ct	SEM
10^3	STI pro	24.3	0.1
10^4	STI pro	21.4	0.2
10^5	STI pro	17.5	0.4
10^6	STI pro	13.8	0.2
10^3	AFSK pro	23.1	0.1
10^4	AFSK pro	20.4	0.3
10^5	AFSK pro	16.5	0.1
10^6	AFSK pro	12.9	0.1

	R^2	slope	Efficiency (%)
AFSK pro	0.995	-3.46	94.6
STI pro	0.997	-3.53	92.1

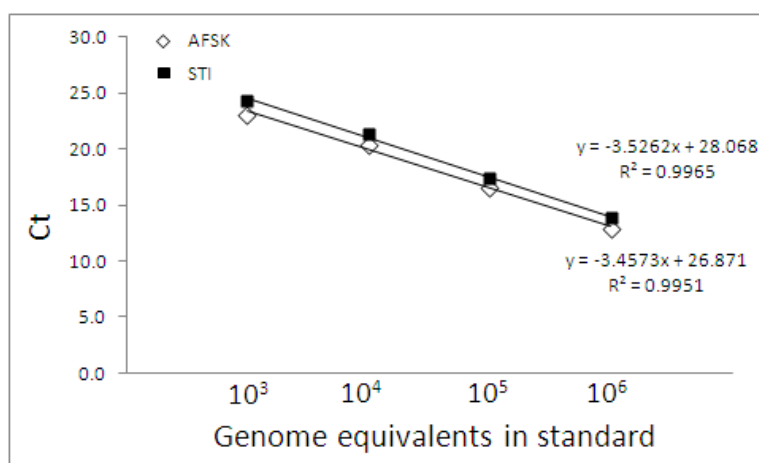
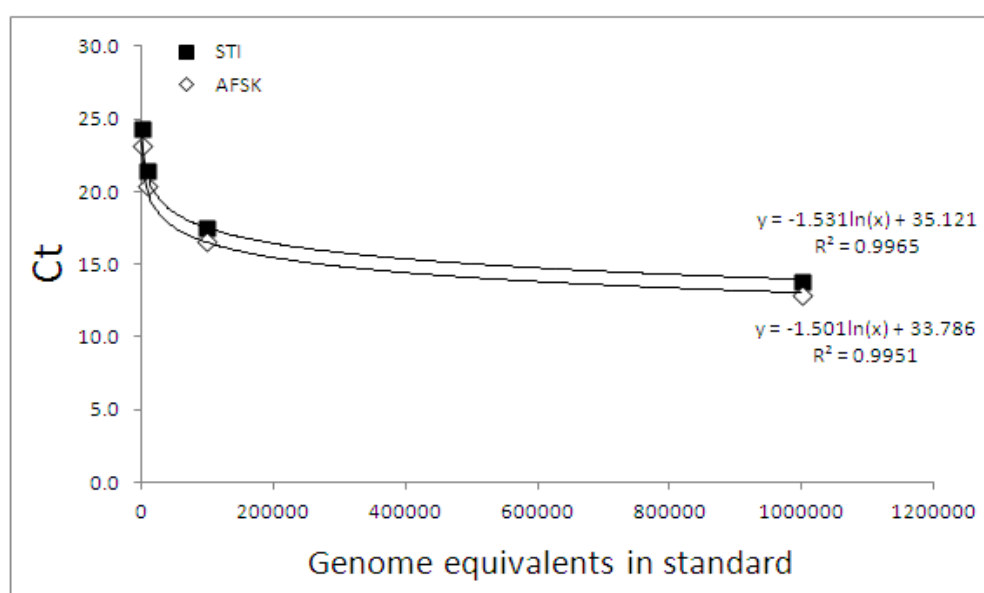


Figure 2.2 Calculation of qPCR reaction efficiency, repeatability and R^2 from gDNA standards.

2.7.6 Copy number calculation

In order to minimise noise resulting from small differences in efficiency between runs and between primer pairs, the delta Ct method was not used when comparing DNA quantities. Instead, copy numbers in each sample were calculated from the set of standards for each primer pair in each run using the following method: the Ct values of the standards were plotted against the number of genome equivalents in that standard, then a logarithmic trendline with the formula $y = -\text{slope} \cdot x + y\text{-intercept}$ was added using Excel. Power was calculated using the formula $P = (\text{slope} - Ct) / (2.303 \cdot y\text{-intercept})$, such that the copy number in the sample was 10^P .



Genome equivalents		Mean Ct
1000	10 ³	23.1
10000	10 ⁴	20.4
100000	10 ⁵	16.5
1000000	10 ⁶	12.9
1000	10 ³	24.3
10000	10 ⁴	21.4
100000	10 ⁵	17.5
1000000	10 ⁶	13.8

Sample	Ct	Power	Copy number	Primer pair
0U	16.5	5.3	10 ^{5.3}	STI
10U	18.2	4.8	10 ^{4.8}	STI
50U	18.0	4.9	10 ^{4.9}	STI
100U	18.9	4.6	10 ^{4.6}	STI
0U	15.5	5.3	10 ^{5.3}	AFSK
10U	16.8	4.9	10 ^{4.9}	AFSK
50U	16.2	5.1	10 ^{5.1}	AFSK
100U	16.9	4.9	10 ^{4.9}	AFSK

Figure 2.3 Copy number calculation from logarithmic gDNA standard curves.

2.8 PCR-targeted mutagenesis (REDIRECT® technology)

Mutant strains were generated by REDIRECT® technology, a form of PCR-targeted deletion, according to the method developed by Datsekno and Wanner (2000) and Gust *et al.* (2003), using the manual available from http://strepdb.streptomyces.org.uk/cgi-bin/tn_inserts_info.pl

Briefly, a cosmid comprising the Supercos1 backbone containing an approximately 40 kb insert of *S. coelicolor* genomic DNA which included the gene to be deleted was transformed into a recombination-proficient strain of *E. coli* (BW25113/pIJ790). The entire open reading frame of the gene was replaced with an apramycin-resistance cassette using the λ-RED recombination system. This mutant cosmid was then transferred into *S. coelicolor* by conjugation from a methylation-deficient strain of *E. coli* (ET), whereupon a double-recombination event replaced the genomic copy of the gene with the resistance cassette. Replacement was confirmed by PCR.

2.8.1 Cosmids, template and primers used

All cosmids used in this study were sourced from the library held at the John Innes Centre.

Table 2.8 Cosmids used for PCR-targeted deletion in this study.

Gene	Cosmid	Co-ordinates of sequence replaced by cassette	Distance between gene and edge of insert
SCO5592	St7A1	6097613.....6097852	2 kb
<i>sco2950</i> (HupA)	StE59	3206403.....3206684	8 kb
<i>sco3375</i> (Lsr2)	StE126	3740100.....3740435	18 kb
<i>sco4076</i> (Lsr2)	StD25	4470261.....4470578	15 kb

The pIJ773 template was used for all constructs as this replaces the gene with a cassette containing Apramycin, which does not need to be excised (“scarred”) in order for complementation with plasmids using Hyg selection.

Gene-specific primers were designed in accordance with the REDIRECT® guidelines (Gust *et al.*, 2003) to replace the entire coding region of each gene to be deleted. These primers included 39 bp of sequence homologous to the sequence flanking the gene to be deleted combined with 19 - 20 bp of sequence homologous to the universal primer binding sites flanking the apramycin cassette on the pIJ773 plasmid. Primer sequences are listed in section 2.14.

2.8.2 Introduction of *S. coelicolor* cosmids into *E. coli* BW25113/pIJ790

E. coli BW25113 is a recombination-proficient strain which carries a chloramphenicol resistance marker and the arabinose-inducible λ -RED recombination genes on plasmid papproximately IJ790 which has a temperature-sensitive origin of replication (which cannot replicate above 30°C). This strain was made electrocompetent using the procedure detailed in section 2.5.11, then transformed by electroporation with around 100 ng of the cosmid to be targeted. Transformants were selected by spreading cells onto L-agar containing 50 µg/ml carbenicillin, 50 µg/ml kanamycin and 25 µg/ml chloramphenicol and incubating overnight at 30°C.

2.8.3 Replacement of the gene on the cosmid in *E. coli*

Cassettes were amplified by REDIRECT® PCR as detailed in section 2.6.4, then gel purified to remove the template plasmid pIJ773. *E. coli* BW25113/pIJ790 containing the cosmid to be targeted was grown in 10 ml L broth containing 50 µg/ml carbenicillin, 50 µg/ml kanamycin and 25 µg/ml chloramphenicol overnight at 30°C with shaking at 250 rpm. The following morning, 10 ml SOB (minus Mg) containing the same antibiotics and 10 mM sterile L-arabinose was inoculated with 1% of the overnight culture. This was incubated with shaking at 30°C until OD₆₀₀ approximately 0.6 (approx. 3-4 hours), whereupon the cells were made electrocompetent as per section 2.5.11. A 50 µl aliquot of cell suspension was incubated in an electroporation cuvette on ice with 150 ng gel-purified apramycin cassette amplified from pIJ773 with the relevant primers for that gene. Electroporation was carried out as per section 2.5.11. Cells were recovered for 60 min in 750 µl L, then one 20 µl and one 730 µl aliquot of this mixture were each plated onto L agar plates containing 50 µg/ml apramycin and 100 µg/ml carbenicillin. Replacement of the gene on the cosmid was confirmed by PCR using a pair of primers external to the gene. As BW25113 supports multiple copies of these cosmids, wild-type and mutant cosmids were often found to be present in the same cell. In these cases a cosmid prep was made as per section 2.5.2 and retransformed into DH5α or BW25113 to give a strain harbouring only the mutant cosmid.

2.8.4 Transfer of the mutant cosmids into *S. coelicolor* by conjugation

S. coelicolor A3(2) carries a methyl-specific restriction system which restricts incoming DNA which has been methylated by the Dam, Dcm and Hsd modification systems of *E. coli* (Gonzalez-Ceron *et al.*, 2009), therefore the cosmid had to be passaged through a non-methylating *E. coli* strain. The mutant cosmid was transformed into ET12567/pUZ8002 by electroporation (as before) and selected using apramycin and carbenicillin only, then confirmed again by PCR. The cosmid was then transferred to *S. coelicolor* M145 by intergeneric conjugation. ET12567/pUZ8002 harbouring

the mutant cosmid was grown overnight in L containing 25 µg/ml chloramphenicol, 50 µg/ml apramycin and 50 µg/ml kanamycin. The following morning, 1% of this culture was used to inoculate 10 ml L containing 25 µg/ml chloramphenicol and 50 µg/ml kanamycin which was grown until OD₆₀₀ reached 0.4-0.5. Cells were washed twice with 10 ml of L to remove antibiotics and finally resuspended in 1 ml L. 10 µl dense spore suspension (approximately 10⁸ spores) were added to 500 µl 2xYT broth and heat shocked at 50°C for 10 min, then allowed to cool. 0.5 ml of the *E. coli* cell suspension was mixed with 0.5 ml of the heat-shocked *S. coelicolor* spores and centrifuged briefly, then resuspended in 50 µl residual liquid. A dilution series from 10⁻¹ to 10⁻⁶ was made, with each step being made in a total volume of 100 µl ddH₂O. The entirety of each dilution was plated onto SFM agar + 10 mM MgCl₂ without antibiotics and incubated at 30°C for 16-20 hours. Each plate was then overlaid with 1 ml of a solution containing 0.5 mg nalidixic acid and 1.25 mg apramycin and incubation continued at 30°C until colonies had formed dark grey spores. Colonies were patched onto DNA plates containing 50 µg/ml apramycin with and without 10 µg/ml kanamycin. Exoconjugants found to be apramycin-resistant and kanamycin-sensitive were tested by PCR using primers external to the gene being replaced. Exoconjugants found to have the cassette but not the wild-type gene were streaked for single colonies twice on SFM containing 50 µg/ml apramycin, before being made into spore preps.

2.8.5 Verification of mutant *Streptomyces* strains

Finished strains were tested again for sensitivity to 10 µg/ml kanamycin on DNA plates and clones showing any growth were rejected. Replacement of the gene on the chromosome was confirmed by *Streptomyces* colony PCR using three pairs of primers: one pair internal to the gene (to confirm that it had been removed), one pair internal to the apramycin cassette (to confirm that it had been introduced) and one pair external to the gene (to confirm that the distance between them had increased to match the size of the apramycin cassette). A sample PCR confirmation is shown in Figure 2.5, which corresponds to clone Δ4076.G1. 2-5 clones of each mutant were produced.

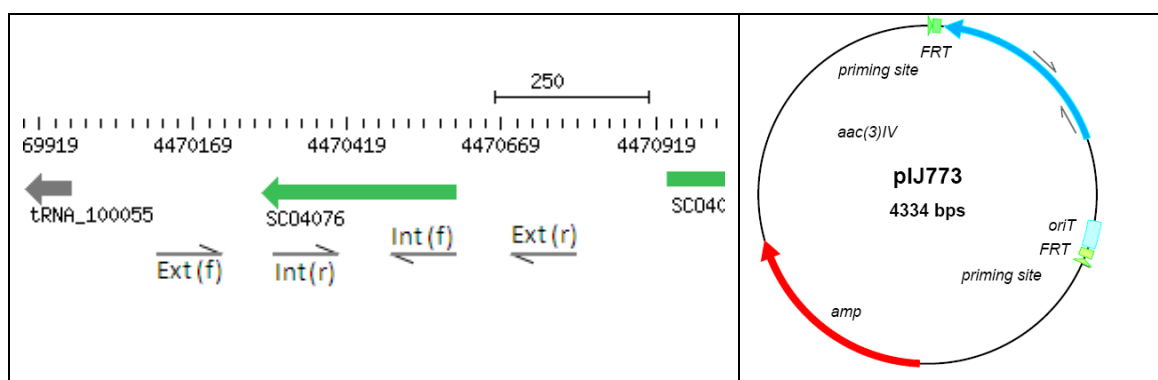


Figure 2.4 Sample primer locations for PCR confirmation of mutant construction.

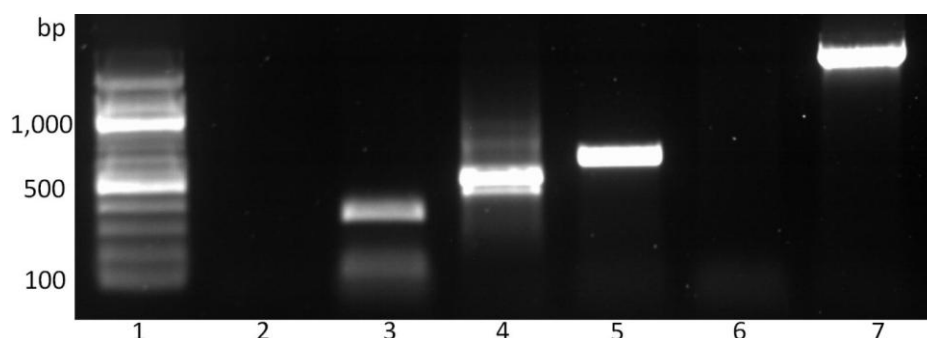


Figure 2.5 Sample PCR results confirming PCR-targeted replacement of SCO4076 with the apramycin cassette. Lane 1: NEB 1 kb ladder; lanes 2, 3 and 4: M145 gDNA template; lanes 5, 6 and 7: $\Delta 4076$ gDNA template; lanes 2 and 5: apramycin cassette internal primers (Apra-cassette-f/r); lanes 3 and 6: SCO4076 internal primers (NdeI-4076-f/4076-HindIII-r); lanes 4 and 7: SCO4076 external primers (BB4076test-f/r).

2.8.6 Complementation of the mutant phenotypes

Complementation of mutant strains was carried out *in trans* by introduction of the vector pMS82 containing the deleted open reading frame with its native promoter. This vector integrates at the ϕ BT1 *attP* site so that it is present in one copy per chromosome. All constructs were confirmed by Sanger sequencing. Constructs and pMS82 (empty vector) were conjugated into the respective mutant or WT strain by conjugation via ET12567/pUZ8002 and uptake was confirmed by PCR using a primer pair internal to the gene in question.

Table 2.9 Genomic co-ordinates of fragments used to complement $\Delta hupA$, $\Delta 3375$ and $\Delta 4076$.

Mutant strain complemented	Genomic co-ordinates of insert used for complementation in pMS82
$\Delta hupA$	3207075.....3206684
$\Delta 3375$	3740655.....3740435
$\Delta 4076$	4470794.....4470578

2.9 General protein methods

2.9.1 1D gel electrophoresis (SDS-PAGE)

Samples were boiled for 5 min in Laemmli SDS loading buffer before being run on precast RunBlue 16% acrylamide SDS-PAGE gels (Expedeon) in RunBlue SDS Tris-Tricine Run Buffer (Expedeon), with RunBlue Prestain Markers (Expedeon). Gels were run at 120 V through the stacking gel and at 160 V through the resolving gel, until the dye front reached the bottom edge of the gel. Gels were rinsed briefly in de-ionised water to remove excess buffer then stained in the Coomassie-based stain InstantBlue™ overnight. InstantBlue™ was removed by rinsing twice in de-ionised water before the gels were photographed using a digital camera.

2.9.2 Dialysis

Recombinant His-tagged protein was dialysed into protein storage buffer following purification, while nucleoid fractions were dialysed into DNase I digestion buffer (during DNase I digestion) and then water (immediately preceding methanol precipitation). 2 L of the appropriate buffer was prepared and pH-adjusted at 4°C in a large plastic beaker. Protein samples were pipetted into a pre-wetted 5 ml Float-a-Lyzer (MWCO 3.5-5 kDa; Spectrum Laboratories). Tubes were stirred gently in the buffer overnight at 4°C, then removed to a second, pre-chilled 2 L beaker of buffer in the morning and dialyzed for a further 2-3 hours.

2.9.3 Methanol precipitation

Typically 0.9 ml of dilute nucleoids were concentrated by precipitation with methanol for 1D electrophoresis and 5 ml dilute nucleoids were precipitated for 2D electrophoresis. The volumes given here apply to an 0.9 ml sample. First the proteins were denatured by boiling for 30 min with 50 µl β-mercaptoethanol and 50 µl 20% SDS. Once cool, the sample was mixed with 2.4 ml MeOH and 0.8 ml chloroform then vortexed briefly. 3.2 ml ddH₂O was added and it was vortexed again. The sample was then centrifuged at 5,000 rpm for 5 min and the majority of the upper (aqueous) phase was discarded. 1.5 ml MeOH was added and the tube was inverted three times. Precipitated proteins were pelleted by centrifugation at 5,000 rpm for 30 min at 4°C. The supernatant was removed and the pellet was allowed to air dry for 10 min before being resuspended in Laemmli SDS loading buffer for 1D gel electrophoresis, or in DryStrip rehydration buffer (GE Healthcare) for 2D gel electrophoresis.

2.9.4 2D gel electrophoresis

5 ml dialysed, DNase I-digested nucleoids (approx. 15-20 µg protein) were methanol-extracted and resuspended in 125 µl DryStrip rehydration buffer (GE Healthcare) containing 1% IPG buffer (GE Healthcare). This was applied to a DryStrip pH3-11 IEF strip (GE Healthcare) and rehydrated overnight, then separation was performed for 6469 V-h with a maximum voltage of 300 V. IPG strips were equilibrated for 15 minutes in buffer containing 10 mg/ml DTT and then for a further 15 minutes in buffer containing 25 mg/ml iodoacetamide to alkylate cysteine residues. The strip was applied to the top of a pre-cast NuPage 12% Bis-Tris gel (Invitrogen) and run at 200 V for 35 minutes. Gels were fixed in 50% MeOH/7% AcOH, stained in SyproRUBY and destained in 30% MeOH/5% AcOH. Spots were picked robotically, digested with trypsin (section 2.10.1) and subjected to peptide mass fingerprinting by MALDI-ToF (section 2.10.2).

2.9.5 Purification of His-tagged recombinant proteins by nickel affinity chromatography

Recombinant proteins were purified by cloning into vector pET15b (Novagen) which possesses an N-terminal in-frame 6xHis tag. Purification was carried out on an Äkta FPLC. *E. coli* BL21(DE3)pLysS strains (New England Biolabs) freshly transformed with the expression vector pET15b containing the appropriate insert was used to inoculate a flask containing 500 ml of L broth containing with 50 µg/ml carbenicillin and 12.5 µg/ml chloramphenicol, which was incubated at 37°C with shaking for around 3 hours until OD₆₀₀ reached 0.5, whereupon IPTG was added to a final concentration of 0.1 mM and growth was continued at 28°C for a further 4 hours with shaking. Cells were harvested by centrifugation and frozen at -80°C until needed (no more than four weeks later). The pellet from 500 ml culture was thawed on ice in 20 ml His buffer A in which two tablets of complete mini EDTA-free protein inhibitor cocktail (Roche Diagnostics Ltd) had been dissolved. Cells were disrupted by three passes through a French pressure cell at 1,000 psi then clarified by centrifugation at 10,000 rpm in a Sorvall SS-34 rotor for 30 min. The supernatant was further clarified by syringe filtration through an 0.22 µm membrane. A 1 ml HiTrap HP Ni Sepharose column (GE Life Sciences) was washed thoroughly with 15 ml His buffer B and then equilibrated with 15 ml His buffer A. Filtered lysate was loaded onto the column at a flow rate of 0.5 ml / min, washed with around 20 ml His buffer A until the trace flattened and then typically eluted using a 0-100% gradient over 40 min. 2 ml fractions were collected and the presence and purity of the recombinant protein was assessed by loading 10 µl of each fraction onto an SDS-PAGE gel which was stained in InstantBlue™. Pure fractions were pooled and dialyzed again protein storage buffer then stored at -80°C until needed.

2.9.6 His-tag removal using thrombin

The N-terminal 6xHis tag used to purify recombinant SCO5592 was cleaved off using thrombin (Novagen). For batch removal, a 50 µl reaction was set up which contained 1x Novagen thrombin cleavage buffer and approximately 15 µg recombinant His-5592. 0.1 µl of Novagen thrombin at a 1/10 dilution from the purchased stock (representing 0.01 U) was added and the reaction was incubated at room temperature. A 10 µl aliquot was removed at 2, 3, 5, 7 and 9 hours and mixed with 10 µl 2x Laemmli loading buffer to stop the reaction. Alternatively on-column cleavage was attempted using a recirculating peristaltic pump. Approximately 3.75 mg purified protein was loaded onto a 1 ml HisTrap HP nickel column (GE Life Sciences) and washed with 18 ml 1x Novagen thrombin cleavage buffer to remove unbound protein. 0.2 U thrombin (Novagen) was diluted in 0.8 ml 1x Novagen thrombin cleavage buffer and introduced to the system, which was then closed with a wet connection and recirculated at a flow rate of approximately 0.5 ml/min at RT for 2 hours.

2.10 Mass spectrometry

2.10.1 Trypsin digestion

Gel slices containing the protein/s of interest were cut into small pieces and washed twice in 200 mM ammonium bicarbonate/ 50% acetonitrile (pH 8). After destaining was complete the pieces were shrunk in 100% acetonitrile for 10-15 minutes then air dried for 10 minutes and reswelled in 50 mM ammonium bicarbonate/ 10 mM DTT. Proteins were alkylated for 30 minutes at room temperature in 50 mM ammonium bicarbonate/ 100 mM iodoacetamide. Gel pieces were washed twice more in 200 mM ammonium bicarbonate/ 50% acetonitrile, then shrunk again in 100% acetonitrile and reswelled in 10 mM ammonium bicarbonate containing 10 ng/µl modified porcine trypsin (Promega). After a 3 hour digestion at 37°C an equal volume of 5% formic acid was added.

2.10.2 Matrix-assisted laser-desorption ionisation time of flight (MALDI-ToF)

SDS-PAGE bands containing proteins to be identified by MALDI-ToF were excised with a clean scalpel and submitted to Dr Gerhard Saalbach at the John Innes Centre proteomics facility for trypsin digest and peptide mass fingerprinting.

2.10.3 ESI-Orbitrap MS/MS

After in-gel trypsin digestion, peptides were extracted with 5% formic acid/50% acetonitrile, dried down, and re-dissolved in 0.1% TFA. For LC-MS/MS analysis, a sample aliquot was applied via a nanoAcquity™ (Waters, Manchester, UK) UPLC™-system running at a flow rate of 250 nL min⁻¹ to an LTQ-Orbitrap™ mass spectrometer (Thermo Fisher, Waltham, MA). Peptides were trapped using a pre-column (Symmetry® C18, 5 µm, 180 µm x 20 mm, Waters) which was then switched in-line to an analytical column (BEH C18, 1.7 µm, 75 µm x 250 mm, Waters) for separation. Peptides were eluted with a gradient of 3-38% acetonitrile in water/0.1% formic acid at a rate of 0.67% min⁻¹. The column was connected to a 10 µm SilicaTip™ nanospray emitter (New Objective, Woburn, MA, USA) attached to a nanospray interface (Proxeon, Odense, Denmark) for infusion into the mass spectrometer. The mass spectrometer was operated in positive ion mode at a capillary temperature of 200 °C. The source voltage and focusing voltages were tuned for the transmission of MRFA peptide (m/z 524) (Sigma-Aldrich, St. Louis, MO). Data dependent analysis was carried out in orbitrap-IT parallel mode using CID fragmentation on the 6 most abundant ions in each cycle. The orbitrap was run with a resolution of 30,000 over the MS range from m/z 350 to m/z 1800 and an MS target of 10⁶ and 1 s maximum scan time. Collision energy was 35, and an isolation width of 2 was used. Only mono-isotopic 2+ and 3+ charged precursors were selected for MS2. The MS2 was triggered by a minimal signal of 1000 with an AGC target of 3x10⁴ ions and 150 ms scan time using the chromatography function for peak apex detection. Dynamic exclusion was set to 1 count and 30 s exclusion with an exclusion mass window of ±20 ppm. MS scans were saved in profile mode while MSMS scans were saved in centroid mode.

2.10.4 MS/MS data analysis

Raw files obtained with the LTQ-Orbitrap™ mass spectrometer were processed with MaxQuant version 1.2.2.5 (Cox and Mann, 2008); <http://maxquant.org>) to generate re-calibrated pkl-files which were used for a database search using an in-house Mascot 2.3 Server (Matrixscience, London, UK). Searches were performed on the SPTreMBL database (sptrembl20111116) with taxonomy set to *Streptomyces coelicolor* (8371 sequences) using trypsin/P cleavage with 2 missed cleavages, 6 ppm precursor tolerance, 0.5 Da fragment tolerance, carbamidomethylation (C) as fixed, and oxidation (M) and acetylation (N-terminus) as variable modifications. The pkl-files from the three gel slices of each sample were combined in a merged Mascot search including a decoy database search. Mascot results were exported using a peptide expect value threshold of 0.05 and a significance threshold of 0.01 resulting in a false discovery rate (FDR) of 0.84% for 74N, 1.09% for 27N, and 1.3% for 44N. The data (raw files and Mascot dat files) is available in the PRIDE database (Martens *et al.*, 2005) (www.ebi.ac.uk/pride) under accession numbers 27029-27034. The data was converted using PRIDE Converter (Barsnes *et al.*, 2009) (<http://code.google.com/p/pride-converter>).

2.11 Next-generation sequencing

DNA samples were incorporated into a paired-end Illumina sequencing library and sequenced to give read lengths of 100 bp using Illumina sequencing technology at the Genome Analysis Centre, Norwich Research Park. Four samples multiplexed into one lane produced a total of 167 million reads which were processed to remove poor-quality sequence towards the end of each read. Sequence reads were aligned against the *S. coelicolor* genome using Bowtie by Dr Govind Chandra and visualised in International Genome Browser.

2.12 Microscopy

2.12.1 Light microscopy

Samples were placed on 76 x 26 mm glass slides (VWR international) and covered with a glass cover slip (18 x 18 mm; VWR international). Slides were observed at 40 times magnification with a Eclipse E800 (Nikon) in bright-field mode and images were captured using a Coolpix 995 digital camera (Nikon).

2.12.2 Cryo-scanning electron microscopy

High-resolution SEM was carried out by Kim Findlay at the John Innes Centre using a Zeiss Supra 55 VP FEG SEM with a Gatan Alto 2500 cryo system at an accelerating voltage of 3 kV.

2.13 Nucleoid isolation

Several different protocols are available for preparing nucleoids by lysozyme-detergent lysis (Cunha *et al.*, 2001; Drlica *et al.*, 1999; Murphy and Zimmerman, 1997). The Drlica method was selected as this gives a high yield of membrane-free nucleoids with a high retention of nucleoid proteins, while the Sarfert method was selected as it can be used successfully on *Streptomyces*.

2.13.1 10–30% linear sucrose gradient preparation

10% and 30% sucrose solutions were prepared in ddH₂O and syringe filtered through an 0.22 µm membrane to remove debris, then incubated at 4°C for 1-2 hours. A gravity-fed sucrose gradient pourer was used to make either 12 ml gradients in 15 ml Corex tubes (Drlica method) or 24 ml gradients in 30 ml Corex tubes (Sarfert method), taking care not to let bubbles disturb the flow of sucrose. Gradients were prepared not more than 2 hours before use and stored at 4°C in the dark until needed.

2.13.2 Drlica method of nucleoid isolation

E. coli DH5α was grown overnight at 37°C in 10 ml L without antibiotics. The following day 500 µl of this was used to inoculate 50 ml L and grown for 3-4 hours, until the culture had reached OD₆₀₀ of around 0.4 (i.e. mid-log phase). The culture was briefly chilled on ice then harvested by centrifugation at 4,000 rpm for 5 minutes. Cells were washed once in nucleoid prep digestion buffer then resuspended in 1 ml of that buffer. 250 µl of 100 mM Tris pH 8/50 mM EDTA containing 4 mg/ml lysozyme was added to the resuspended cells and incubated on ice for 60 seconds, at which time the sample was mixed with 1,250 µl chilled Drlica lysis buffer and incubated at room temperature for 5 minutes until the turbidity had visibly reduced. After no more than 10 minutes, 1 ml of the resulting lysate was layered on top of a 10 ml 10-30% linear sucrose gradient in a 15 ml Corex tube. Gradients were centrifuged at 10,000 rpm in a Sorvall SS-34 rotor for 30 minutes. 500 µl fractions were collected from the top of the gradient using a cut-off 1 ml pipette tip. The DNA concentration of each fraction was estimated with a NanoDrop spectrophotometer (NanoDrop Technologies, DE). Nucleoid-containing fractions were dialysed overnight into DNase I digestion buffer according to the method described in section 2.9.2. 10 µl (100 U) Roche DNase I was added per 5 ml of pooled nucleoid fractions and mixed by gentle inversion after 2 hours of dialysis. Following 16 hours of dialysis the sample was centrifuged at 13,000 rpm in a micro centrifuge to remove insoluble debris.

2.13.3 Sarfert method of nucleoid isolation

50 ml mycelium was harvested by centrifugation at 3,000 rpm for 5 min at 4°C in a Sorvall GS3 rotor then washed once in an equal volume of nucleoid prep digestion buffer. This was centrifuged as before and resuspended in 5 ml nucleoid prep digestion buffer containing one tablet of complete mini EDTA-free protease inhibitor cocktail (Roche). 50 µl of nucleoid prep digestion buffer containing 10 mg/ml lysozyme was added and the mixture was incubated at 4°C for 90 seconds. 5 ml Sarfert lysis buffer was added and the mixture was incubated for 20 mins at 10°C with occasional gentle inversion. The NaCl concentration was then adjusted to 200 mM and the mixture was incubated at room temperature for a further 10 mins until slight visible clearing had occurred, indicating lysis. 4 ml of lysate was loaded onto a 24 ml 10-30% linear sucrose gradient in a 30 ml Corex tube and centrifuged at 10,000 rpm in a Sorvall SS-34 rotor for 60 minutes. 1 ml fractions were collected from the top of the gradient using a cut-off 1 ml pipette tip. The DNA concentration of each fraction was estimated with a NanoDrop spectrophotometer (NanoDrop Technologies, DE). Nucleoid-containing fractions were dialysed overnight into DNase I digestion buffer according to the method described in section 2.9.2. 10 µl (100 U) Roche DNase I was added per 5 ml of pooled nucleoid fractions and mixed by gentle inversion after 2 hours of dialysis.

2.14 Oligonucleotides

Table 2.10 Oligonucleotides used in this study. Asterisks indicate primers used for sequencing.

Primer name	Sequence (5' – 3')	Amplify
Gene amplification primers		
Asp718-2950	ATATATATGGTACCGTACGGGACGCGGCGAAGG	Asp718-P _{HupA} -HupA-HindIII (for insertion to pMS82)
2950-HindIII	ATATATATAAGCTTCTACTTGCCCTTGCGGCT	
Asp718-3375	ATATATATGGTACCGGTTTATTTACCGACGTGG	Asp718-P ₃₃₇₅ -3375-HindIII (for insertion to pMS82)
3375-HindIII	ATATATATAAGCTTTTCAGCCGTTGGCCTGCTCG	
Asp718-4076	ATATATATGGTACCTCGGAAAGATCCCGGATCG	Asp718-P ₄₀₇₆ -4076-HindIII (for insertion to pMS82)
4076-HindIII	ATATATATAAGCTTTACCGCGCCTCGGCGAACG	
NdeI-2950	AAAAAACATATGATGAACCGCAGTGAGCTGGTGG	NdeI-HupA-HindIII (for insertion to pIJ10257)
2950-HindIII	AAAAAAAAGCTTCTACTTGCCCTTGCGGCTTCC	
NdeI-3375	AAAAAACATATGGTGGCACAGAAGGTTCAAGTCC	NdeI-3375-HindIII (for insertion to pIJ10257)
3375-HindIII	AAAAAAAAGCTTCGTACGAGCAGGCCAACGGCTGA	
NdeI-4076	AAAAAACATATGGTGGCGCAGCGTGTCTGGTGC	NdeI-4076-HindIII (for insertion to pIJ10257)
4076-HindIII	AAAAAAAAGCTTTACCGCGCCTCGGCGAACG	
NdeI-5556	AAAAAACATATGGTGAACAAGGCGCAGCTCGTAG	NdeI-5556-NdeI (for insertion to pET15b)
5556-BamHI	AAAAAACATATGCTACTTCTTGCGGGCGGTGGC	
NdeI-5592	ATATATATCATATGCTCGAGGAGGCTCT	NdeI-5556-BamHI (for insertion to pET15b)
5592-BamHI	ATATATATGGATCCTCAGCGGACGTGGTCCACG	
Test primers and sequencing primers		
pIJ10257(f) *	ACGTCCATGCGAGTGTCC	Across MCS of pIJ10257
pIJ10257(r)	CCAAACGGCATTGAGCGTC	
T7 promoter *	TAATACGACTCACTATAGGG Invitrogen primer #69348-3	Across MCS of pET15b
T7 terminator	GCTAGTTATTGCTCAGCGG Invitrogen primer #69337-3	
TeF *	AGGGGGATGTGCTGC	Across MCS of pGem-T
TeR	ACTCATTAGGCACCCAG	
pMS82-RSeq *	CTAGTAGTTCCTTCGTACC	Across MCS of pMS82
Apra-cassette(f)	GGAACCTATGAGCTCAGCC	Anneal within <i>aac(3)IV</i>
Apra-cassette(r)	TGATCGAGGCCCTGCGTGC	Anneal within <i>aac(3)IV</i>
REDIRECT primers		
BB2950KO(f)	GAACGATCCGCGCAGTCCCCGGCACTCCAAGC- GGAGCGCTCTAGAATTCCGGGGATCCGTCGACC	Cassette to generate Δ _{hupA}
BB2950KO(r)	ACGCCTCAACGGCAAGAAGAACACGGGAGTA- ACAACATGTCTAGATGTAGGCTGGAGCTGCTTC	
BB3375KO(f)	CCCCCGCCCTCAAGTTATCGATGAAAGG- AAATCCGGTGATTCCGGGGATCCGTCGACC	Cassette to generate Δ ₃₃₇₅
BB3375KO(r)	CGGCCGATGGCGCCCGCGGTGGCGGGG- TCGGGTGATCATGTAGGCTGGAGCTGCTTC	
BB4076KO(f)	TCATCGCGGAGATGCCGACGAAGGGACC- GATATTCTGTGATTCCGGGGATCCGTCGACC	Cassette to generate Δ ₄₀₇₆
BB4076KO(r)	CGTCCCGCCGGACACCGGCTCGCCTCGC- GCACGGCTCA TGAGGCTGGAGCTGCTTC	
BB5556KO(f)	GCCCTGTG CCCGTTCCGCCGCTGCCGCGGGTG- CC CCTGTCTAGAATTCCGGGGATCCGTCGACC	Cassette to generate Δ _{hupS}
BB5556KO(r)	CACATTCTGTTCTGGCGATGTCTCCGAAGGGGA- AGACGTGTCTAGATGTAGGCTGGAGCTGCTTC	
BB5592KO(f)	GGCCGCCCGCAGTCTCCGCGTCCGAGG- CCTGAGCATGATTCCGGGGATCCGTCGACC	Cassette to generate Δ ₅₅₉₂
BB5592KO(r)	GTGGCCTTCCCGGCCGAGCCGGTGCTG- CGATTCTCATGTAGGCTGGAGCTGCTTC	
Test primers to confirm PCR-targeted deletion		

BB2950test(f)	TCGCGGGATGTCGGGCTTCGGA	Across SCO2950
BB2950test(r)	GATGCCGTGGACCCGGGAGGT	
BB5556test(f)	GCGGT TATATCTGCCTAACGTGCG	Across SCO3375
BB5556test(r)	GACGGGTCCCGGACACGCCG	
BB3375test(f)	GCCGGGGGTTTGTGTTTTTC	Across SCO4076
BB3375test(r)	AACCGCCCTGGATCTCGTC	
BB4076test(f)	GGACACGCGGGCAGATTTAG	Across SCO5556
BB4076test(r)	CTGATCCCATCGGACCATCG	
BB5592test(f)	GGCGAGGACGAGGGCAAGGG	Across SCO5592
BB5592test(r)	CTGTGCTTGCTCTCCTGTCGG	
DNase-seq qPCR primers		
sti1(f)	CATTGCGCCAGTTCTGTTCTC	Anneal within <i>sti1</i>
sti1 (r)	CCCCACGGGTTTCGGTAAC	
afsk (f)	TTCTTTGCCCGGAATCAGGC	Anneal within <i>afsk</i>
afsk (r)	GACCGAAGTGCAGCGTCTC	
HaeIII-seq qPCR primers		
hrdB(f)	CCCGAGGGTACCGAGAAC	Anneal within <i>hrdB</i>
hrdB(r)	GACTACCTCAAGCAGATCGG	
sti1(f)	CACGGCGACCCGAGTGC	Anneal within <i>sti1</i>
sti1 (r)	CTTCGCCAACGAGTGCGTG	
afsk (f)	GAGGCGCTCCAGTCGAT	Anneal within <i>afsk</i>
afsk (r)	TCTGACCATGACGAACGTC	
actIIORF4(f)	GTCGCCGAGCGTCTGAT	Anneal within <i>actIIORF4</i>
actIIORF4(r)	CTGTCCACCGTACCGTGC	

Chapter 3 – Proteomic survey of the *S. coelicolor* nucleoid

3.1 Introduction

In order to assess the contributions of *Streptomyces* NAPs to secondary metabolism via the structure and function of the nucleoid, it was first necessary to identify the main players. As discussed in section 1.2, a number of NAPs have been identified in *Streptomyces* as homologues of NAPs discovered in *E. coli* or *Mycobacterium* but we expect that the list is incomplete. Two copies of HU are present, one of which resembles *E. coli* HU and one of which uses a novel lysine-rich tail to compact nucleoids in spores (Salerno *et al.*, 2009) and may also protect DNA against oxidative damage (Takatsuka *et al.*, 2011). Two paralogues of the H-NS-like protein Lsr2 and one copy of sIHF (Yang *et al.*, 2012) are present but not yet well characterised, both of which were originally identified in *Mycobacterium*. Multiple genes annotated as members of the Lrp/AsnC family can be found but it is not known whether any of these are truly global regulators, also equivalents of Fis and Hfq have never been found.

It is not practical to carry out a genetic screen for NAPs as they have such diverse functions and no easily screenable mutant phenotype. Therefore a proteomic approach was taken as a way of identifying credible candidates which could then be investigated further. We aimed to identify the set of NAPs which are used in the vegetative tissues of *Streptomyces* which could then be tested for pleiotropic effects on secondary metabolism, in particular on expression from cryptic clusters.

This chapter describes the use of gel-based and mass spectroscopic methods to identify the abundant proteins of the *S. coelicolor* nucleoid and the *in silico* analysis used to refine this list down to a set of candidate NAPs for further investigation.

Objectives:

- 1) Establish a protocol for isolating nucleoids from *S. coelicolor* grown in liquid culture.
- 2) Identify the “12 kDa protein” reported by Sarfert *et al.* (1983).
- 3) Identify a set of plausible candidate NAPs for further investigation.
- 4) Identify candidates to investigate as global RNA chaperones.
- 5) Use the datasets produced to look for evidence of post-translational modification.

3.2 System and approach

Several methods are available for enriching DNA-binding proteins from crude cell lysates, including affinity chromatography using heparin or DNA. For example Park *et al.* (2009) performed a DNA-affinity capture assay (DACA) using the promoter regions of the *S. coelicolor* antibiotic regulators *actII-ORF4* and *redD*. The set of proteins found to bind these sequences included several transcription factors as well as three RNA polymerase subunits, both DNA gyrase subunits, DNA topoisomerase I, HupA and siHF, as well as a number of cytosolic proteins most likely to be contaminants. However this method is not optimal for identifying NAPs because these proteins may not have a strong enough affinity for short DNA sequences under such artificial conditions, particularly if they bind co-operatively or have a preference for supercoiled DNA. To circumvent these deficiencies, the enrichment method selected for the work described in this chapter was whole-nucleoid isolation as this was expected to capture the most complete set of important nucleoid-attached proteins in a somewhat quantitative manner.

The most common preparative method of nucleoid isolation involves brief digestion of cells with lysozyme followed by detergent-induced lysis into a buffer containing positively-charged counterions such as Na⁺ ions or spermidine. An alternative approach is lysis induced by osmotic shock of spheroplasts but this typically results in a much lower yield of nucleoids (Cunha *et al.*, 2001). Many protocols are available which have been optimised for different species such as *E. coli* (Drlica *et al.*, 1999), *B. subtilis* (Guillen *et al.*, 1978), *C. crescentus* (Evinger and Agabian, 1977), *M. tuberculosis* (Basu *et al.*, 2009), *S. hygroscopicus* (Sarfert *et al.*, 1983) and even eukaryotic organelles (Garrido *et al.*, 2003; Kuroiwa and Suzuki, 1981). Following lysis the nucleoids are separated from cell envelopes and debris by centrifugation partway through a sucrose gradient.

The sedimentation profile and protein content of nucleoids varies between species, making comparison of results difficult, so the following criteria were chosen as a measure of success: nucleoids should migrate further into a sucrose gradient during ultracentrifugation than soluble DNA or protein alone; nucleoids should be associated with RNA, whose removal using RNase I affects the sedimentation profile (because RNA is thought to have a stabilizing role in isolated nucleoids); nucleoids should contain predominantly HMW DNA; most crucially, nucleoids should contain a plausible set of proteins enriched in known NAPs, global regulators, ribosomal proteins and known or predicted transcription factors. As detailed in section 1.2, the NAPs most likely to be seen were deemed to be HupA (SCO2950) and siHF (SCO1480), with lesser amounts of Lsr2 (SCO3375 and SCO4076), CRP (SCO3571), HupS (SCO5556) or Dps (SCO0596, SCO1050, SCO5756).

Bacteria grown in liquid culture were used because we wished to find novel NAPs which regulate secondary metabolite production under conditions relevant to industrial fermentation, which is carried out in liquid culture. An additional advantage is the removal of variability caused by morphological differentiation, which does not occur in liquid media in this species. The medium TSB:YEME34% (see section 2.2.3) was selected to encourage dispersed growth with reproducible timing of antibiotic production.

3.3 Gradient controls and Drlica method for preparing *E. coli* nucleoids

In order to ensure that general steps such as gradient preparation and nucleoid handling were being carried out correctly, I first used the method of Drlica *et al.* (1999) to isolate nucleoids from *E. coli* DH5 α and determined that they were highly enriched in NAPs.

To determine where free protein and DNA would migrate in a sucrose gradient, 1 ml aliquots of ddH₂O containing either 100 mg of BSA or 12.5 μ g of 1 kb NEB DNA ladder were applied to separate 15 ml 10 - 30% linear sucrose gradients (section 2.13.1) and centrifuged at 10,000 rpm in a Sorvall SS-34 rotor for 45 mins. All DNA remained in the top two fractions (i.e. the sample volume loaded), showing that LMW - MMW naked DNA was unable to enter the gradient under these conditions (Figure 3.1,A – closed diamonds). A proportion of the BSA was able to travel through the first five fractions (Figure 3.1, A – open circles), showing that small amounts of soluble protein can enter a short way into the gradient.

Nucleoids prepared according to the method of Drlica *et al.* (1999) were prepared from exponentially-growing cultures of *E. coli* DH5 α (section 2.13.2) by application of cell lysate to a 15 ml 10 - 30% sucrose gradient and centrifuged at 10,000 rpm in a Sorvall SS-34 rotor for 60 minutes. As expected, these nucleoids were found to run approximately halfway into the sucrose gradient in either a single or a double opalescent band which was too viscous to be measured on the NanoDrop without prior dilution. In order to remove excess DNA and sucrose, 5 ml of nucleoid-containing fractions were loaded into a Float-a-Lyzer (MWCO 3.5-5 kDa; Spectrum Laboratories) and dialysed against 2 l of DNase I digestion buffer at 4°C. After 2-3 hours, 10 μ l Roche DNase I (100 U) was added to the Float-a-Lyzer which was mixed by gentle inversion and dialysis was continued overnight with one change of buffer. After 16 hours the samples were centrifuged for 1 minute at 7,000 rpm in a micro centrifuge and the proteins in the resulting pellet and supernatant were resolved on a RunBlue 12% SDS-PAGE gel then stained with the Coomassie-based stain InstantBlue (Expedeon). A selection of bands present in one lane but not the other

were excised and identified by trypsin peptide mass fingerprinting (sections 2.10.1 and 2.10.2), which are marked in Figure 3.1 with arrows. The supernatant contained large amounts of small proteins (Figure 3.1, C) including Fis, Hfq and HU. A strong band seen at around 35-40 kDa was assumed to be DNase I added during preparation. The pellet (Figure 3.1, C) did not contain NAPs but instead contained larger proteins of which two were identified by MALDI-ToF as primary metabolic enzymes. As these proteins were not solubilised by the DNase I digest they most likely represented insoluble debris which had co-sedimented with the nucleoids. The 1D SDS-PAGE profile of proteins was the same whether a single or double nucleoid peak was seen in the sucrose gradient profile (not shown).

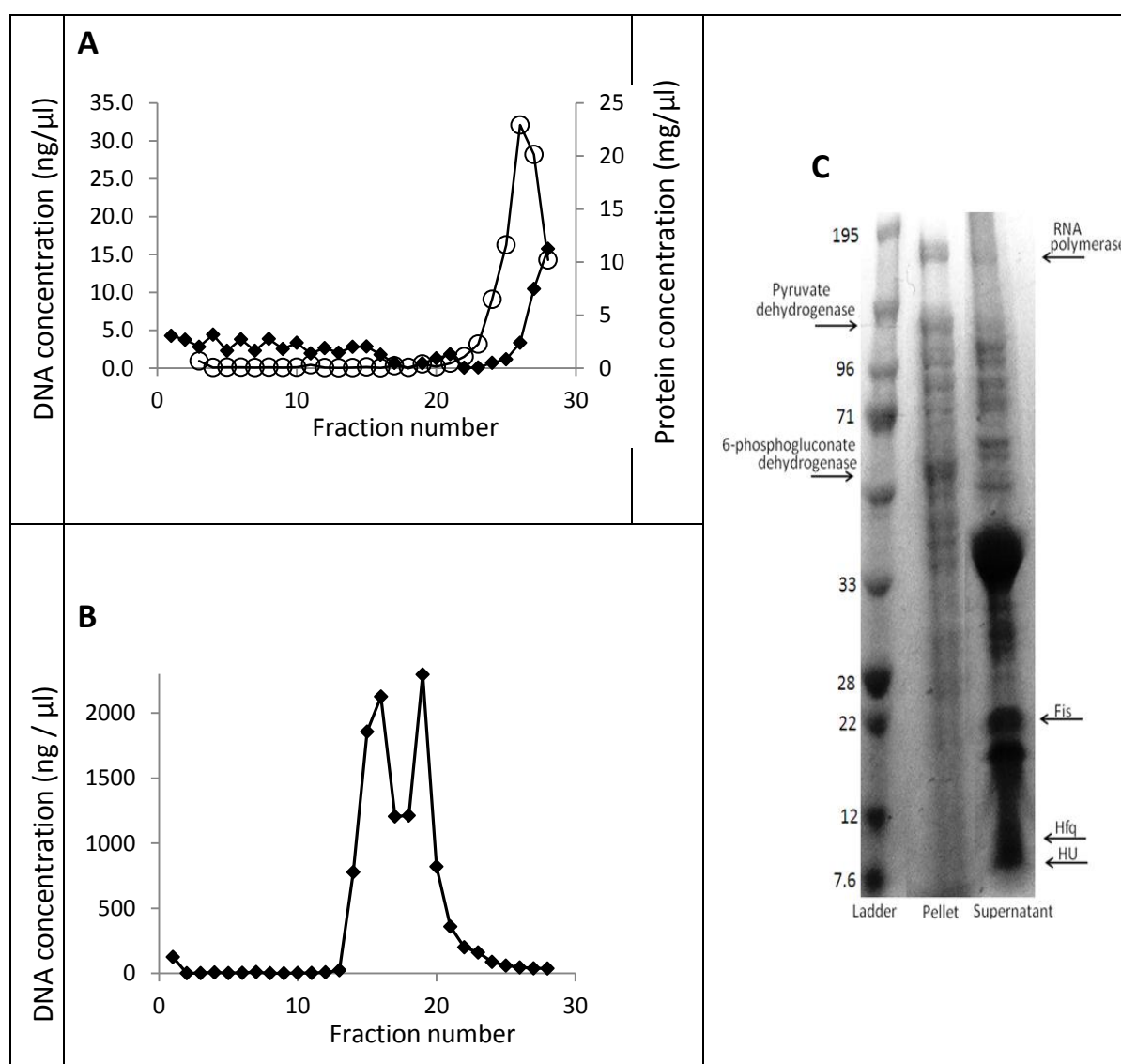


Figure 3.1 Validation of general nucleoid isolation procedures in *E. coli*. A: Distribution on a 10 - 30% linear sucrose gradient of naked DNA and soluble protein. Open circles = BSA; closed diamonds = 1 kb NEB DNA ladder. B: distribution on a 10 - 30% linear sucrose gradient of *E. coli* nucleoids isolated using the Drlica method. Fraction 1 is the bottom of the gradient and fraction 29 is the top fraction in both cases. C: protein content of recovered *E. coli* nucleoids as analysed by 1D SDS-PAGE. Proteins identified by MALDI-ToF are labelled with arrows.

Not all the proteins detected in these samples were expected to be DNA-binding or to have a specific role in the nucleoid. For example ribosomal proteins co-sediment with nucleoids (Ohniwa *et al.*, 2011). Also some protein chaperones and metabolic enzymes involved in primary metabolism are so abundant in the cell that some degree of contamination is likely to be unavoidable, either by inadvertent co-sedimentation or perhaps from the retention of nascent proteins on the ribosomes.

3.4 Isolation of nucleoids from *S. coelicolor*

Initially the Drlica method of nucleoid isolation was attempted on *S. coelicolor* protoplasts prepared freshly from 50 ml liquid cultures grown to late vegetative phase (see section 2.3.5). In these gradients DNA was seen in the top three fractions only (Figure 3.2, left), indicating that the procedure had liberated only degraded DNA fragments and not intact nucleoids. Attempts to optimise the method by using non-protoplasted mycelium or varying the buffer conditions or degree of lysozyme digestion produced similar results. Modified versions of the lysozyme-detergent lysis protocol are sometimes used for Gram positive bacteria (Basu *et al.*, 2009; Guillen *et al.*, 1978) which include an incubation of 20 - 30 minutes at 10°C after the lysozyme treatment but before the addition of 0.2 – 1 M NaCl. In these protocols the addition of NaCl is thought to be the point of lysis. A protocol for isolating nucleoids from the related species *S. hygroscopicus* was previously described by Sarfert *et al.* (1983) which produced nucleoids whose protein content was dominated by a 12 kDa species whose identity was not known. This method was optimised for *S. coelicolor* grown to 44 hours (late vegetative / early stationary phase) by varying the degree of lysozyme digestion: the original method used 0.04 mg/ml lysozyme for 60 seconds while the optimum for *S. coelicolor* was found to be 0.1 mg/ml lysozyme for 90 seconds. Nucleoids were successfully obtained with this method (Figure 3.2, right, closed diamonds), with the DNA concentration of fractions forming a broad peak around the middle of the gradient profile. The DNA-rich fractions were rich in both high molecular weight (>30 kb) DNA and RNA and the position of the most DNA-rich fraction moved upwards in response to brief digestion with RNase I (Figure 3.2, right, open diamonds), consistent with RNA having a role in compacting isolated nucleoids.

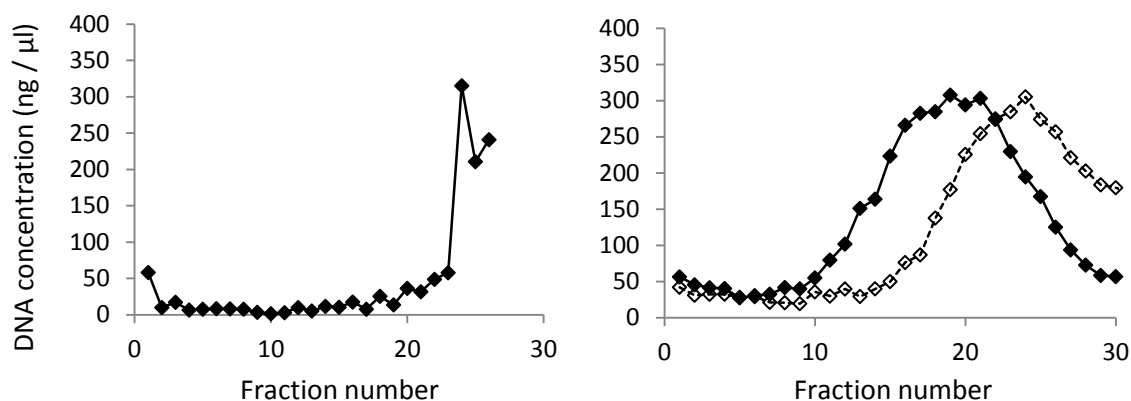


Figure 3.2 Left: sucrose gradient centrifugation profile obtained by applying the Drlica method. Right: sucrose gradient centrifugation profile obtained by applying the Sarfert method of nucleoid isolation to *S. coelicolor* mycelium. Without RNase I treatment: closed diamonds; with RNase I treatment: open diamonds.

Nucleoid fractions were pooled in a 5 ml Float-a-Lyzer® with a MWCO of 5 kDa and dialysed overnight against DNase I digestion buffer. DNase I digestion was found to solubilise all proteins in the nucleoid fractions, perhaps due to a lower level of membrane contamination than in *E. coli* preps, so the centrifugation step was omitted.

One pair of 30 ml gradients generally supplied sufficient nucleoids for one 2D gel or LC-MS/MS experiment. However the quantity of nucleoids recovered was limited by the amount of material which could be loaded onto each sucrose gradient without causing disruption. The release of nucleoids was less efficient in *S. coelicolor* than *E. coli*, resulting in a much lower overall yield per gradient.

3.5 Determination of *S. coelicolor* nucleoid protein content by 1DGE and 2DGE

Sarfert *et al.* (1985) reported a single protein – “S protein” – associated with the *S. hygroscopicus* nucleoid. However, the identity of this and other nucleoid proteins present in smaller quantities were not established. Separation of nucleoid proteins on a 1D 12% SDS-PAGE gel in this study produced a pattern of bands which was highly similar to that published by Sarfert *et al.* (1983), with the bulk of the protein forming a broad band around 12 kDa corresponding to the “S protein”. Additional bands were seen at ~9 kDa, ~15 kDa, ~17 kDa and ~19 kDa, with many more faint bands and smudged regions due to the complexity of the mixture. A number of these were excised and subjected to trypsin peptide mass fingerprinting (sections 2.10.1 and 2.10.2) but in most cases multiple non-significant hits were returned, suggesting the bands contained a mixture of too many proteins to be analysed in this manner.

2D electrophoresis in combination with peptide mass fingerprinting has proven to be a valuable tool in *Streptomyces* proteomic research (Hesketh *et al.*, 2002; Kim *et al.*, 2005a; Kim *et al.*, 2005b) therefore this method was used to separate the proteins more fully. Nucleoid proteins were separated in two dimensions (see section 2.9.5) with the first dimension being charge (IEF strip pH 3 – 11, GE Healthcare) and the second dimension being size (12% Bis-Tris PAGE, Invitrogen). Gels were stained with SYPRO® Ruby because the Coomassie-based stain InstantBlue (Expedeon) was found to interact with the IPG buffer as well as the proteins, creating excessive background staining. As expected, a large proportion of the protein was contained in one large spot of approximately 12 kDa on the right hand side of the gel, indicating that the protein/s in that spot were highly basic (Figure 3.3). Three biological replicates were performed with good agreement between the spot patterns. Spots were picked from one representative gel then digested with trypsin and identified by peptide mass fingerprinting. Surprisingly, the proteins identified did not include any of the predicted NAPs but rather an assortment of DNA-binding and ribosomal proteins, plus several tellurium-resistance proteins, chaperones and the putative anti-anti-sigma factor BldG. Many of these would certainly be found within a nucleoid but they would not be expected to be the most abundant proteins. As with the 1D gels, a significant proportion of spots picked and subjected to peptide mass fingerprinting (again using trypsin and MALDI-ToF) returned multiple or non-significant hits. In particular the extremely large spot adjacent to spot 21 (which presumably represents “S protein”) could not be identified, suggesting that the mixture was still too complex.

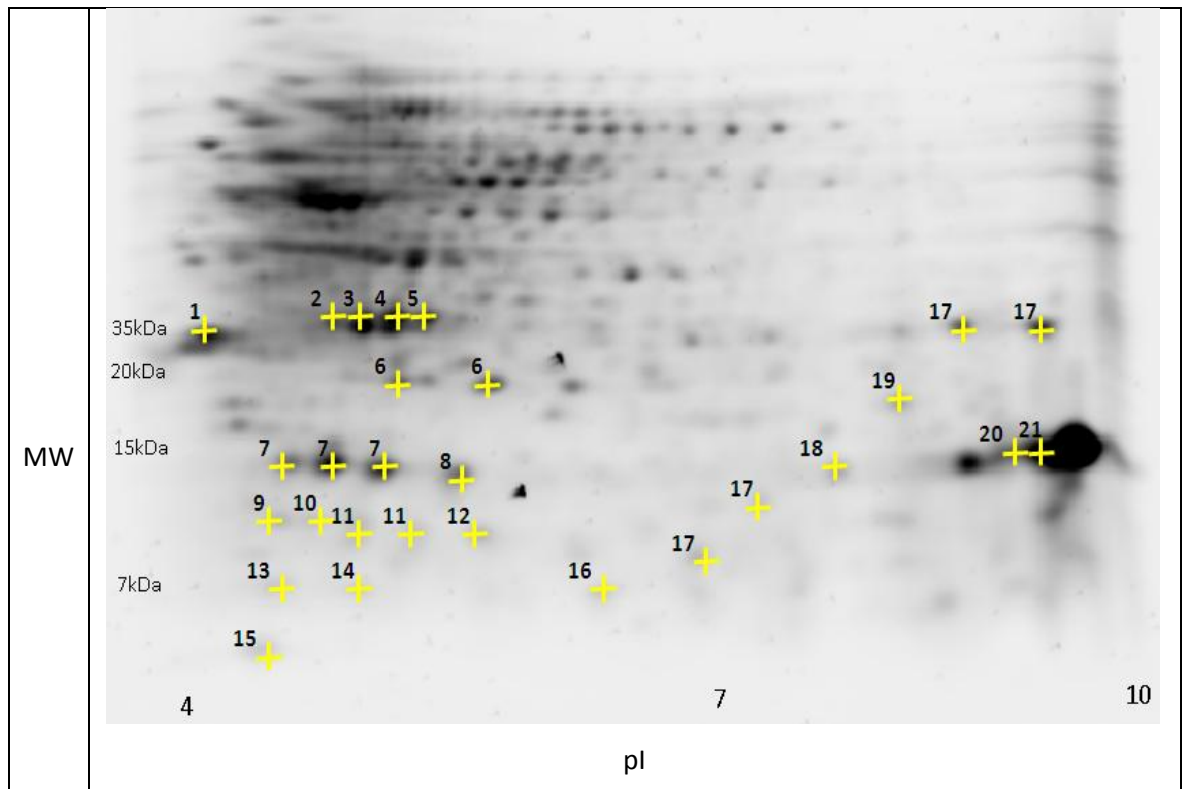


Figure 3.3 Representative 2D gel of proteins in *S. coelicolor* nucleoid fractions. 1st dimension: IEF separation over the pI range 3 - 11. 2nd dimension: SDS-PAGE separation on a 12% Bis-Tris gel (Invitrogen). Molecular weight and pI are an approximation based on the properties of the proteins identified in different regions. Spots identified by peptide mass fingerprinting (MALDI-ToF analysis of trypsin-digested peptides): 1 - SCO4277, tellurium-resistance protein; 2 - SCO4762, GroEL; 3 - SCO4442, transcriptional regulator (AbaA); 4 - SCO3198, transcriptional regulator (deoR-family); 5 - SCO4827, malate dehydrogenase; 6 - SCO4614, nucleotide-binding protein YajQ; 7 - SCO3767, tellurium-resistance protein; 8: SCO2612, nucleoside diphosphate kinase; 9 – SCO0722, putative hydrolase; 10 – mixture of SCO5880, transcriptional activator, RedY, and SCO3767, tellurium-resistance protein; 11 – SCO3549, anti-anti-sigma factor BldG; 12 – SCO0682, Topoisomerase IB; 13 – SCO3793, Yjqa; 14 – SCO4295, DNA-binding cold-shock protein; 15 – SCO3599, aminopeptidase; 16 – SCO0876, Holliday junction resolvase; 17 – mixture of SCO3906, RpsF, and SCO4716, RpsH, and SCO4714, RplE, and SCO4717, Rpf; 18 – mixture of SCO4184, aerial mycelium formation protein AmfC, and SCO3571, transcriptional regulator CRP; 19 – SCO5676, aminotransferase GabT; 20 – SCO5077, transcriptional regulator ActVA2; 21 – SCO4295, DNA-binding cold-shock protein.

3.6 Experimental design for LC-MS/MS of nucleoid fractions

As many protein spots could not be identified using the 2D electrophoresis approach, and because the set of proteins which were successfully identified did not include the expected proteins HupA and SlhF, another approach was designed which combined the same nucleoid isolation procedure with LC-MS/MS of peptides produced by trypsin digestion of large gel slices cut from 1D SDS-PAGE gels of whole nucleoid fractions.

In order to study a given subcellular fraction, a comparison is generally made between its protein composition and that of a whole-cell lysate or different fraction. However this approach may not be optimal when studying NAPs because of their very great abundance in a large cellular structure and because the amounts in the cytoplasmic “pool” are unknown. A study was published after the bulk of this work had been completed where proteins from isolated nucleoids of *E. coli*, *P. aeruginosa*, *B. subtilis* and *S. aureus* were analysed by LC-MS/MS in a very similar experiment to this (Ohniwa *et al.*, 2011). This group subtracted suspected contaminants by comparing the composition of the nucleoids against cell envelope and “top matter” fractions (material remaining at the top of the sucrose gradient following ultra-centrifugation), then classifying proteins which were at least three times more abundant in the nucleoid fractions than the others as “csNAPs” (contaminant-subtracted NAPs). While this removed a reasonable proportion of the presumed contaminants, such as porins and metabolic enzymes, it also excluded the major NAPs H-NS and Hfq. In this study we aimed to identify promising candidate NAPs so we chose not to compare the nucleoid samples with other fractions (a “ruling out” approach) but rather to prioritise the most promising candidates from the full lists of detected proteins as they stand (a “ruling in” approach).

The mass spectroscopic approach used in this study was electrospray ionisation coupled to an LTQ Orbitrap mass spectrometer (Thermo Fisher) which combines an LTQ linear ion trap with an Orbitrap mass analyzer. The methods used are described in section 2.10.

The method of quantification used in this study was the Exponentially Modified Protein Abundance Index (emPAI) score. This method of quantification was found to be accurate for proteins spiked into LC-MS experiments (Ishihama *et al.*, 2005) but is based on coverage rather than peak intensity so artefacts are possible: saturation effects can be seen for highly abundant proteins, notably ribosomal proteins (Ishihama *et al.*, 2005), but Ohniwa *et al.* (2011) found that emPAI scores for NAPs such as HU, Dps and H-NS were correlated with their abundances as determined by Azam *et al.* (2000). emPAI scores are not readily comparable between samples so both emPAI score and rank within the sample are given when comparing quantities.

3.7 Nucleoid isolation for LC-MS/MS

In order to identify as complete a set of the major NAPs as possible, mycelium was harvested from rich liquid medium at three time points (Figure 3.4) representing early vegetative growth (27 h; mycelium appears beige-white), late vegetative growth/onset of secondary metabolism (44 h; mycelium appears red) and late stationary phase (74 h; mycelium appears blue-purple). The sucrose gradient profiles for nucleoids harvested at 27 h and 74 h were somewhat different from the one harvested at 44 h, with a narrower peak which may reflect a lower degree of nucleoid heterogeneity during those phases. The protein content at all three time points appeared broadly similar on a 1D gel. Three gel slices were cut from each lane covering molecular weights of 5 - 11 kDa, 11 - 13 kDa and 13 - 31 kDa respectively, in order to focus on the small proteins as NAPs are generally around 7 – 17 kDa. The peptides detected in the three slices were later reconstituted into a single set of results per time point during the Mascot search.

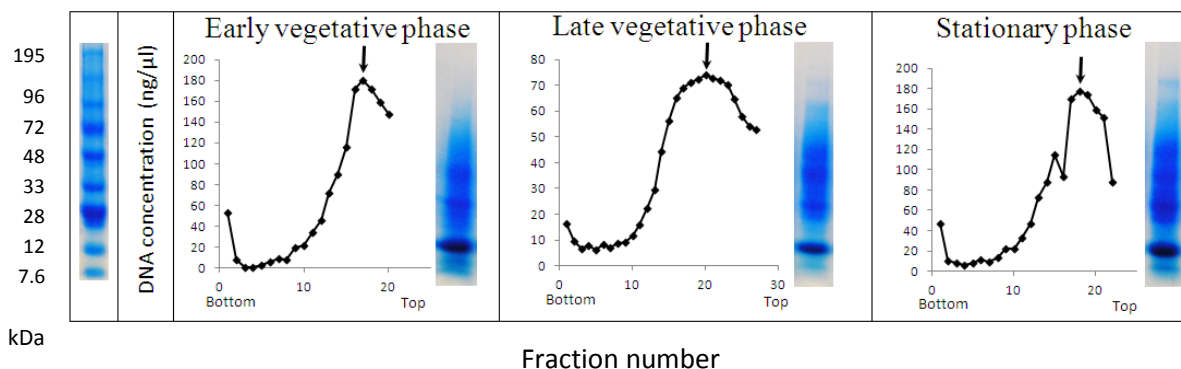


Figure 3.4 Isolation of spermidine nucleoids from *S. coelicolor* at early vegetative, late vegetative and stationary phase. Sucrose gradient profiles and 1D SDS-PAGE analysis of nucleoids harvested at each time point. Fractions taken forwards for LC-MS/MS analysis in subsequent sections are marked with an arrow. Growth phases sampled were: early vegetative phase (27 hours; left), late vegetative phase (44 hours; middle) and stationary phase (74 hours, right). Protein marker: 7.6 kDa, 12 kDa, 22 kDa, 28 kDa, 33 kDa, 48 kDa, 72 kDa, 96 kDa, 142 kDa and 195 kDa.

3.8 Protein identification and quantification

Proteins were identified as described in sections 2.10.3 and the raw files and Mascot dat files are available in the PRIDE database under accession numbers 27,029 – 27,034. The false discovery rates were 1.09% (27 hr), 1.30% (44 hr) and 0.84% (74hr) respectively. Putative functions were assigned where possible from the genome annotation given at strepdb (<http://strepdb.streptomyces.org.uk/>) or from BLAST searches against NCBI's non-redundant protein database.

The total number of proteins positively identified in each sample (Figure 3.5) were 627 (27 h, early vegetative phase), 884 (44 h, late vegetative phase) and 624 (72 h, stationary phase) respectively, representing around 10% of the *S. coelicolor* theoretical proteome (which is thought to encode 8,371 proteins in total). Of these proteins 11%, 15% and 15% respectively were annotated as known or putative transcriptional regulators, whereas a previous experiment using whole-cell lysate analysed by 2DGE/MALDI-ToF showed that known regulators made up only 6% (47/770) of the species identified in whole-cell lysates (Hesketh *et al.*, 2002), showing that enrichment for DNA-binding proteins had occurred. The true number of transcriptional regulators present was likely to be slightly higher as some may not be annotated. The number of lipoproteins and outer membrane proteins was extremely low, showing that nucleoids had been cleanly separated from the cell envelope.

	Early vegetative	Late vegetative	Stationary
Total number of proteins identified	627	884	624
Number and proportion of proteins annotated as transcriptional regulators	70 (11%)	130 (15%)	93 (15%)

Figure 3.5 Total number of proteins and transcriptional regulators identified in *S. coelicolor* nucleoid fractions isolated during early vegetative, late vegetative and stationary phase.

3.9 Proteins detected in the *Streptomyces* nucleoid

Within this section rank refers to the position of a protein within the total list of proteins identified. All proteins were assigned to the following five functional categories based on their annotated function: 30S/50S ribosomal proteins (R), chaperones/cold-shock proteins/redox-protective proteins (C), DNA-/RNA-binding proteins (D), proteins of unknown function (U) and enzymes/cytosolic function proteins/membrane proteins (E). The abundances of proteins varied

widely, with emPAI scores ranging from 53.86 down to 0.02, so that the protein composition at each time point could not be calculated from the number of species in each category. As a rough estimate of composition within each category, the summed emPAI values of its members were calculated as a percentage of the sum of all emPAIs scores at that time point (Figure 3.6). Overall, the mixture of proteins was comparable in composition to that observed by Ohniwa *et al.* (Ohniwa *et al.*, 2011) from nucleoids of *E. coli*, *B. subtilis* and *P. aeruginosa*. Most categories were of similar size in each sample: chaperones/cold-shock proteins/redox-protective proteins (C; 6 - 10%), DNA-binding proteins (D; 6 - 14%) and proteins of unknown function (U; 15 - 17%) did not vary considerably over time. The proportion of ribosomal proteins (R) was greatest during early vegetative growth (49%), less during late vegetative growth (32%) and lowest during late stationary phase (17%), reflecting the reduction in translational capacity in later stages of growth. Proteins known to be upregulated at later developmental time points, such as ActVA2 (SCO5077), ActVA4 (SCO5079) and SapA (SCO0409), were found to be more abundant during stationary phase than earlier time points. The proportion of DNA-binding proteins (D) shown in Figure 3.6 as proportion of summed emPAI (6%/9%/14%) is lower than the proportion of proteins annotated as transcriptional regulators shown in Figure 3.5 (11%/15%/15%) because the former reflects the abundance of each protein species while the latter does not.

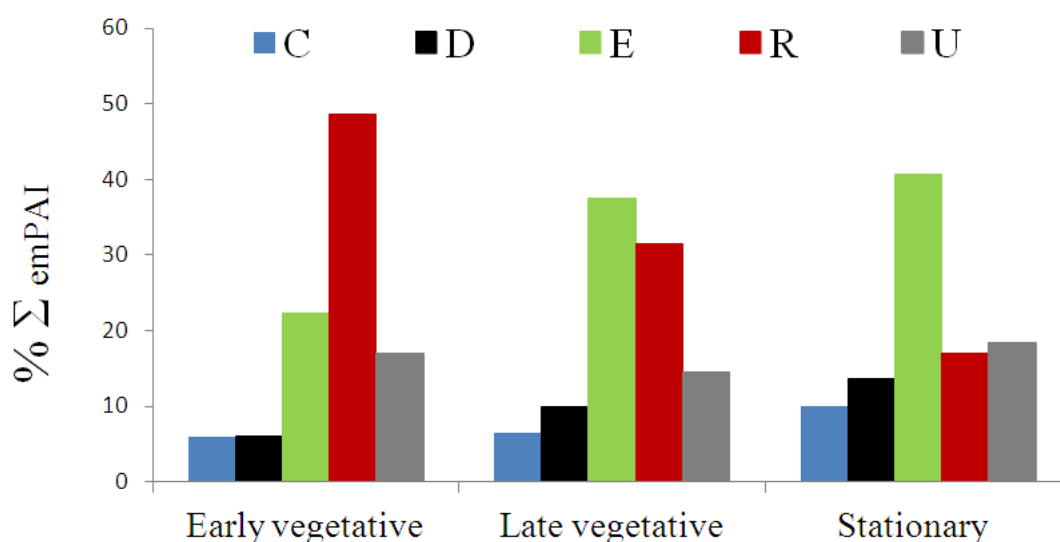


Figure 3.6 Approximate abundance of major protein categories identified in *S. coelicolor* nucleoid fractions at three time points. The value for each category represents the summed emPAI score of its constituent members, calculated as a percentage of the summed emPAI scores for all categories. Blue (C) = chaperones/cold-shock/redox-protective proteins; black (D) = DNA-/RNA-binding proteins; green (E) = enzymes/cytosolic function proteins/membrane proteins; red (R) = 30S/50S ribosomal proteins; grey (U) = proteins of unknown function.

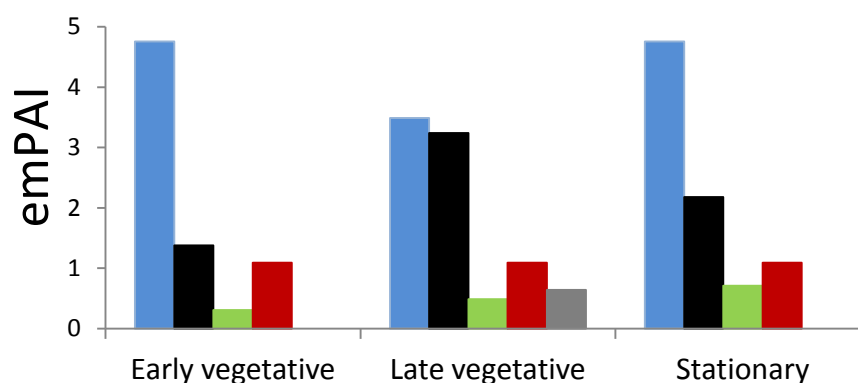


Figure 3.7 emPAI scores for known *S. coelicolor* NAPs at three time points. Blue: sIHF; black: HupA; green: HupS; red: Lsr2 (SCO3375); grey: Lsr2 (SCO4076).

As expected, HupA (black) was present in all three samples (Figure 3.7) but its relative abundance at late vegetative phase (emPAI = 3.24/37th) was far higher than at early vegetative phase (emPAI = 1.38/110th) or stationary phase (emPAI = 2.18/50th), which is consistent with its major role in vegetative growth (Azam and Ishihama, 1999). sIHF (blue) was also present in all three samples but was present at greatest abundance at stationary phase (emPAI = 4.76/12th), with less present during early vegetative phase (emPAI = 3.49/30th) or late vegetative phase (emPAI = 4.76/30th). IHF from *Salmonella* is likewise known to be upregulated during stationary phase (Azam and Ishihama, 1999; Mangan *et al.*, 2006). Of the two paralogues of Lsr2, SCO3375 (red) was detected at all three time points with reasonable abundance (emPAI = 1.09/129th; emPAI = 1.09/137th; emPAI = 1.09/104th) and SCO4076 (grey) was detected only at a low level during late vegetative phase (emPAI = 0.64/201st). The spore-specific HU homologue HupS (green) was detected at low levels at all time points but rose slightly during stationary phase (emPAI = 0.31/287th; emPAI = 0.49/261st; emPAI = 0.71/129th) with no peptides from the lysine-rich tail being detected at any time point, raising the possibility that it could be cleaved off in the nucleoids of liquid cultures as the tail is readily detected on recombinant versions. The developmental global regulator BldD was detected at high abundance, particularly during stationary phase (emPAI = 1.26/114th; emPAI = 2.12/69th; emPAI = 2.67/40th). Other important developmental regulators from the Bld, Whi and Wbl protein groups were detected at extremely low levels if at all, presumably because they are not strongly expressed in these non-sporulating cultures.

As seen by Ohniwa *et al.* (2011), many proteins involved in redox chemistry were found including multiple superoxide dismutases, thiol-specific antioxidant proteins and AhpC. The “armour hypothesis” suggests that these proteins are specifically associated with the nucleoid to protect the DNA from oxidative damage. Indeed, MnSOD of *E. coli* has been shown experimentally to associate non-specifically with DNA as a “tethered antioxidant” (Steinman *et al.*, 1994). Several

genuine NAPs (Dps and the HupS homologue Mdp1) have an additional redox function protecting DNA from Fenton chemistry. Tethered antioxidants are also known to be present in mitochondrial and chloroplast nucleoids (Kienhofer *et al.*, 2009; Myouga *et al.*, 2008).

3.10 emPAI threshold for classifying annotated transcriptional regulators as NAPs

The previous section produced a list of proteins (category D) annotated as transcriptional regulators, ranked by emPAI score. Only a subset of these will be NAPs or global regulators but the expected number is unknown. *E. coli* is generally reported to have around 10-20 NAPs but the list quoted is different between research groups, with no guarantee that all important *E. coli* NAPs have been discovered. The *Streptomyces* genome is particularly large and has a high proportion of transcriptional regulators in comparison to other species so we could not rule out the possibility that it also has more NAPs. Indeed, gene duplication within this lineage has generated two paralogues of HU (Salerno *et al.*, 2009), two of Lsr2 and three of Dps (Facey *et al.*, 2009). In order to visualise the data more clearly, the emPAI scores for all proteins in group D were ranked and displayed as a column chart, with the positions of known NAPs marked by arrows (Figure 3.9). It became apparent that there was no sound statistical basis for separating global from local regulators. Rather than a pair of normally-distributed groups (one representing NAPs/global regulators and one representing local regulators), the emPAI distribution was instead found to follow a power law, i.e. there are a small number of dominant species and a large number of low-abundance species which form a “long tail”, with no meaningful average for either group (Figure 3.8). A conservatively-placed threshold of the top 25 most abundant proteins consistently included HupA, siHF, CRP, BldD and the Lsr2 protein SCO3375.

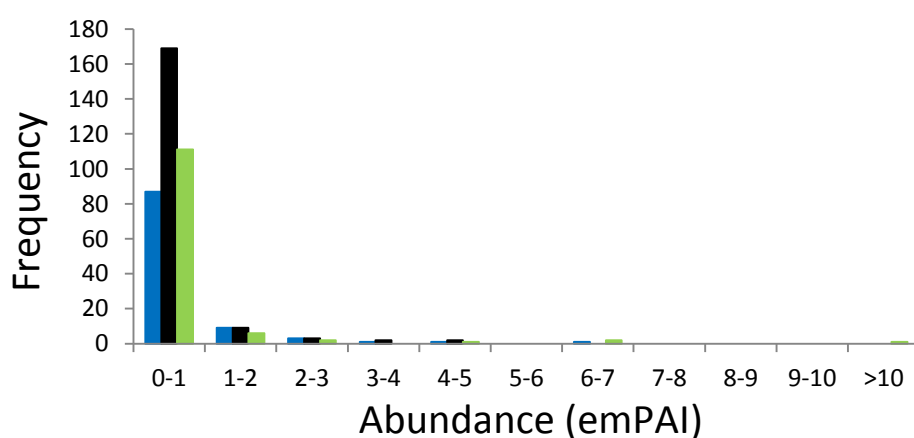


Figure 3.8 Protein abundances (emPAI scores) were found not to follow a normal distribution: most proteins were detected at a very low abundance (emPAI <1) while a very small number had a high abundance (emPAI >5). Blue: early vegetative phase; black: late vegetative phase; green: stationary phase.

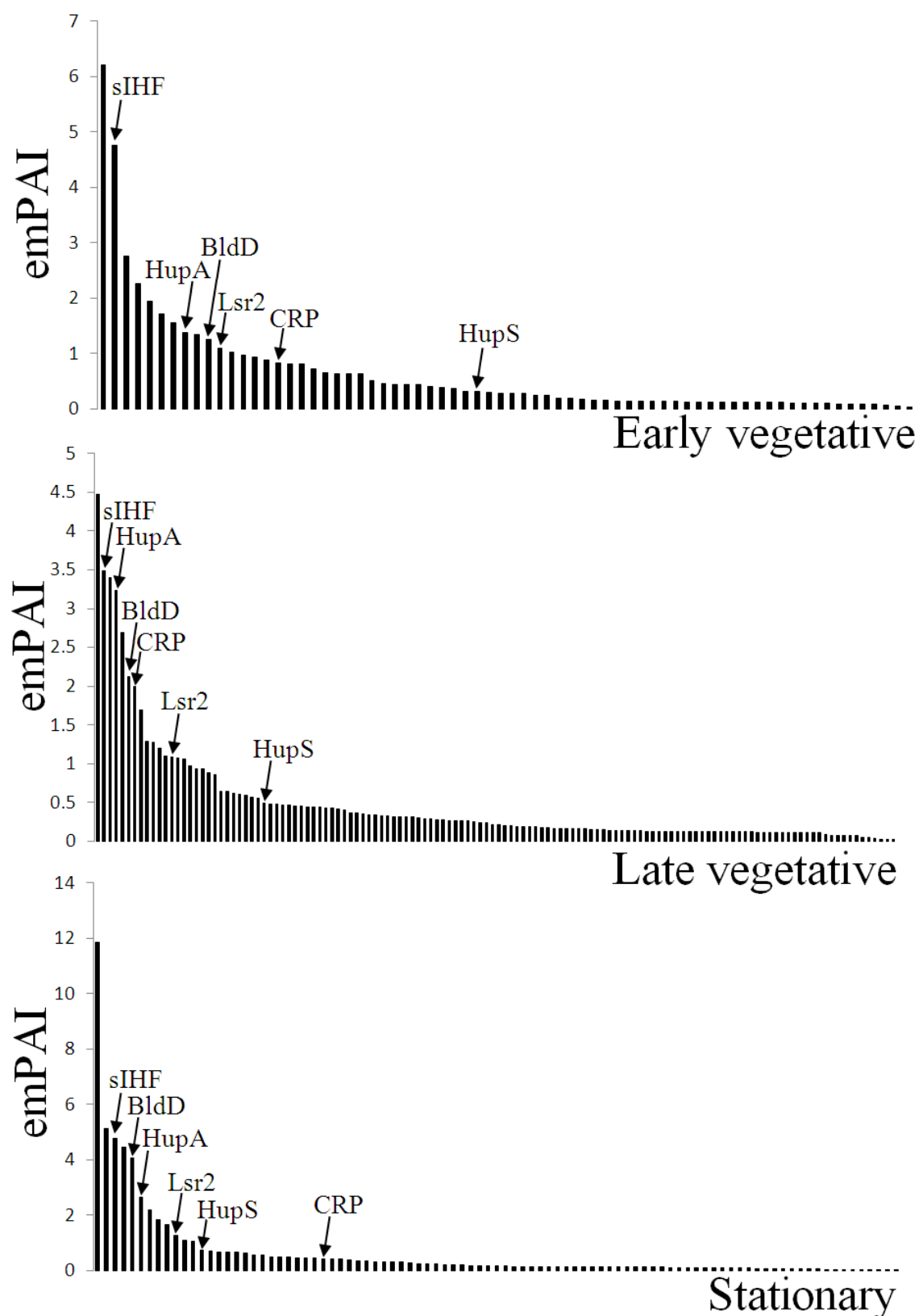


Figure 3.9 All DNA-binding proteins (category D) at each time point were ranked by abundance (emPAI score) and found to follow a power law distribution. Arrows indicate the positions of known *S. coelicolor* NAPs (sIHF, HupA, LsrA, HupS) and global regulators (CRP, BldD).

3.11 Prediction of nucleic acid binding ability

A NAP may contain a common DNA-binding domain such as a helix-turn-helix (e.g. Fis, Lrp) or alternatively may have an unusual DNA-binding domain (e.g. HU, H-NS). Therefore some of the proteins originally classified here into category U (uncharacterised) may represent the most interesting novel NAPs. In order to avoid discounting these, all category U proteins with an emPAI score greater than that of the 25th most abundant protein in category D (DNA-binding) were assessed for potential DNA-binding ability using *in silico* prediction tools. The helix-turn-helix prediction tool provided by NPS@ (Combet *et al.*, 2000), available at http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hth.html), did not identify any additional motifs which were not already present in the genome annotation. Another tool, DNAbinder (Kumar *et al.*, 2007); available at <http://www.imtech.res.in/raghava/dnabinder/index.html>), was used to estimate the likelihood of binding ability from amino acid composition. For example proteins with a higher proportion of lysine, arginine and glutamic acid are predicted to be more likely to bind DNA or RNA. The SVM model trained on the alternate dataset was selected as this was developed using full-length protein sequences. Of the proteins from category (U) with a suitable emPAI score, 25 were found to have predicted DNA-binding ability and were added to the list of potential NAPs. The top 25 proteins from the amalgamated list were then taken forward as the most promising candidates. It was not possible to classify the candidates further into DNA-binding and RNA-binding proteins because these prediction tools are not sensitive enough to discriminate between the two. Also several NAPs such as StpA (Mayer *et al.*, 2007) bind both substrates.

3.12 Candidate NAPs

This analysis generated a list of proteins of high abundance which contained either a recognisable DNA-binding motif or had an amino acid composition typical of nucleic acid binding proteins. In this section rank refers to the position of a protein in the top 25 amalgamated list of category D and category U proteins. The annotated name or function, SCO number, TrEMBL accession number, molecular weight, pI and rank at each time point are listed in Figure 3.10 for the most promising candidates.

Name/ function	SCO no.	TrEMBL no.	MW (kDa)	pI	Early veg. phase	Late veg. phase	Stat. phase
sIHF *	1480	Q9KXR9	11.5	10.4	4th	2nd	3rd
TerB	3767	Q9F2L4	16.6	5.3	1st	8th	11th
YajQ	4614	Q9F2U7	18.0	5.5	2nd	3rd	12th
AsnC-family	2140	Q9X7Z9	10.1	4.8	18th	1st	4th
HupA *	2950	P0A3H5	9.9	9.5	16th	4th	7th
WXG-100	5725	O86644	11.5	5.2	7th	7th	x
BldD *	1489	O52732	18.2	6.6	19th	9th	6th
HtH regulator	1839	Q9RJ29	7.6	10.5	x	19th	1st
RNA-binding	5592	P0A4Q4	8.7	9.9	x	6th	9th
YjqA	3793	Q9F325	13.6	7.1	x	11th	5th
Uncharacterised	5783	O69959	17.1	7.1	17th	5th	x
Response regulator	3013	Q9KYW8	25.2	6.0	8th	14th	ND
Regulator	4232	Q9L0Q9	17.8	5.7	6th	22nd	x
BdtA regulator	3328	Q9WX06	7.8	8.7	10th	x	14th
Uncharacterised	6482	Q9ZBJ2	16.1	5.0	3rd	x	x
Uncharacterised	2911	Q9S2H0	10.1	4.8	5th	x	x
Uncharacterised	2093	Q9S2W3	18.4	5.6	9th	x	x
Lysine-rich	4199	Q9FCH0	22.8	9.3	x	x	10th
CRP *	3571	Q9XA42	24.6	6.2	x	10th	x
LuxR regulator	0204	Q9RI42	25.4	5.1	12th	x	x
Lsr2 *	3375	Q9X8N1	11.7	6.8	20th	18th	13th
Putative regulator	5607	O69895	6.6	7.3	x	ND	25th
HupS *	5556	P0A3H7	22.3	11.2	x	x	16th
PadR regulator	1366	Q9AK96	22.8	7.2	x	16th	17th
AsnC regulator	4493	Q9KYP0	17.9	5.6	x	15th	20th

Figure 3.10 The 25 most promising candidate NAPs/global regulators found in *S. coelicolor*, arranged approximately by abundance. Rank at each time point represents its position within the amalgamated category D/category U generated in section 3.11. “X” indicates that the protein was detected at a lower abundance at that time point while “ND” indicates that the protein was not detected at all at that time point. The top 10 at each time point are shaded pale grey. The six known NAPs are marked with asterisks.

The key NAPs HupA, Lsr2 and siHF plus the global regulator BldD were in the top 25 at all three time points, while the late-developmental HupS was seen in the top 25 only during stationary phase (16th). The highly-conserved but poorly-characterised RNA-binding host factor YajQ (Teplyakov *et al.*, 2003), which was previously reported as being required to support bacteriophage ϕ 6 replication in *Pseudomonas syringae* (Qiao *et al.*, 2008), was abundant at all time points (2nd/3rd/12th), however the function of this protein in chromosomal gene expression and nucleoid architecture is not clear. A second protein annotated as an RNA-binding protein of unknown function, SCO5592, was very abundant at late vegetative phase (6th) and stationary phase (9th).

The Lsr2 proteins were present in lower amounts than perhaps might have been expected for an H-NS equivalent. While SCO3375 (20th/18th/13th) was in the top 25 at each time point, it was never in the top 10. This could be due to inaccuracy in quantitation or it could be a sign of redundancy with another H-NS-like protein present. *Mycobacterium* has one copy of Lsr2 alongside a second non-canonical H-NS protein which shares low sequence identity with *E. coli* H-NS (Sharadamma *et al.*, 2010) and so it is possible that another as-yet-undiscovered H-NS-like protein exists in *Streptomyces*. The other paralogue of Lsr2, SCO4076, was not found to be present in significant quantities in these nucleoids. The mutant phenotype of SCO4076 suggests that it has an important role in spore nucleoids, in particular in spore heat resistance and spore nucleoid segregation (see sections 5.11 and 5.12), so may simply be absent from these cultures due to lack of differentiation in liquid medium.

The presence of CRP in the top 10 during late vegetative phase supports the conclusion of Piette *et al.* (2005) that it is a global regulator in *Streptomyces*. Rex (SCO3320), the redox-sensing repressor thought to work in concert with CRP (Brekasis and Paget, 2003), was not abundant at any time point.

None of the three Dps proteins were detected at any time point during this study, suggesting that they do not have as strong a role in stationary phase liquid cultures as the *E. coli* version does. However several other stationary-phase proteins were found instead. HupS (16th) is a paralogue of HU with a long lysine-rich tail which binds DNA with high affinity and protects it from DNase I (Kumar *et al.*, 2010), UV (Salerno *et al.*, 2009) and harmful Fenton chemistry (Takatsuka *et al.*, 2011). Another protein with a strikingly similar lysine-rich tail (SCO4199, 10th) was found to be highly upregulated during stationary phase (Figure 3.11). This protein was of no currently recognised family and could represent a new class of NAP with a wholly novel mode of action. Similar lysine-rich tails have evolved independently on other proteins such as HU from

Deinococcus radiodurans (Grove, 2011) and the *Streptomyces* proteins SCO2075 (a developmental regulator), HrdB (the major vegetative sigma factor) and several other DNA-binding proteins (SCO2599, SCO0653, SCO6503, SCO5750). In each case the “core” domains of the proteins are conserved between *Streptomyces* species while the lysine-rich tails share amino acid composition rather than sequence similarity. The helix-turn-helix regulator SCO1839 ($x/19^{\text{th}}/1^{\text{st}}$) was also highly upregulated in stationary phase. This protein is very small (7.6 kDa) so presumably consists of a DNA-binding domain alone, with no ligand-binding domain.

```
>SCO4199
MAITDDL R K T F S D P T P L Y F A A G T A D L A F Q Q A K K V P V L V E Q L R A E A P A R I D A V R N T D P K A V
Q E K A T A R V K E T Q G S L Q A K A N E L F G N L D A D F K K L G E T A Q D L A L R S V G V A A E Y A V K A R E T Y E
K V A E R G E H A V R T W R G Q A A D G I E D L A E D V E D L A V A V E P A K P A P A R P A T P K P A T A K S E P V R V
T E D R S A A A K K T P A P R K A P A K K N T T A K K T T P P A K
```

Figure 3.11 The amino acid sequence of the uncharacterised protein SCO4199 shows C-terminal lysine-rich repeats similar to HupS or histone H1. Basic residues are shaded in yellow.

Some of the other highly-abundant DNA-binding proteins could be tentatively grouped into a particular class of transcription factor (e.g. IciR, TetR, DeoR) and two of them may be part of two-component signalling systems as they are each adjacent to a sensor kinase which is presumably their partner. SCO0204 is a response regulator annotated as belonging to the LuxR family and is located immediately upstream of a sensor kinase (SCO0203) which has been shown to regulate not only SCO0204 but also the orphan response regulator SCO3818, which is 65% identical at the amino acid level to SCO0204 (Wang *et al.*, 2009). SCO3013 is the second member of a four-gene operon which also includes a sensor kinase (SCO3012), a possible lipoprotein (SCO3011) and a possible translation initiation factor (SCO3014). This trio of genes is thought to form a three-component system homologous to the *mtrAB-lpqB* system in *Mycobacterium* (Hoskisson and Hutchings, 2006), of which the response regulator is essential for growth and survival (Zahrt and Deretic, 2000). SCO3328 (BdtA) is known to be repressed by BldD (Elliot *et al.*, 2001). SCO4232 is a singleton gene about which nothing is known. SCO3793 is a homologue of the uncharacterised *Bacillus* protein YjqA, whose function is not known. SCO5607 is annotated as belonging to the GntR family of regulators. It does not form an operon but is part of a cluster of prophage genes (SCO5605-SCO5620). SCO1366 is part of an operon with two ABC transporter components. The hypothetical protein SCO5783 shows similarity to TrkA and appears to form a two-gene operon with a transmembrane transporter. SCO4493 is an AsnC-family regulator which is part of a cluster containing several enzymes and may be a target for BldA regulation as it contains a TTA leucine codon.

Not all the candidates listed will be genuine NAPs: any abundant contaminants with a reasonably high proportion of lysine and arginine will be score highly in this system. For example SCO6482, SCO2911 and SCO2093 are uncharacterised proteins which may prove to be non-DNA-binding proteins. Several candidates were annotated with a function not previously associated with DNA-binding ability, including homologues of tellurium-resistance proteins such as SCO3767, which is a homologue of *Klebsiella* TerB whose solution structure demonstrated one very basic face and one very acidic face (Chiang *et al.*, 2008). It is not immediately obvious how this would bind DNA or have NAP activity so it may simply be an abundant chaperone. Also while SCO2140 was highly abundant and is annotated as a transcriptional regulator of the Lrp/AsnC family, it was found to be much shorter than the majority of these proteins. On closer inspection the protein shows homology only to the ligand-binding domain of AsnC, i.e. it is truncated, and does not contain an obvious DNA-binding domain. DNAbinder predicted that this protein lacked DNA-binding ability. Perhaps it acts via protein-protein interactions or perhaps it is not a transcriptional regulator at all.

3.13 Detection of candidate NAPs in Actinomycetes and other lineages

NAPs are often highly lineage-specific, with some being limited to a particular genus or family. Even NAPs with homologues in most lineages may be present in multiple copies which have diverged to acquire specialised features and function. We are particularly interested in NAPs which are specific to the Actinobacteria or those shared between *Streptomyces* and subclades relevant to natural product discovery, such as the Streptosporangineae (*Microbispora*, *Planomonospora*) or the Micromonosporaceae (*Salinispora*, *Actinoplanes*).

The presence and sequence conservation of 25 proteins found in previous sections to be both abundant at multiple time points and predicted to have nucleic-acid binding ability were investigated with the help of Professor Keith Chater and Govind Chandra (both John Innes Centre, Norwich) using reciprocal tBLASTn to detect orthologues of *S. coelicolor* genes in the genomes of representative species from the following groups: Corynebacterineae (*Corynebacterium*, *Gordonia*, *Rhodococcus*, *Tsukamurella*, *Segniliparus*, *Mycobacterium*, *Nocardia*), Streptosporangineae (*Streptosporangium*, *Nocardiopsis*, *Thermobifida*, *Thermomonospora*), Propionibacterineae (*Propionibacterium*, *Nocardioides*, *Kribbella*), Pseudonocardineae (*Thermobispora*, *Saccharopolyspora*, *Saccharomonospora*, *Amycolatopsis*, *Actinosynnema*), Frankineae (*Frankia*, *Nakamurella*, *Acidothermus*, *Geodermatophilus*), Catenulisporineae (*Catenulispora*), Streptomyicineae (*Kitasatospora*, *Streptomyces*), Micrococcineae (*Tropheryma*, *Leifsonia*, *Clavibacter*, *Beutenbergia*, *Cellulomonas*, *Brachybacterium*, *Kytococcus*, *Intrasporangium*, *Jonesia*, *Microbacterium*, *Arthrobacter*, *Micrococcus*, *Renibacterium*, *Rothia*, *Xylanimonas*, *Sanguibacter*), Micromonosporineae (*Micromonospora*, *Salinispora*), Kineosporiineae (*Kineococcus*), Actinomycineae (*Arcanibacterium*, *Mobiluncus*), Glycomycineae (*Stackebrandtia*), Others (*Bifidobacterium*, *Gardnerella*, *Atopobium*, *Cryptobacterium*, *Eggerthella*, *Olsenella*, *Rubrobacter*, *Conexibacter*), Enterobacteriaceae (*E. coli* only) and Bacillaceae (*B. subtilis* only). For some genera a single species was used while multiple species or accessions were used for widely-studied genera such as *Streptomyces* and *Mycobacterium*. Identity at the amino acid level, rather than at the nucleotide level, was used to allow for the greatly variable GC contents between genomes.

Proteins were classified as ubiquitously present (present in >95% of species; green), patchily distributed (present in 50 - 95% of species; P; blue), very sparsely distributed (present in <50% of species; VS; dark red) or absent (no orthologues detected; A; black). The list of species used for comparison was not exhaustive and each species within a family gave slightly different identity and coverage, therefore an estimate of “typical % identity” was made for each group where

members did not deviate from that value by more than 10%. In groups where identity varied by more than 10% between species, an approximate range is given.

The presence of multiple paralogues of HU in both *S. coelicolor* and several target genomes complicated the results somewhat, as HupA was apparently not detected in *E. coli* while HupS was detected as the orthologue of HU β .

Some proteins were ubiquitously distributed in every lineage with considerable sequence identity, for example YajQ (SCO4614), the response regulator SCO3013 and CRP (SCO3571) were highly conserved within the genus *Streptomyces* (70 - 100% ID), within multiple other Gram positive species with reasonable identity (40 - 70% ID) and also detected in *E. coli* (40% ID).

sIHF and SCO3375 were ubiquitous within Actinobacterial species but absent from *E. coli* (as expected), as were regulator BdtA (SCO3328), regulator SCO4232, TerB (SCO3767) and putative RNA-binding protein SCO5592. SCO3013 and its neighbouring sensor kinase SCO3012 are apparently orthologous to the response regulator of the two-component signalling system BaeRS, which in *E. coli* is responsible for a multi-drug-resistance phenotype (although the genes under this system's control are not the same between species). SCO4493 (Lrp/AsnC-family) appeared to be a genuine orthologue of the *E. coli* protein AsnC, which is a locally-acting paralogue of the NAP Lrp, while SCO2140 did not have an orthologue outside the Gram positives.

BldD (SCO1489), the putative regulator SCO1839 and the uncharacterised proteins SCO4199, SCO2093, SCO2911, SCO6482, SCO5783 and SCO5725 were ubiquitous amongst *Streptomyces* species but somewhat less well conserved within the wider group of the Actinomycetes and absent from *E. coli*. The regulators SCO1366, SCO5607 and SCO0204 and the unknown function protein YjqA (SCO3793) were very patchily distributed even within the genus *Streptomyces* and are therefore less likely to be important NAPs.

! THIS FIGURE TO BE PRINTED SEPARATELY !

(this is just an image of the spreadsheet which will be printed from Excel, with the same header and page numbering as surrounding pages.)

SCO no.	Annotated function	Streptomyces		Actinobacteria		Other Gram +ves		E. coli	
		Distribution	% ID	Distribution	% ID	Distribution	% ID	Distribution	% ID
1480	sIHF	U	100	U	50-70	P	50	A	
3767	TerB	P	80-90	VS	50-70	A		A	
4614	YajQ	U	100	U	70	P	50-70	Present	40
2140	Putative regulator (Lrp/AsnC family)	U	80-90	P	50-60	P	40	A	
2950	HupA	U	100	P	50	P	40	A*	
5725	UF	U	80	P	50-90	VS	40	A	
1489	Putative DNA binding protein BldD	U	100	P	70	A		A	
1839	Putative transcriptional regulator	U	100	P	50-80	A		A	
5592	UF	U	100	U	70	U	40	A	
3793	Yjqa	P	90	P	40	VS	40	A	
5783	UF	U	50-80	P	40	VS	30	A	
3013	Putative two-component response regulator	U	100	U	80	P	60	Present	40
4232	Putative transcriptional factor regulator	U	100	U	80	U	30-60	A	
3328	UF	U	100	U	40-80	P	40	A	
6482	UF	U	80	P	50	A		A	
2911	Uncharacterised	U	80	P	60-70	VS	30-40	Present	40
2093	Uncharacterised	U	80-90	P	40-50	VS	40	A	
4199	Lysine-rich UF	U	60-100	P	30	VS	30	A	
3571	Putative transcriptional regulator CRP	U	90	U	60	U	40	Present	30
204	Putative regulator (luxR family)	P	80	P	60	A		A	
3375	Lsr2 protein	U	90	U	50	A		A	
5607	Putative transcriptional regulator	VS	50	VS	40	A		A	
5556	HupS	U	90	P	50	VS	40	Present *	40
1366	PadR regulator	P	90	P	40-50	P	40	Present	40
4493	Putative regulator (Lrp/AsnC family)	U	90	VS	50	A		Present	30

Figure 3.12 Distributions of the most promising 25 putative NAPs identified in the previous section among different lineages of bacteria. Distribution is classified as ubiquitous (U; green), patchy (P; blue), very sparse (VS; dark red) or absent (A; black). Typical identity (%) is given at the amino acid level.

3.14 Post-translational modifications

In addition to novel NAPs, we hoped to discover novel post-translational modifications in this dataset. However very few modifications were identified when the data from the LC-MS/MS experiments were re-searched to take into account lysine acetylation, serine/threonine phosphorylation or arginine methylation. None of the modifications found were on the known or candidate NAPs. However it was very possible that modifications were present but were insufficiently abundant or stable under these conditions to be detected with significance.

In order to simplify the mixture of nucleoid proteins and thereby increase the number of peptides per protein, in an effort to detect rare species, another LC-MS/MS experiment was performed using much finer gel slices. Nucleoids were prepared in the same way as in section 3.7 from mycelium grown to late stationary phase (96 hours), when acetylation was expected to be greatest based on experiments carried out in *E. coli* (section 1.4.1). These were resolved on a 16% SDS-PAGE gel (RunBlue; Expedeon) and cut into fine slices representing 0.25 – 1 kDa per slice. Slices expected to contain the major known NAPs (HupA, sIHF) or the lysine-rich stationary phase protein (HupS, SCO4199) were submitted for LC-MS/MS using the same method as before. In this experiment methionine oxidation, N-terminal acetylation and ϵ -lysine acetylation were the only variable modifications used in the Mascot search. In one slice the coverage of sIHF was 49% but no acetylated peptides were observed. However there was no coverage of the first 33 amino acids forming the lysine/arginine-rich N-terminal region (Figure 3.13).

MALPPLTPEQRAAALEKAAAARRERA EVKNRLK**HSGASLHEVIKQGQENDVIG**
KMKVSALLESLPGVGKVR AKQIMER**LGISESR**RVR**GLGSNQIASLER**EFGSTGS

Figure 3.13 Coverage of sIHF (SCO1480) seen in an LC-MS/MS experiments on a thin (0.5 - 1 kDa) gel slice cut from an SDS-PAGE gel of *S. coelicolor* nucleoids. Matched peptides are shown in bold red.

The lack of observed modifications in these experiments does not mean the modifications are not present. Proteomics experiments designed to search for novel modifications generally use a modification-specific enrichment step following trypsin digest, for example acetylated peptides can be enriched using anti-acetyllysine antibodies (Guan *et al.*, 2010; Yu *et al.*, 2008) while phosphopeptides can be enriched using metal oxide affinity chromatography (Parker *et al.*, 2010; Thingholm *et al.*, 2006). It would be beneficial to repeat this experiment with such an enrichment step, or in the presence of HDAC inhibitors such as sodium butyrate to prevent enzymatic removal of acetylation marks during processing.

3.15 Discussion

In this chapter whole nucleoid isolation was used in combination with LC-MS/MS as a discovery tool for identifying candidate novel NAPs/global regulators. While independent confirmation of these candidates was not possible due to time constraints, the known NAPs sIHF, HupA and Lsr2 (SCO3375) all scored highly in this system, as did the global regulators BldD and CRP, which suggests that the approach was successful. A number of proteins not currently identified as NAPs/global regulators also scored highly and would be excellent candidates for investigation. Candidates for early vegetative phase NAPs/ global regulators included the response regulators SCO3013 and SCO0204 as well as a homologue of the poorly-characterised *E. coli* protein YajQ. Candidates for novel NAPs/ global regulators important during late vegetative phase included SCO5592, the hypothetical protein SCO5783 and the truncated Lrp/AsnC-family protein SCO2140. Candidates for novel stationary phase NAPs/ global regulators included the helix-turn-helix regulator SCO1839, the putative DNA-binding protein SCO0596, the uncharacterised protein YjqA (SCO3793) and the lysine-rich uncharacterised protein SCO4199. Two further proteins, SCO3767 (TerB) and SCO5725 (WGX-100 family), were highly abundant and predicted to have DNA-binding ability but belong to classes not known for DNA-binding ability so may be contaminants.

Not all of the homologues of known NAPs were found to be abundant here. For example none of the three paralogues of Dps were seen to be the dominant NAP during stationary phase in vegetative tissue. They are known to be very important in genome segregation in later development (Facey *et al.*, 2009) so perhaps over the course of evolution they have become restricted to the developmental tissues (aerial hyphae, prespore compartments, spores) and their protective functions in vegetative tissue are either obsolete or carried out by the lysine-rich proteins HupS and SCO4199. Also the H-NS-like protein SCO4076 was only detected at one time point (late vegetative), suggesting that it has a specialised role which is distinct from SCO3375.

Some of the proteins identified as likely NAPs were from unexpected classes (SCO3767 and SCO5725) and may turn out to be false positives which are simply contaminants containing many lysine residues. For example SCO3767 TerB is a homologue of *Klebsiella* TerB whose NMR structure (PDB accession 2jxu) revealed one very basic face and one very acidic face. It is not obvious how this would bind DNA or have NAP activity so it may be some sort of abundant molecular chaperone. It is somewhat reminiscent of the RNA-binding proteins shown in Pruijn (2005), albeit smaller and lacking an obvious central pore. The presence of so many abundant tellurium-resistance proteins in the nucleoids was particularly intriguing as none have been described as NAPs and little is known about them. In Enterobacteriaceae the tellurium-resistance phenotype is conveyed by the same R plasmids as a phage-inhibition phenotype (Taylor and Summers, 1979) but it is not known whether this is mediated by a DNA-binding protein. SCO2140 is similarly a mystery as it has no obvious DNA-binding domain, despite belonging to a well-known class of transcription factors.

One protein that was unexpectedly detected was HupS, the paralogue of HU with a lysine-rich tail. This has previously been described as a spore-specific protein (Salerno *et al.*, 2009) based on non-detection of a translational fusion, but it was detected here in moderate quantities. We speculate that HupS may be produced only in stressed, pellet-form mycelium and not in the individual hyphae previously observed. Another possibility is that only a truncated version is produced in vegetative tissues (in which case the fluorophore of their translational fusion would naturally be absent): no peptides from the lysine-rich tail were detected at any time point and the protein did not run at the expected size on 1D gels.

The list of candidate NAPs presented here is not intended to be an exhaustive catalogue of the contents of the *Streptomyces* nucleoid. More importantly, these results may not be an accurate reflection of the nucleoid composition in other tissues of fully-differentiating cultures grown on solid media. There are also likely to be other NAPs induced by specific conditions not tested here, as is the case with YgiP from *E. coli* which is only seen under anaerobic conditions. However, there are a number of candidates worthy of further investigation.

It will be exciting to find out whether these are genuine novel NAPs and, if so, what their roles are. Perhaps some of them perform a Fis-like role or respond to ligand binding to help *Streptomyces* respond to sets of environmental cues in the soil which contrast with those faced by *E. coli*. To establish these as NAPs would require the following information: evidence of association with the nucleoid *in vivo* (for example by translational fusion to a fluorophore),

evidence of a large regulon (~5% of the genome) and evidence that the protein modulates chromatin structure and/or nucleoid compaction.

We were somewhat disappointed not to discover more evidence of modification. However most proteome-wide surveys of post-translational modification use a modification enrichment step following tryptic digest, which we did not. This work is ongoing in our lab.

Of the two approaches used, LC-MS/MS was clearly superior to the gel-based method, probably because it was able to handle the complexity and unusual migration patterns of NAPs far better. The method could be applied to other species for which a well-annotated genome is available. The mixture of proteins detected in these nucleoids was broadly similar to those observed in similar nucleoid preparations obtained from other bacterial species (Ohniwa *et al.*, 2011), being highly enriched in NAPs and ribosomal proteins but also containing cytosolic enzymes. The nucleoids in these experiments contained much less contamination from outer membrane proteins than seen in *E. coli* nucleoids (Kim *et al.*, 2004), perhaps because *S. coelicolor* nucleoids which have not separated cleanly from the cell envelope are trapped within the mycelium while nucleoids from unicellular species like *E. coli* which are associated with remnants of cell envelope are still able to sediment normally.

One of the main limitations of this method was the very low yield of nucleoids recovered relative to species such as *E. coli*, most likely to be a result of the manner in which this species grows in liquid culture. *S. coelicolor* grows as mycelia with multinucleate compartments which may hamper the release of nucleoids, either because the nucleoids become entangled with each other or the cell wall. Furthermore, the mycelium aggregates to form large (~2 mm) pellets which may restrict the action of the lysozyme to only the surface of each pellet resulting in large amounts of non-lysed material. Changes in the composition of the media which caused increased clumping also resulting a total loss of nucleoid yield.

Summary:

- 1) A protocol for isolating nucleoids from *S. coelicolor* mycelium was developed which can be used for other applications in the McArthur lab.
- 2) The 12 kDa “S protein” reported by Sarfert *et al.* (1983) was revealed to be a complex mixture of proteins rather than a single dominant species.
- 3) A set of candidate NAPs was obtained from the LC-MS/MS data which included the known NAPs sIHF, HupA and Lsr2 plus the global regulators BldD and CRP. Several other intriguing novel candidates were detected, including several highly abundant helix-turn-helix regulators and a lysine-rich protein (SCO4199) from an unknown protein class which was induced in stationary phase.
- 4) Several highly abundant proteins of unknown function were found which could represent missing parts of the sRNA-handling machinery, including SCO5592 (discussed further in chapter 6).

Chapter 4 – $\Delta hupA$ and $\Delta hupS$ mutants of *S. coelicolor*

4.1 Introduction

Streptomyces species possess two paralogues of HU: HupA resembles *E. coli* HU α/β while HupS has a long lysine-rich tail which contributes to its strong affinity for DNA (Kumar *et al.*, 2010) and its ability to protect the DNA from harmful Fenton chemistry (Takatsuka *et al.*, 2011). It has been reported that HupA is a major component of the nucleoid during vegetative growth while expression of HupS is induced in sporogenic aerial hyphae and particularly spores, where it has a role in spore nucleoid compaction, based on microscopical studies using translational fusions to *egfp* (Salerno *et al.*, 2009). This is reflected in the reported phenotypes of their knockout mutants: a $\Delta hupA$ mutant generated in *S. lividans* grew more slowly in liquid medium (Yokoyama *et al.*, 2001) while a $\Delta hupS$ mutant generated in *S. coelicolor* had heat-sensitive, weakly-pigmented spores (Salerno *et al.*, 2009). However it was also previously reported that the HupS protein was upregulated in the $\Delta hupA$ mutant (Yokoyama *et al.*, 2001), perhaps suggesting that there may be some degree of redundancy between the two. In chapter 3 HupA was detected in great abundance at all time points in nucleoids isolated from liquid cultures and, surprisingly, HupS was also detected in moderate quantities during stationary phase. *S. coelicolor* does not sporulate in liquid medium so its role here was unclear. While a small number of altered traits have been described for $\Delta hupA$ and $\Delta hupS$ mutants, their phenotypes have not yet been fully characterised in any species of *Streptomyces* and a $\Delta hupA$ strain of *S. coelicolor* has never been described. We wished to discover whether these mutants had any phenotypic changes which could potentially be of use for manipulating secondary metabolism, and what potential drawbacks may exist with this approach. This chapter explores the phenotypic effects of genetically deleting HupA and HupS from *S. coelicolor*, with particular respect to growth rate, development and antibiotic production. It also investigates the distribution of the different forms throughout the Actinobacteria.

Aims:

- 1) Generate a $\Delta hupA$ knockout in M145, complement it *in trans* and assess its phenotype.
- 2) Investigate the phenotype of a $\Delta hupS$ knockout.
- 3) Investigate the phylogenetic distribution of HU homologues within the eubacteria.
- 4) Use the information generated in 1) – 3) to infer their respective roles.

4.2 Strains

HU mutants were created using the REDIRECT[®] protocol developed at the John Innes Centre (manual available from <http://streptomyces.org.uk/redirect/index.html>, also see section 2.8). Briefly, the gene of interest was replaced in its entirety with a cassette containing an apramycin resistance gene by PCR-targeted mutagenesis. Replacement was confirmed by PCR with multiple primer pairs (section 2.8.5, fig 2.5), confirming that they were double-crossover recombinants. All mutant lines were created in the strain M145, which is a plasmid-cured “wild-type”.

Two independent clones of the $\Delta hupA$ strain ($\Delta hupA.V$ and $\Delta hupA.Y$) and two independent clones of the $\Delta hupS$ strain ($\Delta hupS.2$ and $\Delta hupS.3$) were tested and found to behave similarly, therefore only one clone of each ($\Delta hupA.V$ and $\Delta hupS.2$) are shown in this chapter.

Complementation of the $\Delta hupA$ strain (B1001) was carried out using plasmid pIJ11404 which contained the *hupA* gene (*sco2950*) under its native promoter in the vector pMS82, creating the complemented strain B1002. Two clones of $\Delta hupA.V$ harbouring the complementation plasmid (B1002.2 and B1002.21) were tested and found to behave similarly, therefore only one (B1002.2) is shown in this chapter.

Unfortunately it was not possible to construct a plasmid with which to complement the $\Delta hupS$ strain. Repeated attempts to clone the gene into the vectors pIJ86 or pIJ10257 (both under the strong constitutive promoter *ermE**p), pIJ8600 (under the thiostrepton-inducible promoter *tipAp*), and pMS82 (under its native promoter) were unsuccessful in *E. coli*. However it was easily cloned into very stringent or non-expressing vectors such as pGEM-T Easy (Promega) and pET15b (Novagen). Attempts are currently being made to replicate the complementation strategy used in Salerno *et al.* (2009), which used a much larger complementation fragment spanning multiple genes.

Wild-type (M145) strains harbouring empty versions of plasmid pMS82 were found to grow largely as M145, but produced somewhat more actinorhodin on R5+ and DNA media. These are not shown here.

In order to test whether over-expression of HupS was toxic to *E. coli*, plasmid pET15b containing either the non-toxic putative RNA-binding protein *sco5592* (see chapter 6) or *hupS* (*sco5556*) was freshly transformed into *E. coli* BL21(DE3) pLysS and an equal number of cells were plated onto L agar with 100 µg/ml carbenicillin in the presence or absence of 0.1 mM IPTG. With IPTG induction, colonies producing SCO5592 were smaller but just as numerous as without production, indicating that the cells tolerated the presence of the protein despite the large amount of energy being diverted into its production. With IPTG induction, cells producing HupS failed to produce colonies (Figure 4.1). It is not known whether HupS is toxic in a conventional sense or whether as a dormancy-related protein it simply slows growth to a great extent, perhaps by binding to sites in *E. coli* which alter expression of genes essential for growth. A similar effect was seen on the growth of *E. coli* and *M. smegmatis* when the *M. tuberculosis* orthologue of HupS (Mdp1) was over-expressed (Matsumoto *et al.*, 2000).

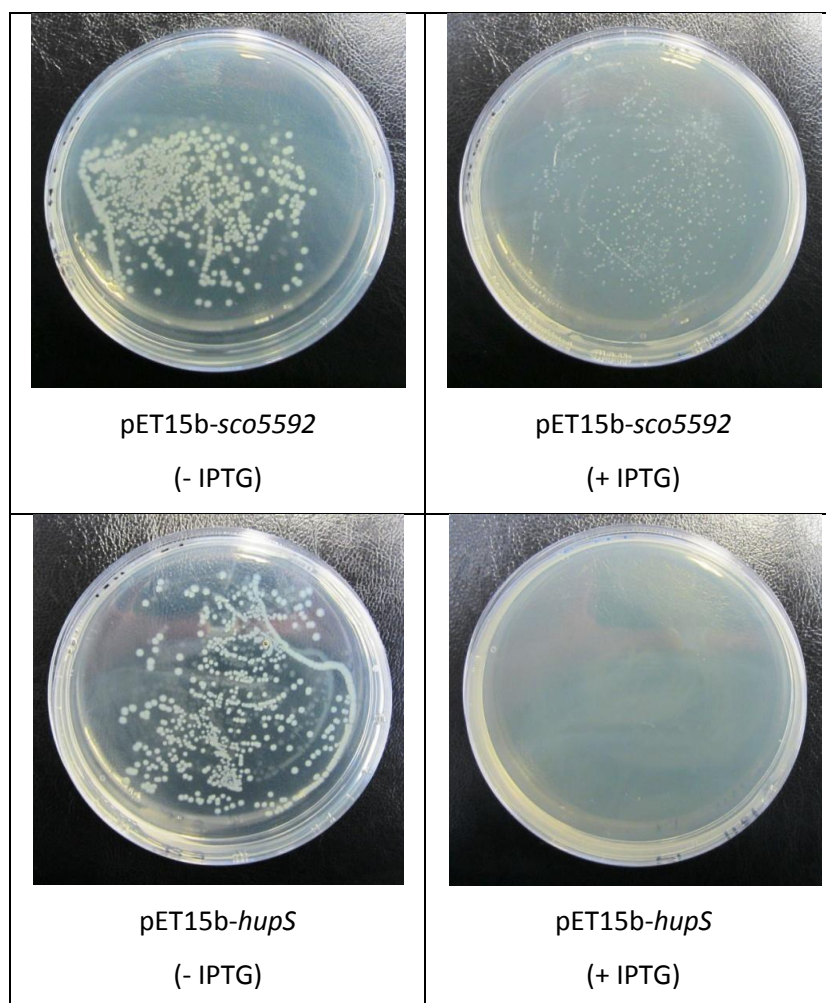


Figure 4.1 Growth of *E. coli* BL21(DE3) pLysS harbouring pET15b-*sco5592* or pET15b-*hupS*, with and without IPTG induction. Over-expression of the putative RNA-binding protein SCO5592 was tolerated while over-expression of HupS prevented colony formation, indicating toxicity.

4.3 Phenotypes screened

NAP mutants are generally highly pleiotropic so $\Delta hupA$ and $\Delta hupS$ were expected to show multiple phenotypic changes. As an initial screen they were grown on a selection of standard media (recipes are given in section 2.2) which were selected to highlight changes in different processes. SFM medium was selected for studying changes in development as *S. coelicolor* rapidly and robustly completes its life cycle on this medium. R5+ medium was selected to show changes in secondary metabolism as this medium was developed to maximise production and was expected to clearly show deficiencies or early over-production. DNA medium was selected because it gives rapid growth with a lesser degree of differentiation, to more clearly show increased levels of sporulation and antibiotic production. Plates were inoculated by spreading a fixed quantity of spores onto the agar surface with a wooden toothpick, then using fresh toothpicks to streak from this lawn to give single colonies. Plates were incubated at 30°C and photographed over 10 days to monitor phenotypic differences. The size and appearance of mature single colonies was investigated by streaking out a lesser quantity of spores (to give 5 – 15 well-separated colonies per SFM plate) and incubating for two weeks at 30°C.

Growth curves were carried out as described in section 2.3.6, with growth rate and antibiotic production being measured spectrophotometrically. Time points were selected to demonstrate maximum growth rate, timing of transition to stationary phase and final density. The mean of three cultures is shown in each case, with error bars representing the standard error of the mean. For clarity each mutant is plotted separately with the same M145 data for comparison on each graph. The rich liquid medium TSB:YEME34% was selected because it encourages growth in a reproducible and dispersed manner with distinct colour changes.

Several more features were measured to further investigate changes observed in the basic phenotypic assessment. *S. coelicolor* is a mycelial bacterium which grows as pellets in liquid culture, typically 0.1 – 0.5 mm in diameter, therefore changes in growth rate revealed by measuring optical density may be caused either by fewer pellets, by smaller pellets or by some change in their form. Pellet morphology was examined without staining by light microscopy under a 10x or a 40x objective on a Nikon Eclipse E800. Cryo-SEM was used to examine the gross morphology and ultrastructure of spores from week-old colonies, with the help of Kim Findlay. Spore stress tolerance was assessed by heat-shocking equal aliquots of dilute spore suspension at 60°C in a water bath for increasing amounts of time, then plating them onto SFM and counting the reduction in CFUs relative to untreated aliquots. Spores were used in these experiments because mycelium is extremely heterogeneous and cannot be accurately titred.

4.4 Summary of phenotypes observed in comparison to wild-type

ΔhupA

- Actinorhodin overproduction on SFM and DNA
- Increased sporulation on DNA
- Slightly delayed development on R5+
- 10-fold reduction in CFUs/ml of spore preps and low germination rate
- Tighter mycelial pellets which clump severely in TSB:YEME34%
- Reduced production of actinorhodin and undecylprodigiosin in TSB:YEME34%

ΔhupS

- Loss of actinorhodin production on SFM and DNA and in TSB:YEME34%
- Sensitivity of actinorhodin production on R5+ to the presence of proline and NaOH
- Low final density and loose pellet morphology in TSB:YEME34%
- Delayed spore pigment production on SFM
- Mixture of normal spore chains, enlarged spores and thick, undifferentiated aerial hyphae
- Spore sensitivity to heat stress (Salerno *et al.*, 2009)

4.5 Phenotype of $\Delta hupA$ on agar media

On SFM (sporulation medium), $\Delta hupA$ produced actinorhodin earlier and in greater quantities than M145 (Figure 4.2). This was particularly apparent on days 2 and 3 but was still visible by day 6. Development was otherwise normal, with aerial hyphae and grey spore pigment appearing at the same time as for M145. This phenotype was only partially complemented as the quantity of actinorhodin produced by the complemented strain was intermediate between that of M145 and $\Delta hupA$. It would be beneficial to construct additional complementation strains from different independent clones of the mutant strain to determine whether the partial complementation was a unique feature of this clone.

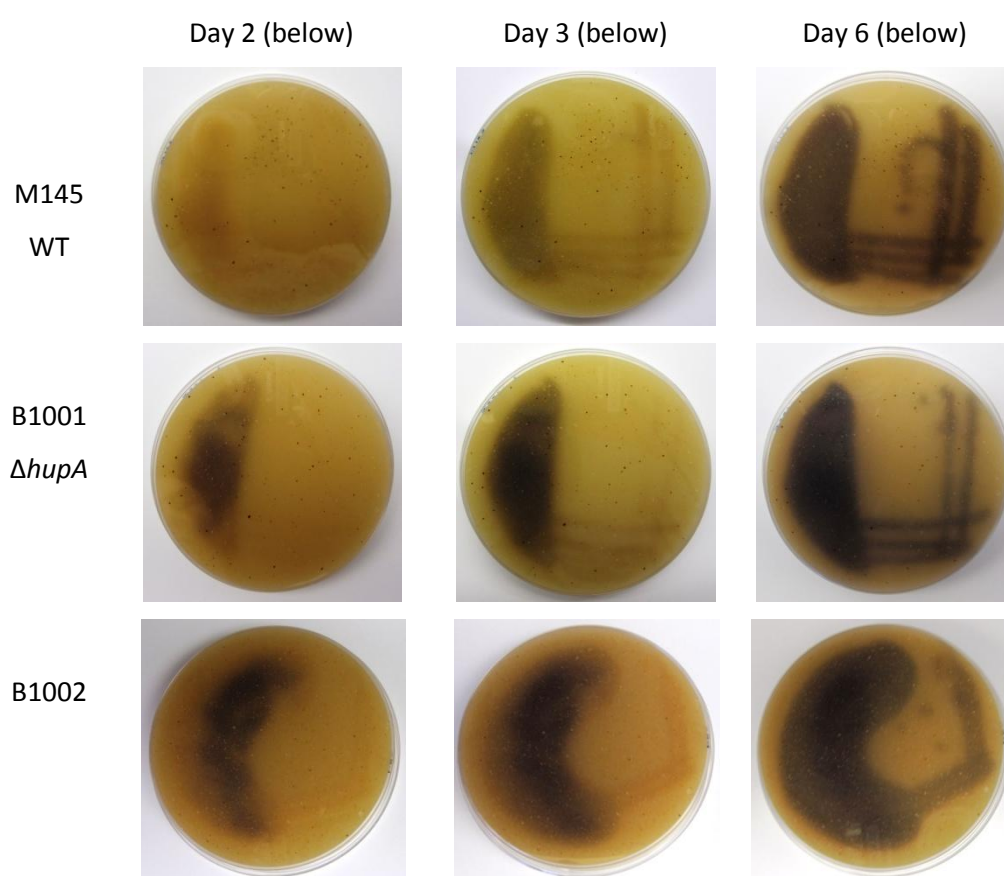


Figure 4.2 Early and increased production of actinorhodin by $\Delta hupA$ grown on SFM medium (sporulation medium). Plates photographed from below. Strain B1002 is a $\Delta hupA$ strain harbouring the complementation plasmid pMS82-HupA.

On R5+ (antibiotic production medium), $\Delta hupA$ developed slightly later than M145 (Figure 4.3). On day 1 M145 was already reddish and patches of aerial mycelium were visible. In contrast, $\Delta hupA$ was still unpigmented and had not begun to differentiate. By day 3 both had sporulated and were secreting actinorhodin but $\Delta hupA$ produced less, particularly from low-density areas. By day 4 both were secreting actinorhodin into the medium in comparable quantities. These phenotypes could be complemented.

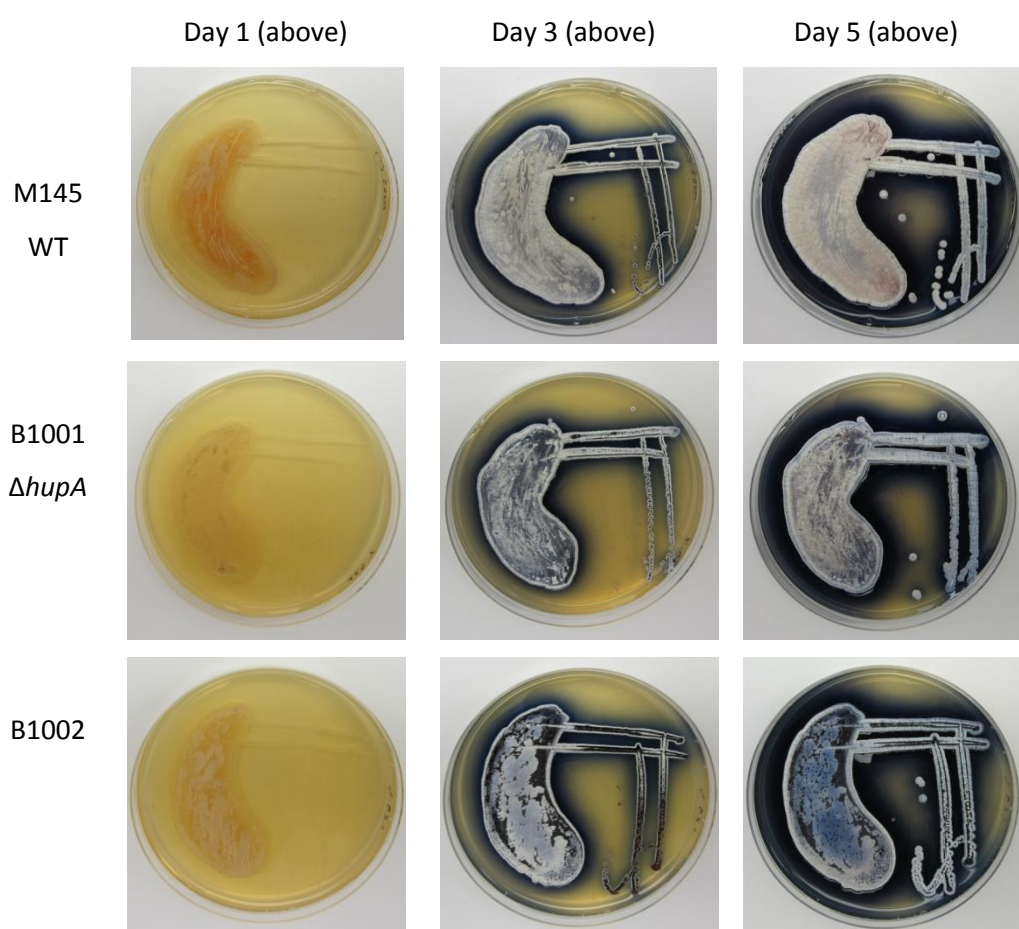


Figure 4.3 Mild developmental delay and actinorhodin reduction of $\Delta hupA$ on R5+ medium (antibiotic production medium). Strain B1002 is a $\Delta hupA$ strain harbouring the complementation plasmid pMS82-HupA.

On DNA (rapid growth medium), $\Delta hupA$ produced large quantities of actinorhodin much earlier than M145 (Figure 4.4). By day 2 the high-density regions had already turned dark blue and by day 3 actinorhodin was being secreted into the medium, while M145 had only turned blue around the tips of the lawn area. Also $\Delta hupA$ had differentiated more completely, with grey spores being visible across the colony surface by day 3 while at the same time point M145 had developed aerial mycelium but there was no evidence of sporulation. The sporulation phenotype was complemented but the actinorhodin phenotype was only partially restored to wild-type: the complemented strain secreted less actinorhodin into the medium than the mutant but the lawn area was still dark blue.

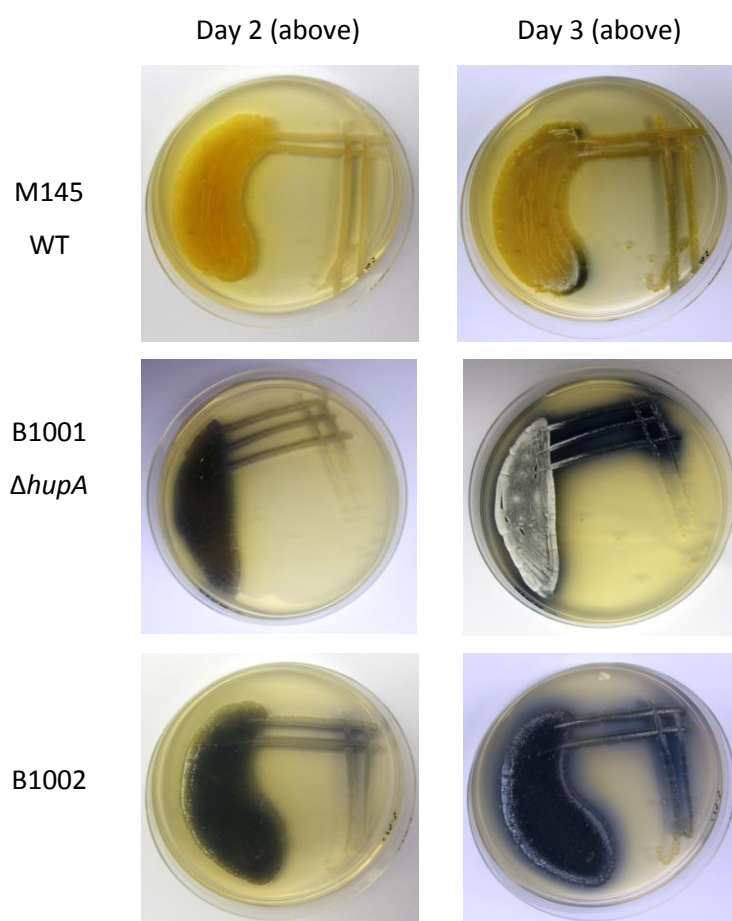


Figure 4.4 Over-production of actinorhodin and greater extent of sporulation of $\Delta hupA$ on DNA medium (rapid growth medium). Strain B1002 is a $\Delta hupA$ strain harbouring the complementation plasmid pMS82-HupA.

4.6 Phenotype of $\Delta hupA$ in liquid media

Growth curves were carried out in 50 ml rich liquid media (TSB:YEME34%) as described in section 2.3.6. Cells were cultured over 6 days at 30°C with shaking. Flasks were siliconised with dimethyldichlorosilane and baffled with a spring in order to encourage dispersed growth. Flasks were inoculated with an equal number of CFUs (approximately 1×10^7), giving an initially higher OD for the poorly-germinating $\Delta hupA$ strain (see section 4.7).

$\Delta hupA$ growth in liquid could not be recorded satisfactorily over the full course of the growth curve as changes in dispersal prevented later readings from being taken. After a variable number of hours growing at a normal rate, the majority of the biomass would stick to the sides of the flasks in a solid ring, with a corresponding sharp fall in OD₄₅₀ of the liquid (Figure 4.5). The timing of this clumping was variable and could happen at any time point between 12 and 100 hours post inoculation (HPI) but, once begun, was completed rapidly (within a few hours). In contrast, wild-type mycelium growth was well-dispersed and any small amounts of matter that got stuck to the side were easily washed back into the medium by gentle swirling of the flask. The effect was present but less severe in unsiliconised flasks, with most of the biomass forming large irregular clumps which did not adhere to the flask walls so strongly.

In Figure 4.5 the mean growth rate of three M145 flasks is compared against the individual growth rates of three flasks of $\Delta hupA$ ($\Delta hupA$ -A, $\Delta hupA$ -B, and $\Delta hupA$ -C). At 17.5 HPI all six flasks were similar. However by 25 HPI the mycelium of flask $\Delta hupA$ -A had clumped and the OD₄₅₀ decreased. The same process was seen in flasks $\Delta hupA$ -B and $\Delta hupA$ -C just after 100 HPI and 75 HPI respectively. A second set of three $\Delta hupA$ (not shown) flasks behaved in the same way, with clumping occurring between 24 and 72 hours.

Antibiotic production was measured for flask $\Delta hupA$ -2 until 120 hours. Levels of both undecylprodigiosin and actinorhodin were severely reduced relative to M145 (Figure 4.6), even prior to clumping. However these measurements were taken from a single flask and may not accurately represent the phenotype of this mutant strain.

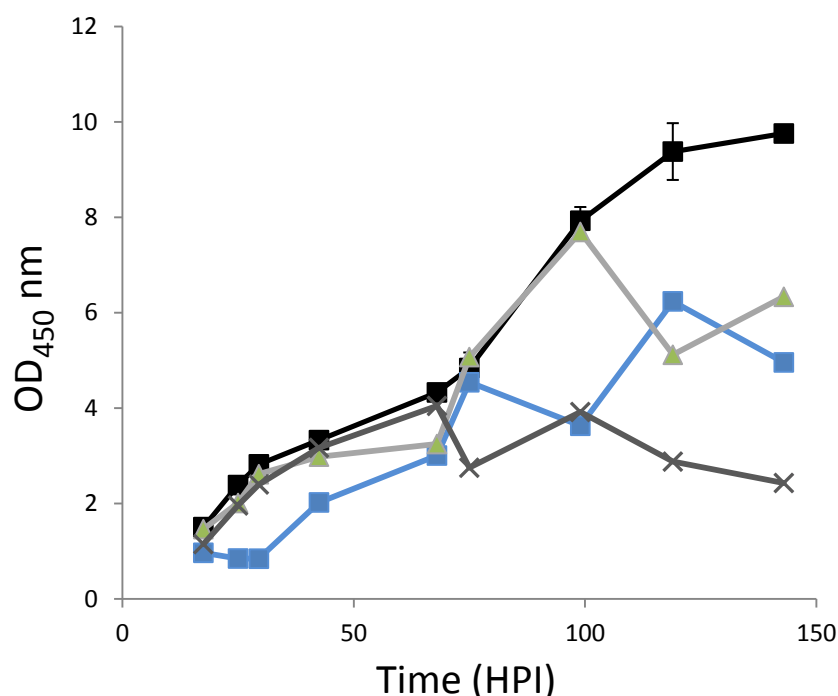


Figure 4.5 Liquid growth curves of M145 (WT) and $\Delta hupA$ in TSB:YEME34%. M145 (black squares) is shown as a single curve, with error bars representing the standard error of the mean for three flasks. Three identical flasks of $\Delta hupA$ are shown as separate curves: $\Delta hupA$ -1 (blue squares); $\Delta hupA$ -2 (light grey triangles); $\Delta hupA$ -3 (dark grey crosses).

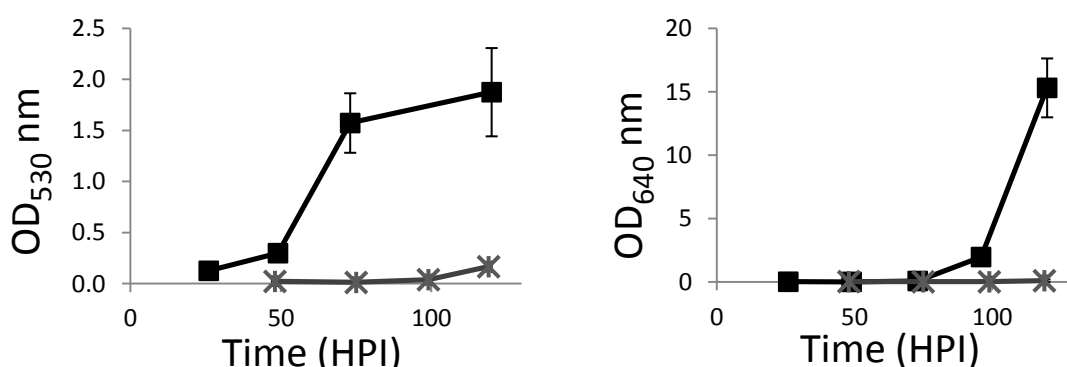


Figure 4.6 Production of undecylprodigiosin (left) and actinorhodin (right) in TSB:YEME34% of M145 and $\Delta hupA$. M145 (black squares) error bars represent the standard error of the mean of three flasks while $\Delta hupA$ (grey asterisks) represents measurements from a single flask ($\Delta hupA$ -2 in Figure 4.5).

The degree of dispersal during liquid growth is known to have strong effects on antibiotic production (Jonsbu *et al.*, 2002), possibly due to quorum sensing effects or changes in the microenvironment of parts of the pellet. For example oxygen and nutrient diffusion to the centre of the pellet would become limited as the size increased. Therefore the appearance of the mycelial pellets prior to clumping was examined by light microscopy. Pellets were resuspended in

water and applied directly to microscope slides without fixation or staining, then examined immediately under a Nikon Eclipse E800 light microscope under a 10x or 40x objective. Approximately 25-30 fields of view were examined for each genotype, with representative images shown here.

M145 pellets were fairly heterogeneous in size and shape. They appeared brown-beige with a soft outline during early vegetative phase and deep red with a lightly lobed surface during late vegetative / early stationary phase. $\Delta hupA$ pellets were of a comparable size but showed some subtle differences. During early vegetative phase the surface appeared firmer, with a sharper outline than M145. During late vegetative phase $\Delta hupA$ was paler pink, corresponding with the reduction of undecylprodigiosin production mentioned previously, and again had a firmer surface with less obvious lobes (Figure 4.7).

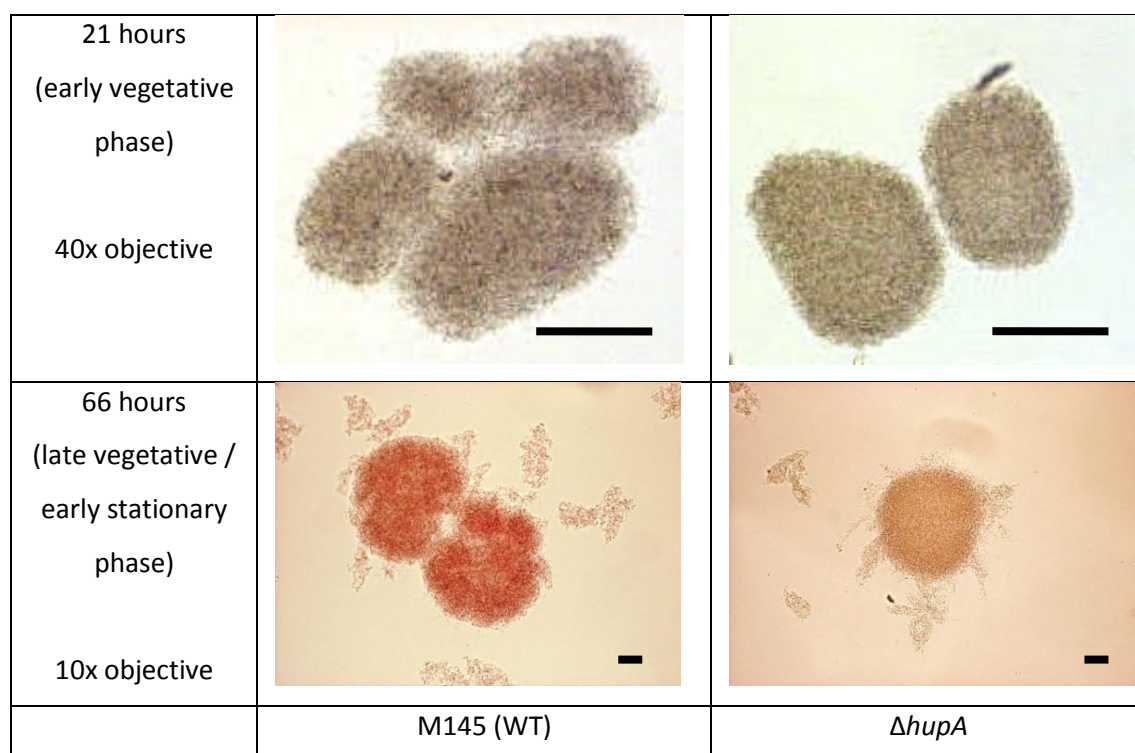


Figure 4.7 Appearance of mycelial pellets of M145 and $\Delta hupA$ grown in liquid culture during early vegetative and late vegetative / early stationary growth phases. Scale bar = 500 μm .

The strong clumping and altered pellet morphology suggest that the adhesive properties of the surface of $\Delta hupA$ pellets may have changed. In *S. coelicolor* static liquid cultures, surface attachment is mediated by hydrophobic fimbriae composed of fibrils of chaplin proteins which are joined to the hyphae by cellulose (de Jong *et al.*, 2009). If excessive fimbriae are being produced by $\Delta hupA$ this could explain the increased severity of the clumping phenotype in siliconised flasks, which become more hydrophobic during the treatment.

4.7 Germination deficiency of $\Delta hupA$

During strain construction it became apparent that spore preps of $\Delta hupA$ gave remarkably low titres when germinated on SFM plates, with only around 10% as many CFUs/ml than M145 (Figure 4.8), an effect which was not seen for other mutant strains. Typically when spore preps are prepared by resuspending a measured wet weight of spores in a proportional volume of 10% glycerol they give a very consistent titre so this change was thought to be a part of the mutant phenotype, particularly as spore preps of the complemented strain (B1002) had a titre similar to that of wild-type.

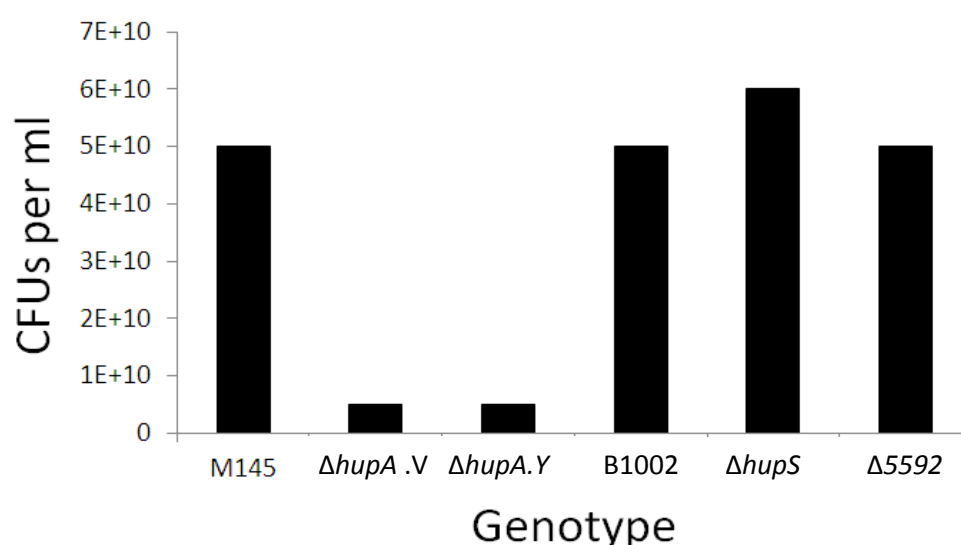


Figure 4.8 $\Delta hupA$ spore preps were found to have a low number of CFUs/ml compared to M145 and two unrelated mutant genotypes. Two independent clones of $\Delta hupA$ are shown here ($\Delta hupA.V$ and $\Delta hupA.Y$) alongside one clone of the complemented $\Delta hupA$ strain (B1002).

Images obtained from Cryo-SEM visualization of samples from 7-day-old colonies grown on SFM showed that spore morphology was indistinguishable from M145 (not shown) so it seemed unlikely that the lower titre could be explained by a defect in spore assembly. $\Delta hupA$ spores were also found to be no more susceptible to heat stress than M145 spores (see section 5.11) so a defect in general robustness of the spores was not indicated. Additionally the density and consistency of the spore preps appeared normal and had been filtered through sterile cotton wool so it seemed unlikely that non-spore debris was present in the preps which could account for the discrepancy. HupA was previously shown to be expressed primarily during vegetative growth (Salerno *et al.*, 2009), therefore the difference was suspected to occur during germination or very early growth. Dormant *S. coelicolor* spores are refractive; upon germination in the presence of calcium this refractility is lost with a characteristic timing as the spores darken and swell, which can be measured as a reduction in OD₆₀₀ of around 10 – 20% (Eaton and Ensign, 1980; Hirsch and Ensign, 1976). Therefore in order to determine whether the timing of $\Delta hupA$

germination was normal, a loss-of-refractility assay was carried out. Dense spore stocks of M145 and $\Delta hupA$ were each diluted 1/20 into 10 ml TX buffer (a minimal medium) which had been pre-warmed to 55°C in water bath. The suspension was heat shocked at 55°C for 10 minutes and then rapidly cooled under running water. Spores were harvested by centrifugation at 13,000 rpm in a micro centrifuge and resuspended in 10 ml TX buffer supplemented with 0.5 mM CaCl_2 and 15 mM glucose. M145 in TX buffer supplemented with 15 mM glucose alone was used as a negative control as these conditions inhibit germination. These suspensions were incubated at 37°C with shaking over 160 hours and the OD_{600} was measured every 10 minutes.

The OD_{600} of M145 germinated in the presence of calcium reduced by approximately 15% over the first 50 minutes, at which point the rate of change levelled off. The OD_{600} of $\Delta hupA$ germinated in the presence of calcium also decreased during the first 50 minutes, but the magnitude of this change was smaller and only around 8% of the optical density was lost. The OD_{600} of M145 germinated in the absence of calcium did not change over the course of the experiment, showing that the changes seen for the spores germinated in the presence of calcium were indeed due to germination and not simply to loss of suspended matter by flocculation of spores.

These results suggest that a proportion of the $\Delta hupA$ spores present are germinating normally while the rest are not germinating at all, with the failure occurring very early in the life cycle (prior to germ tube outgrowth). However this result is preliminary as the experiment was repeated only once and the complemented strain remains to be tested.

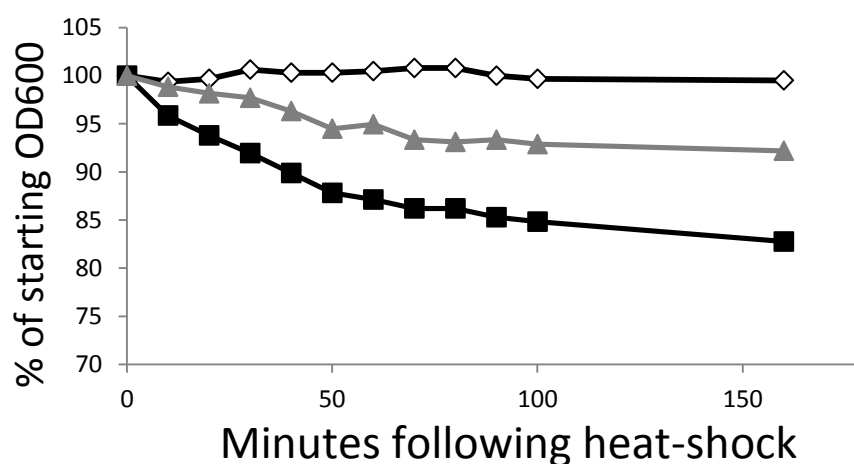


Figure 4.9 A loss-of-refractility assay showed that $\Delta hupA$ spores germinated with similar timing to M145 but with a smaller magnitude of change in optical density. M145 (+ CaCl_2): closed black squares; M145 (- CaCl_2): open black diamonds; $\Delta hupA$ (+ CaCl_2): closed grey triangles.

4.8 Phenotype of $\Delta hupS$ on agar media

On SFM, $\Delta hupS$ initially produced paler spores than M145 (as was previously reported by Salerno *et al.*, 2009) but by day 13 the spores produced by single colonies were significantly darker (Figure 4.10). By day 13 M145 had begun to produce blue guttates on its upper surface while $\Delta hupS$ did not. Also the $\Delta hupS$ single colonies had a more ragged border on day 4 than those of M145 which may have been due to sectoring of colonies.

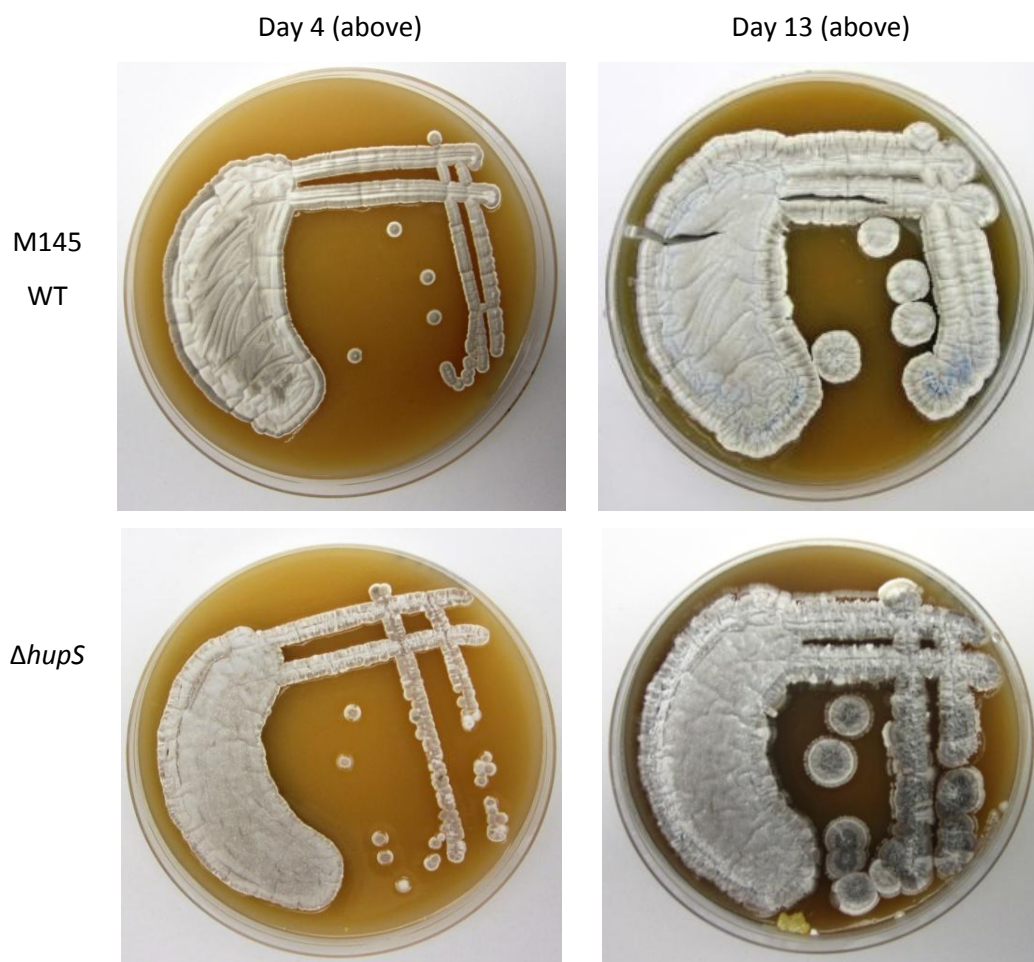


Figure 4.10 The $\Delta hupS$ strain took longer than M145 to produce spore dark spore pigment, but eventually produced darker single colonies than M145.

On R5+ plates, $\Delta hupS$ was marginally paler than M145 on day 1 (Figure 4.11). By day 3 M145 was fully covered with aerial mycelium, some of which had formed dark spores, while $\Delta hupS$ had produced a far less extensive covering of aerial mycelium.

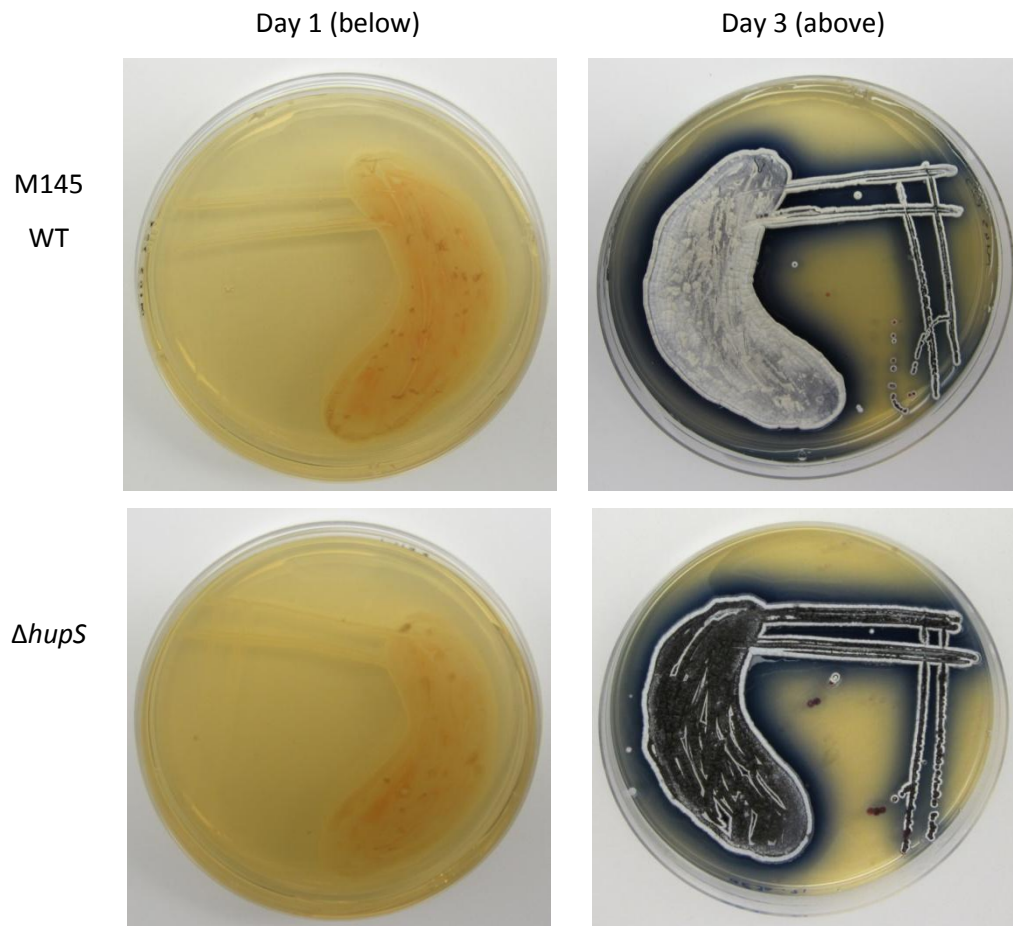


Figure 4.11 Delayed aerial mycelium formation of the $\Delta hupS$ strain on R5+ medium.

On DNA plates, *ΔhupS* appeared similar to M145 on day 2 (Figure 4.12). On day 6 some of the low-density regions and the edges of the confluent lawn of M145 had formed aerial mycelium and were secreting small quantities of actinorhodin while *ΔhupS* had not secreted any actinorhodin.

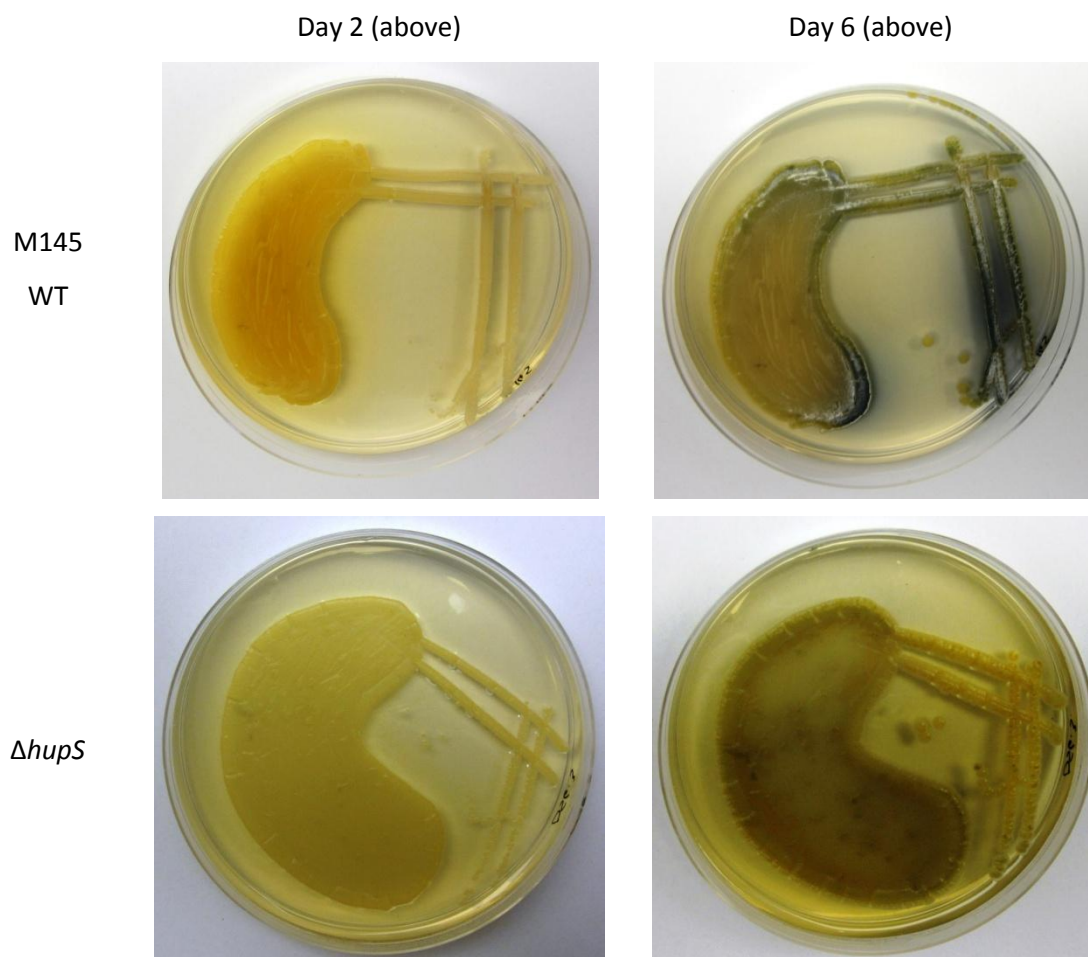


Figure 4.12 Reduced actinorhodin secretion of the *ΔhupS* strain on DNA medium.

4.9 Phenotype of $\Delta hupS$ in liquid media

A growth curve was carried out as described in sections 5.6 and 2.3. $\Delta hupS$ grew initially at a similar rate to M145 but entered stationary phase prematurely at 50 HPI, as compared with around 120 HPI seen for M145, giving a low final optical density of around half that of M145 (Figure 4.13).

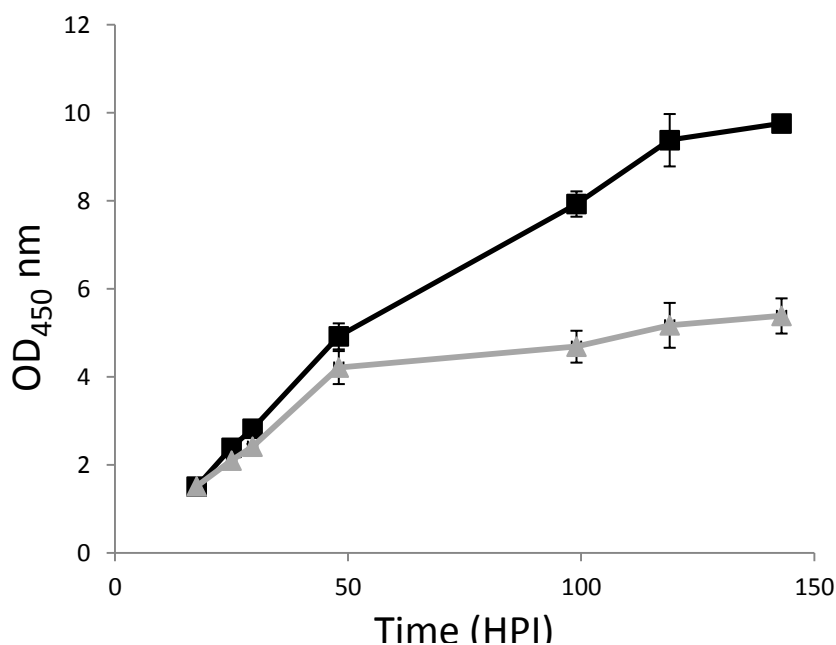


Figure 4.13 Liquid growth curves of M145 (black squares) and $\Delta hupS$ (grey triangles) in TSB:YEME34%. Error bars represent the standard error of the mean for three flasks.

Antibiotic production was also significantly disrupted. Undecylprodigiosin production was delayed by around 50 hours but eventually rose to approximately wild-type levels (Figure 4.14, left). Actinorhodin was not produced in detectable quantities at any time point (Figure 4.14, right).

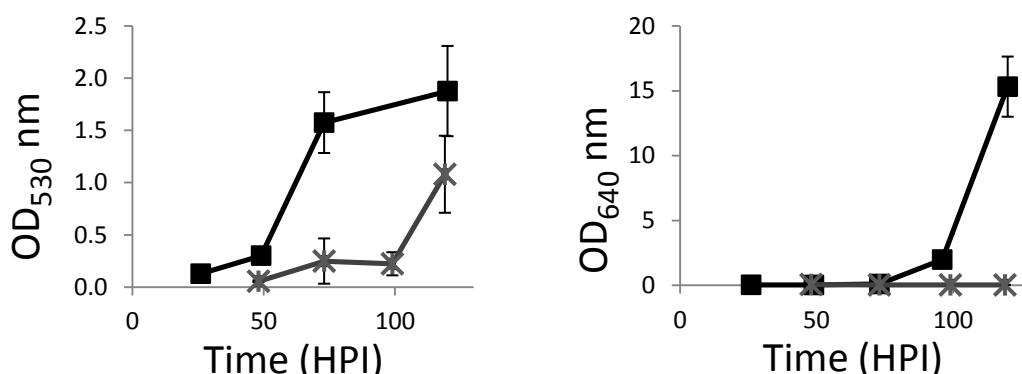


Figure 4.14 Production of undecylprodigiosin (left) and actinorhodin (right) of M145 (black squares) and $\Delta hupS$ (grey triangles) in TSB:YEME34%. Error bars represent the standard error of the mean for three flasks.

Pellet morphology of $\Delta hupS$ was examined as in section 4.6. At 21 hours the $\Delta hupS$ pellets were of normal size but were structurally defective, with a “loosely-packed” appearance and multiple deep fissures crossing each pellet. At 66 hours the $\Delta hupS$ pellets were a more rounded shape than M145 and varied greatly in size. There was also more debris, suggesting a lack of cohesion. The largest of these pellets had begun to turn pink in the centre but were less strongly pigmented than those of M145 (Figure 4.15).

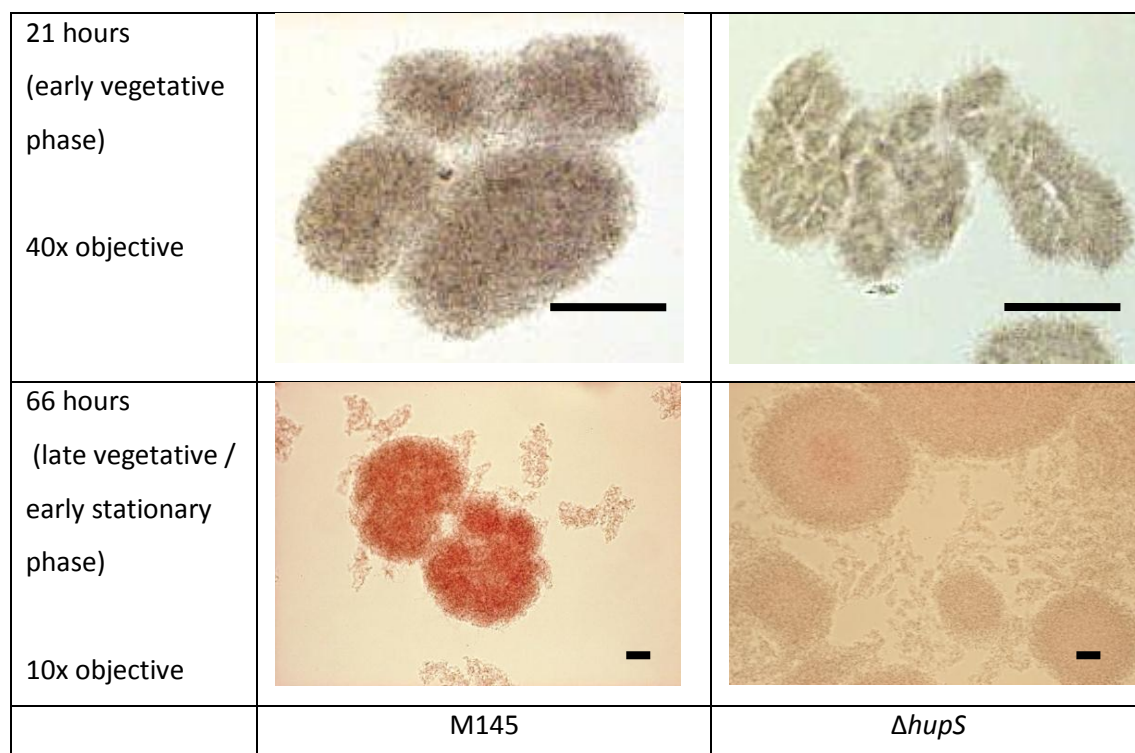


Figure 4.15 Appearance of mycelial pellets of M145 and $\Delta hupS$ grown in liquid culture during early vegetative and late vegetative / early stationary growth phases. Scale bar = 500 μm .

The low structural integrity seen in pellets of $\Delta hupS$ was the opposite of the “firmer” phenotype of the $\Delta hupA$ mutant. It is tempting to speculate that the role of HupS in vegetative mycelium is to positively regulate surface hydrophobicity (e.g. formation of fimbriae) as this could account for both the loss of pellet integrity in a $\Delta hupS$ mutant and the increased adhesiveness of the $\Delta hupA$ mutant, assuming that HupS is upregulated in a $\Delta hupA$ background in *S. coelicolor* as it is in *S. lividans* (Yokoyama *et al.*, 2001). However fimbriae have only been demonstrated to form in differentiating mycelium produced in standing liquid cultures (de Jong *et al.*, 2009) and they are not currently known to have a role in the interior portions of mycelial pellets produced in shaken liquid cultures. It is not yet known whether the unusual pellet morphology also accounts for the reduction in final optical density, but this could be addressed by measuring the total mass of protein per ml of culture for M145 and $\Delta hupS$.

4.10 Effects of R5+ additives on $\Delta hupS$ phenotype

R5+ medium contains a range of additives designed to maximise antibiotic production, which were found to be even more crucial for production in the $\Delta hupS$ mutant than in wild-type: on R5- (R5 without any additives), $\Delta hupS$ secretes absolutely no actinorhodin at any time point and appears dark red, while M145 secretes a small amount of actinorhodin (not shown). In order to examine the relative contributions of each additive, a series of plates were made which each lacked one additive. Photographs were taken from below on day 3. In the absence of CaCl_2 or KH_2PO_4 , both genotypes produced some actinorhodin but less than they did on R5+, with similar amounts produced by each when regions of equal plating density are compared. Without proline, M145 produced very little actinorhodin while $\Delta hupS$ produced absolutely none. Without NaOH, M145 still secreted actinorhodin from all densities while $\Delta hupS$ secreted only a little from densely-plated regions and none at all from single colonies (Figure 4.16). The particularly strong effect of NaOH suggests this phenotype may be pH-dependent, perhaps due to a failure to reabsorb the organic acids secreted during growth (Madden *et al.*, 1996; Viollier *et al.*, 2001), therefore it would be beneficial to repeat this experiment and measure the pH of the medium surrounding the plated region. [zx add Charles thompson references]


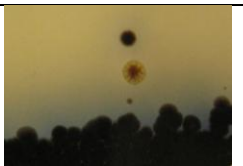
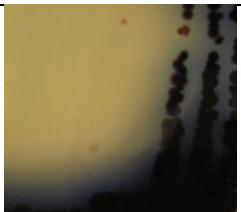
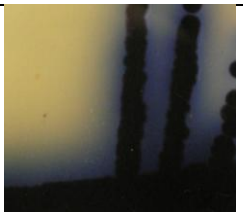
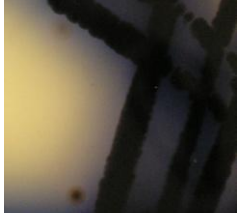


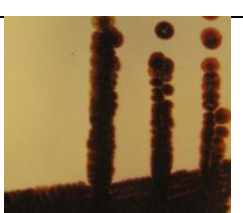
M145	$\Delta hupS$	Additives
		All but KH_2PO_4
		All but CaCl_2
		All but NaOH
		All but proline

Figure 4.16 Actinorhodin secretion of $\Delta hupS$ is more sensitive than M145 to the presence of the R5+ additives proline and NaOH. Plates were photographed from below on day 3.

4.11 Spore morphology of $\Delta hupS$

HupS has a key role in forming compact spore nucleoids which is reflected in the reduced heat tolerance, increased nucleoid diameter and delayed pigmentation of $\Delta hupS$ spores (Salerno *et al.*, 2009). To determine whether the mutant strain showed changes in spore morphology, M145 and $\Delta hupS$ were grown on SFM for seven days to form a confluent lawn before being examined by scanning electron microscopy (SEM) at a magnification of 12,000x (see section 2.12.2) by Kim Findlay. Around 6-10 fields of view were observed for each genotype and representative images are shown here.

After seven days, the majority of M145 aerial hyphae had fully differentiated into spore chains where spores were well-separated and of similar size (Figure 4.17, above, white arrows). Only a very small proportion of spores were significantly larger or smaller than average (Figure 4.17, above, red arrow) which would indicate abnormal events during the septation process.

At the same time point, seven days, $\Delta hupS$ showed a mild defect in spore chain formation. While normal spore chains were abundant, many hyphae were abnormally thick and non-septated (Figure 4.17, below, black arrows). Also a greater proportion of spore chains than WT contained individual enlarged spores (Figure 4.17, below, red arrows). It is interesting that the majority of hyphae are either almost entirely normal (Figure 4.17, below, white arrows) or entirely abnormal, reflecting the high degree of co-ordination during the segregation of the large rod-like prespore nucleoids into condensed single-genome spore nucleoids. It would be beneficial to quantify the proportion of abnormally-sized spores for each genotype by measuring the sizes of a large number of spores, however it may be more difficult to estimate the proportion of thickened hyphae. Zxrephrase?

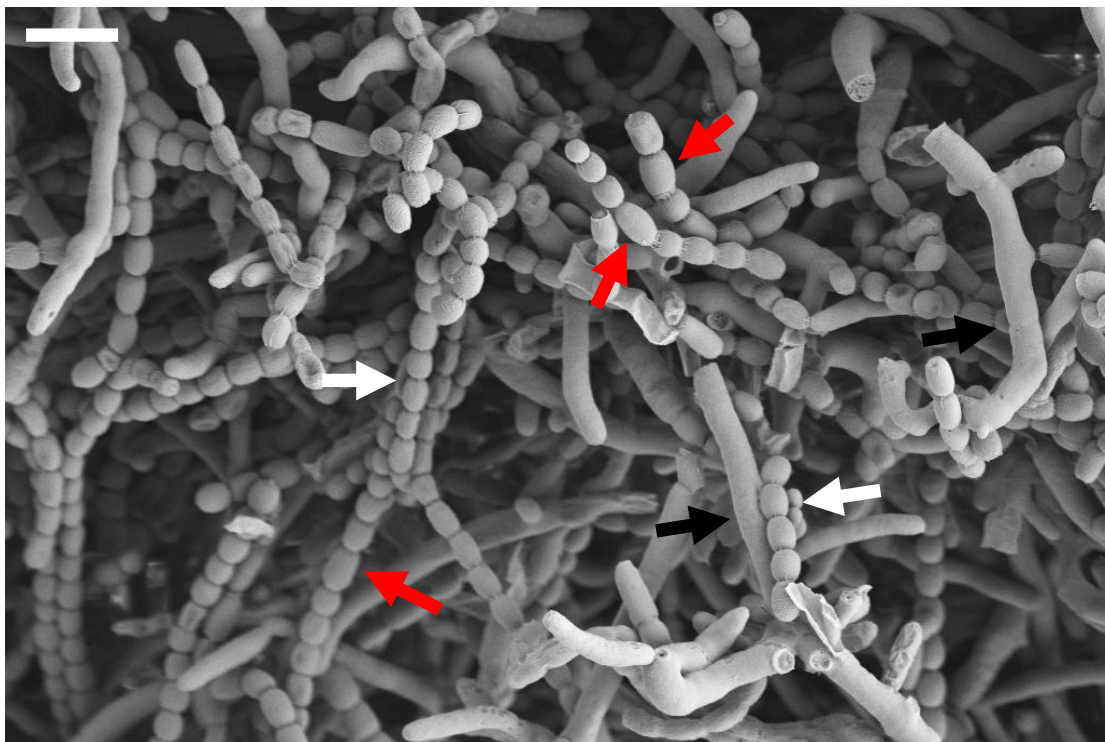
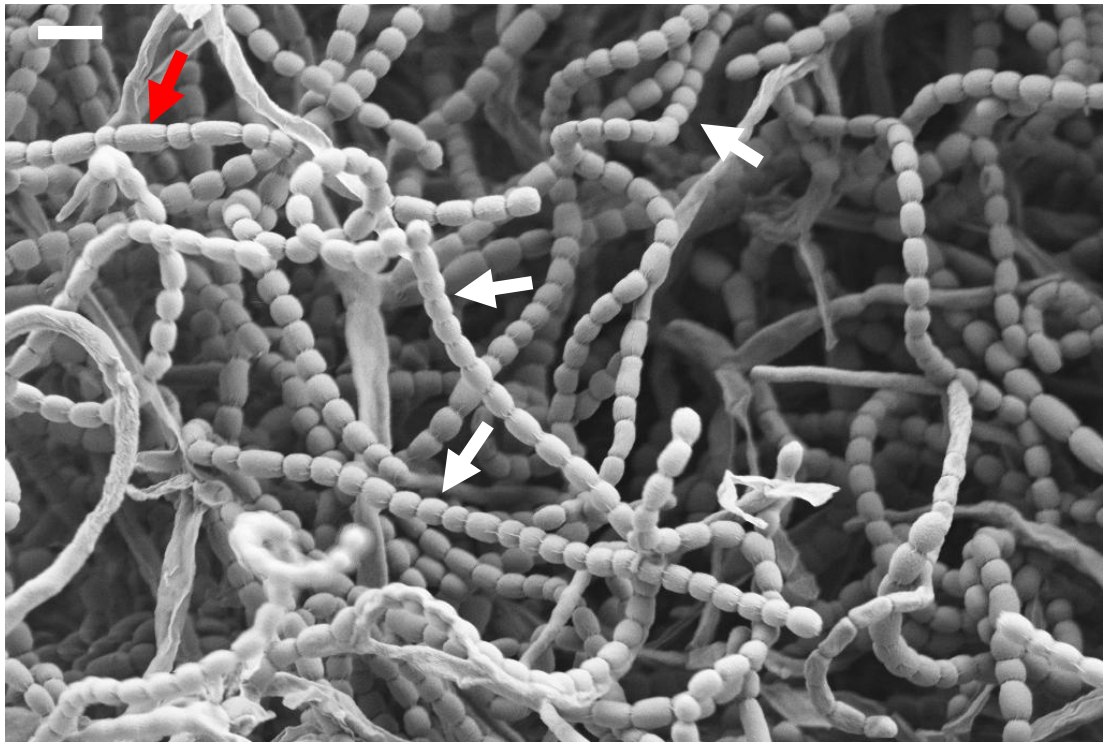


Figure 4.17 Scanning electron micrograph of M145 (above) and $\Delta hupS$ (below) grown on SFM for 7 days at 30°C, showing a mixture of normal spore chains (white arrows), individual enlarged spore (red arrows) and thickened aerial hyphae which have failed to septate (black arrows). Scale bar = 2 μ M.

4.12 *ΔhupA* and *ΔhupS* mature colony morphology

Both HU mutants showed changes in the morphology of liquid-grown pellets, therefore in order to visualise more clearly possible changes in agar-media growth, M145 and both mutants were plated onto SFM at a low density such that each plate contained 16-17 well-spaced colonies. These were incubated at 30°C and photographed on days 8 and 11. *ΔhupA* colonies appeared largely normal, but were marginally smaller than M145 (Figure 4.19, left) and often possessed a prominent central pore below which a split in the agar frequently formed (Figure 4.18, middle). *ΔhupS* colonies were larger and flatter (Figure 4.18, right; Figure 4.19).

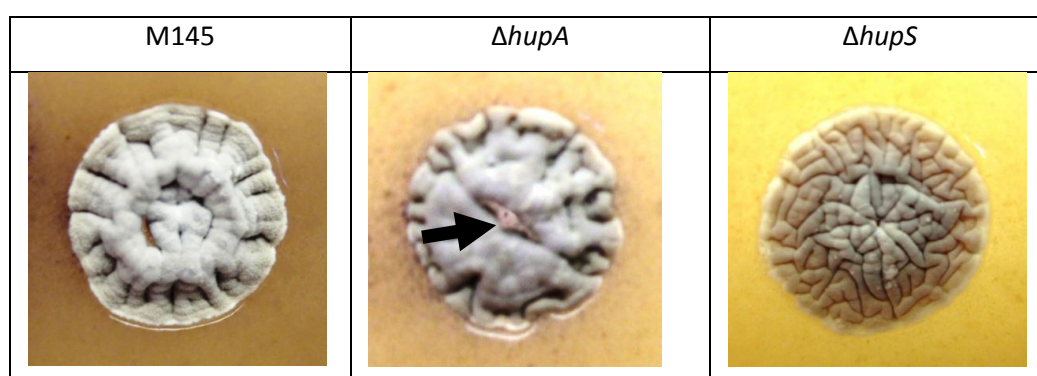


Figure 4.18 Colony morphology of M145 (left), *ΔhupA* (middle) and *ΔhupS* (right) following two weeks of incubation. Views from above (above). The pore of *ΔhupA* is marked with an arrow.

The sizes of well-separated individual colonies were then measured in ImageJ (available from <http://rsbweb.nih.gov/ij/>) from photographs by comparison to the diameter of the petri dish (85 mm). Error bars represent the standard error of the mean, based on 16 - 17 colonies per genotype. On day 11, *ΔhupA* colonies (7.7 mm, SEM = 0.1 mm) were slightly narrower in diameter than M145 (8.2 mm, SEM = 0.2 mm) and *ΔhupS* colonies were considerably larger (11.4 mm, SEM = 0.3 mm).

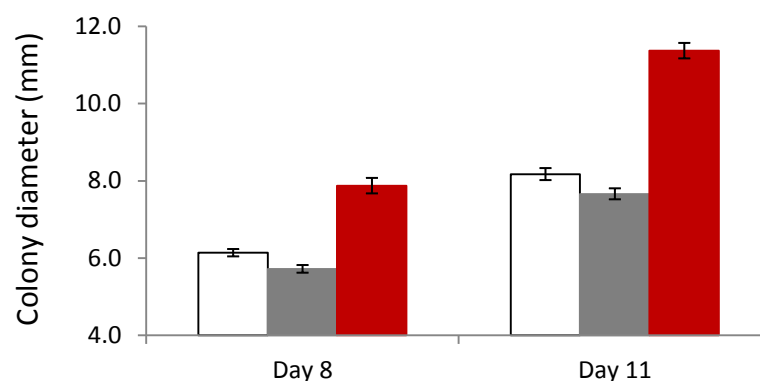


Figure 4.19 Average colony diameters of M145 (white), *ΔhupA* (grey) and *ΔhupS* (red) strains grown on SFM medium for 8 and 11 days. Error bars represent the standard error of the mean for 16-17 single colonies separated by at least 1.5 cm of bare agar. Zxrephrase?

4.13 Phylogenetic analysis of HU distribution among the Actinomycetes

It has previously been reported that the number of paralogues per species is highly variable in this lineage: *Streptomyces*, *Janibacter* and *Salinispora tropica* each have two copies (one short and one long); *Mycobacterium*, *Rhodococcus* and *Salinispora arenicola* each have one long copy; *Micrococcus*, *Thermobifida* and *Bifidobacterium* each have one short copy; *Kineococcus radiodurans* has five copies (one short and four long); *Corynebacterium* lacks HU entirely (Grove, 2011; Salerno *et al.*, 2009). In order to address questions about the evolution of HU within the Actinomycetes, a neighbour-joining phylogenetic tree was assembled from the protein sequences of a large number of Actinomycetes, phages and bacteria from other lineages.

Questions to be answered included: 1) Did the lysine-rich tail arise once or multiple times? 2) At what point in Actinomycete evolution did the lysine-rich tail arise? 3) Did it arise by fusion of HupA to an existing lysine-rich domain? 4) Is there any evidence of horizontal gene transfer (HGT)?

Amino acid sequences were retrieved from NCBI by using the amino acid sequence of *S. coelicolor* HupA (SCO2950; SB1002) as a tBLASTn query against individual complete genomes. This meant that all HU sequences belonging to each species were identified without duplication. Search results were clear-cut, with HU homologues having high (>40%) identity over the majority of the full length of the HU domain and no other plausible hits being returned. HU sequences from non-Actinomycetes were represented by sequences from the Gram negative bacteria *E. coli* and *Legionella pneumophila*, from the basal Gram positive species *Heliobacterium* and *Megasphaera*, and from the Firmicutes *Bacillus*, *Staphylococcus*, *Enterococcus*, *Lactobacillus* and *Erysipelothrix*. IHFa/b from *E. coli* and *L. pneumophila* were selected as a related but likely divergent group as these are homologues of HU which have acquired a novel function. Additionally, 13 HU homologues from bacteriophages were identified by searching the NCBI protein database with the terms “IHF phage”. It was not possible to include an outgroup in this tree because HU is ubiquitous among bacteria, i.e. the tree drawn from this set of species cannot be rooted because the point at which HU first originated is not known. However it was expected that the gene tree drawn from vertically-inherited HU sequences would have a similar topology to a species tree.

To exclude the lysine-rich tails (which were not present in all proteins) from the alignment, sequences were trimmed back to 95 amino acids. Some were slightly less than this length to begin with due to small indels throughout the genes. An alignment and neighbour-joining tree were generated in ClustalW2 and the tree was displayed using Interactive Tree of Life (Letunic and Bork, 2007, 2011) in unrooted mode (Figure 4.21).

SB1002 and ScHupS were found to belong to very distinct clades; the distance between them was as great as that between HupA and Gram negative HU or IHF sequences. This suggests that SB1002 and ScHupS diverged a significant time ago, perhaps followed by rapid evolution of ScHupS after it acquired a new function. Surprisingly not all of the sequences in the HupS-like cluster had lysine-rich tails, suggesting that HupS was under positive selection which caused its sequence to diverge from HupA before acquisition of the tail. The ones which did have tails formed almost a single clade within the tailless versions so it seems likely that the tail arose once. It is therefore not ideal to refer to short HUs as “HupA-like” and long HUs as “HupS-like”. They will be referred to here as HupA-like, tailless HupS-like and tailed HupS-like to reflect the evolution of the core HU domain.

Two intra-generic radiations of tailed HupS-like HUs were seen: the four sequences from *K. radiodurans* and the three from *Amycolatopsis mediterranei* each formed tight clades, showing that they had diverged only recently. As its name suggests, *K. radiodurans* was isolated from a radioactive waste site and is highly resistant to γ -radiation. The extra copies of HupS are likely to contribute to its stress tolerance but may also be developmentally specialized since it produces motile zoospores (Bagwell *et al.*, 2008). The benefit of extra copies of HupS to *A. mediterranei* is less apparent as this is a soil-dwelling species.

Actinosynnema mirum was found to have two short versions of HU, one of which was HupS-like (Actinosynnema-1; 66% ID to Mdp1) and one of which clustered with HUs from Gram negative species and Firmicutes (Actinosynnema-2; 33% ID to Mdp1).

Figure 4.20 Unrooted neighbour-joining tree of HU amino acid sequences from Actinomycetes and representatives from bacteriophages, Gram negative bacteria and Firmicutes.

[attached separately as a pdf because this will be an A3 fold-out page in the thesis]

Figure 4.20 Unrooted neighbour-joining tree of HU amino acid sequences from Actinomycetes and representatives from bacteriophages, Gram negative bacteria and Firmicutes.

The gene tree generated in this chapter did not generate strong evidence for widespread horizontal gene transfer of HupS. However one example may have been identified: *Rhodobacter* phage RcapMu was found to encode an HU which was particularly similar to the tailless HupS sequences of *S. roseum* and *S. tropica* (Figure 4.21). These species both also encode a second copy of HupS, but the other copy from *S. tropica* has a lysine-rich tail while the other from *S. roseum* does not. The other species of *Salinispora* (*S. arenicola*) only has the tailed version so perhaps the additional copy in *S. tropica* was acquired by HGT. *S. arenicola* is geographically very widespread while *S. tropica* is restricted to the Bahamas and co-exists with *S. arenicola*, suggesting that it has acquired a novel ecological niche to which its second copy of HupS contributes.

This work also identified the first example of a tailed HupS sequence outside of the Actinobacteria. *Yersinia* phage phiR1-37 encoded one HU with a tail with the sequence “KKAACKAAPAKAAAPAAKKVVKKGKK” (Figure 4.22). This protein is shorter than the tailed HupS sequences of Actinobacteria, but significantly longer than non-tailed versions, and would be an excellent candidate for biochemical experiments investigating the effect of different tails on HU function.

Mdp1 of *M. tuberculosis* has an asparagine in the second position from the N-terminus which is vital for its growth-slowing properties (Matsumoto *et al.*, 2000). The majority of the tailed HupS homologues surveyed here had an asparagine at this position with the exception of *A. mediterranei* 1-3, *S. tropica*-2 and Yersiniaphage-1.

It has previously been suggested that the lysine-rich tail was acquired by HupS by a gene fusion event (Salerno *et al.*, 2009). However it was not possible to identify a possible donor for the lysine-rich tail by using it as a BLAST or tBLASTn query due to the extreme variability of the lysine-rich sequences even on orthologous proteins within the same genus. The lengths of the tails also varied greatly (Figure 4.22), with *A. mediterranei* HU3 and *Janibacter* spp. HU2 having particularly long tails (170aa and 141aa respectively) and the *Yersinia* phage ϕ R1-37 HU having a particularly short tail (26aa). *S. coelicolor* HupS had a tail of intermediate size (117aa). HUs from the Firmicutes often had a region of low sequence complexity of around 20 residues at the C-terminus with sequences such as “PATKAPAFKPAKALKDAVKAK” (*B. subtilis*, YonN) or “PASKVPAFKAGKALKDAVK” (*S. epidermidis*). These had a similar amino acid composition to the HupS tail but the repeat was of a shorter unit. Some Actinomycete species had a short region rich in lysine and alanine at their C-terminus, for example *A. marinum* and *S. roseum*, which could represent a “proto-tail” comprising a single unit of the repeat (Figure 4.22). Taken together, these features suggest rapid evolution by a mechanism such as strand slippage or unequal

recombination or gene conversion. Perhaps the same mechanism acts on SCO4199 and HrdB which have a similar lysine-rich domain.

Species	Amino acid sequence
<i>Streptomyces coelicolor</i> - HupS	VNKAQLVEAIADKLGGRRQQAADAVDAVLDAVRAVVAGDRVSVTGFGSFEK VDRPARYARNPQTGERVRVKKTSVPRFRAGQGFKDLVSGSKLKPNDIAV <u>KK</u> <u>APKGSLSGPPPTISKAAGKKAAAKKATGAAKTTGAAKKTSAAAKTTAKKT</u> <u>TGAAKTTAKKTTAKKSAAKTTTAAAKTAACKAPAKKATAKKAPAKKSTAR</u> <u>KTTAKKATARKK</u>
<i>Amycolatopsis mediterranei</i> - 3	MANKAQLIEALTERLGDKKAASEAVDGLVDIIIRTVNKGEKVNITGFGVF EKRA RAARTARNPRTGEAVRVKKTNPVAFRAGTTFKDVISG <u>AKKLPKATPVKRTPTA</u> <u>TRAAATTTTRATTPRATTSRTAAAAPKPATTRSTTTTRRATAAKPATTRTPA</u> <u>AKPATTRA KAAPKTTAAKTTA AKAAPPAKAATAAKTTVAKATTA AKSAAAK</u> <u>AATAAKTTA AKAATAAKPTTARKTSAARTPAAKTTAAKTPARRTSAAAKKD</u>
<i>Yersinia</i> phage ϕ R1-37	MATVNDLLAQVQTNLKAQGV DSTKKECGVALEAVLNAIQTVALDQGSIRTAI GTFKRKECAARQAHNPKTGDKVDVPAKITVSFKSYNEVVGE <u>KKAAKKAAPA</u> <u>KAAAPAAKVVKKGKK</u>
<i>Bacillus subtilis</i> (YonN)	MNKTELIAKVAEQGVSKKEGAPSVEKVFDTISEALKSGEKVSIPGFGTFEVRER AARKGRNPQTGEEIDI <u>PATKAPAFKPAKALKDAVKAK</u>
<i>Aeromicrobium marinum</i>	MAISRDLVSALAEKAGTTKTEADAVLSALSDVLLDSVGKGEKVSIPGILSVERV SRAARTGRNPSTGETIEIPAGFGVKVSAGS <u>KLKAAAK</u>
<i>Streptosporangium roseum</i> - 2	MNKKELVDAIADRVGDKKTATEAVNAVLDAIQKAVASGDKVSITGFGAFEMV HKPARTARNPSTGAEIKVAESWGPKFRPGSDFKELVNVEG <u>KKAAKKK</u>

Figure 4.21 Representative HU amino acid sequences showing variability in lysine-rich domain length.

4.14 Discussion

Several novel phenotypic changes were identified in the $\Delta hupA$ and $\Delta hupS$ strains. In some cases the two deletions had opposite effects, for example $\Delta hupS$ was actinorhodin-deficient on most media while $\Delta hupA$ over-produced actinorhodin on SFM and DNA. They also had opposing effects on pellet morphology and mature colony size: $\Delta hupA$ gave compact mycelial pellets and smaller colonies while $\Delta hupS$ gave looser mycelial pellets and larger colonies than M145. This could possibly be due to loss of HupS in the former and upregulation of HupS in the latter to compensate for the loss of HupA as was reported in a *S. lividans* $\Delta hupA$ strain (Yokoyama *et al.*, 2001). Unfortunately it was not possible to construct a strain over-expressing HupS to confirm this as HupS appeared to be toxic. The deletions both had effects on spores but these acted at different stages of the life cycle: the $\Delta hupS$ mutant produced heat-sensitive spores and a proportion of its spore chains had failed to septate properly, while the $\Delta hupA$ mutant produced morphologically normal spores of which 90% failed to develop into colonies. In other words, it appeared that $\Delta hupS$ had trouble transitioning from mycelium to spores while $\Delta hupA$ had trouble transitioning from spores to mycelium. The exact point at which $\Delta hupA$ germination or outgrowth failed is not yet known. It is tempting to speculate that HupA in some way displaces the large amount of HupS attached to the spore nucleoids during germination, but the defect could also be due to an inability to transcribe early-growth genes or even physiological mechanisms such as problems in trehalose storage. A strain over-expressing CRP is also germination-deficient but this is likely to be via a different mechanism as its spores were much seen to be much rounder under SEM (Piette *et al.*, 2005).

Not all the phenotypes were in opposition: both strains were delayed or deficient in actinorhodin production in liquid medium and on R5+ but this is likely due to separate defects in growth which affected actinorhodin production indirectly.

Despite a high degree of conservation, HupA and HupS may have slightly different roles in different *Streptomyces* species. A $\Delta hupS$ strain was constructed in *S. coelicolor* by Salerno *et al.* (Salerno *et al.*, 2009) and by our lab, but could not be constructed in *S. lividans* (Yokoyama *et al.*, 2001). Also the type of growth defect $\Delta hupA$ showed was different between species: in *S. lividans* the mutant grew more slowly in liquid culture while in *S. coelicolor* the spores were defective at some point in germination but thereafter grew at a normal rate until clumping occurred. In both cases inoculation was with a known number of CFUs rather than to a particular OD.

The loss of actinorhodin production on most media in the $\Delta hupS$ strain was partially restored by the additives present in R5+ medium, with a strong effect being seen for proline and NaOH.

Proline is a precursor of undecylprodigiosin, which may act as a shunt for excess proline (Smith *et al.*, 1995) but it does not have a known role in actinorhodin production. NaOH could have a role in actinorhodin production via its effects on pH: at pH 4.5 - 5.5 all actinorhodin is produced mainly intracellularly but at pH 6.0 - 7.5 a portion is converted to γ -actinorhodin and secreted (Bystrykh *et al.*, 1996). It is not obvious why the $\Delta hupS$ mutant would be more sensitive to pH than M145, but responsiveness of a NAP to pH has been seen for *E. coli* H-NS: *in vitro*, the stiffening mode was predominant at pH 6.5 while the bridging mode was predominant at pH 8 (Liu *et al.*, 2010a). In general actinorhodin production is a rather ambiguous measure of phenotype because it responds to so many factors. An increase may signify that part of the pathway has been transcriptionally upregulated, or that more chemical precursors are available, or that the cells are stressed in some way. Conversely a decrease may signify that transcription has been abolished, or that secretion is blocked, or that pH is altered, or that the mutant strain is deficient in responding to one of the other R5+ additives.

While there was insufficient time during this study to examine the respective binding sites of each HU by methods such as ChIP-seq, the information generated in this chapter will be useful for guiding this type of work. Time points have been identified on several media for each mutant where its phenotype differs from the WT. However it has not yet been proven that either of these genes regulates secondary metabolism directly at the level of transcription. While the $\Delta hupA$ mutant over-expressed actinorhodin on more than one medium, it is conceivable that this could be an induction of the general stress response due to defects in DNA repair (which HU participates in) or changes in nutrient utilization.

The low germination rate of the $\Delta hupA$ mutant was particularly interesting. The fact that some spores are able to germinate normally hints at some stochastic effect, for example it is possible that signals are released by the first spores to germinate which cause the remaining spores to germinate at a faster rate. $\Delta hupA$ dormant spores may produce lesser amounts of these signals or be less responsive to them than wild-type.

The phylogenetic distribution of HU sequences within the Actinomycetes proved to be more complex than expected. Many of the short HUs were in fact much more similar at the amino acid level to the core domain of HupS than to HupA. There was an unexpectedly large divide between HupA-like and HupS-like sequences, of a similar magnitude as the difference between *S. coelicolor* HupA and *E. coli* HU α , or *E. coli* HU α and *E. coli* IHF, which suggests either extremely rapid sequence divergence or an HGT event early in Actinomycete evolution. It is also possible that more than one mechanism is in operation. The suggestion of Salerno *et al.* (Salerno *et al.*, 2009)

that both a short and a tailed version of HU existed in the first Actinomycete which were variously lost and duplicated in different lineages is highly plausible, plus there may also have been further instances of HGT to complicate the pattern. For example *S. tropica* appears to have gained a second short HU since its divergence with *S. arenicola*, possibly via a bacteriophage, which is shared by *S. roseum*. Functional NAPs are known to be carried on phages and plasmids as a means of inducing the host to accept them, for example a full-length H-NS was found encoded in a phage of “*Candidatus Accumulibacter phosphatis*” which had been acquired from its host (Skenner et al., 2011) and the Sfh protein from plasmid pSf-R27 is an H-NS-like protein which is required to prevent titration of the host H-NS pool away from its own DNA when such a large (180 kb) plasmid is taken up. The presence of a shorter lysine-rich tail on the HU from *Yersinia* phage ϕ R1-37 was particularly intriguing as this could either represent another instance of HGT or, perhaps more likely, an independent tail acquisition event. This may be an example of repeated evolution as similar but probably non-homologous lysine rich tails exist on the N-terminus of HU from *D. radiodurans* (Ghosh and Grove, 2006) and on several other DNA-binding proteins in *S. coelicolor* including SCO4199 (see figure 3.11) and HrdB.

From the results presented in the chapter it appears that the majority of phenotypic changes seen in $\Delta hupA$ and $\Delta hupS$ are opposites. This supports the notion of the pair having evolved developmental specialization, although the phenotypic changes of a mutant in liquid culture suggests that the role of HupS may not be strictly limited to differentiated tissues.

Summary:

- 1) The *S. coelicolor* chromosomal *hupA* and *hupS* genes were replaced with apramycin resistance cassettes using REDIRECT technology (PCR-targeted mutagenesis). $\Delta hupA$ was complemented in trans using plasmid pMS82 containing the gene under its own promoter but $\Delta hupS$ was not successfully complemented because its apparent toxicity interfered with construct creation.
- 2) $\Delta hupA$ had a low spore germination rate although its spores appeared morphologically normal. It produced actinorhodin precociously on two agar media but was deficient in actinorhodin production in liquid culture. It had somewhat tighter mycelial pellets than WT and was prone to severe clumping during liquid culture, suggesting that it may have been producing some sort of secreted polysaccharide.
- 3) $\Delta hupS$ had a deficiency in actinorhodin production on most media which was particularly responsive to the R5+ additives proline and NaOH. It grew to a low final density in liquid culture, produced larger colonies than M145 and had an unusually loose mycelial pellet morphology.
- 4) Phylogenetic analysis revealed that the lysine-rich tail of HupS was acquired after the divergence of the its HU domain from that of HupA, so that tailless HupS-like proteins are found. One possible instance of HGT into *S. tropica* via a phage was identified which may have contributed to its restricted geographical range. Also one instance of a *Yersinia* phage containing a partially-tailed HU was found, which most likely represents an independent tail acquisition event.

Chapter 5 – Δ *lsr2* mutants of *S. coelicolor*

5.1 Introduction

The Lsr2 proteins were selected for investigation because they are equivalent to the key NAP H-NS, which is known to repress large horizontally-acquired tracts such as pathogenicity islands in the Enterobacteriaceae (Lucchini *et al.*, 2006). It was anticipated that deletion of one or both *lsr2* genes in *S. coelicolor* could cause derepression of undecylprodigiosin and actinorhodin in a manner that would be readily applicable to secondary metabolites of other Actinomycetes. Conversely it was expected that strains over-expressing Lsr2 might show reduced antibiotic production.

S. coelicolor has two copies of *lsr2*. The molecular weights of the encoded proteins are very similar (11.6 kDa and 11.7 kDa) but the pI of SCO4076 (9.6) is notably higher than that of SCO3375 (6.5), suggesting that they may have different properties. Very little is known about either protein, only that SCO3375 was found to have one phosphorylated threonine (Parker *et al.*, 2010) and may be a target of a eukaryotic-like serine/threonine kinase. In chapter 3, SCO3375 was found to be abundant at all time points in liquid culture while SCO4076 was not highly abundant at any time point, perhaps because SCO4076 has a greater role in differentiated tissues rather than vegetative growth.

The mutant phenotype of the single *Mycobacterium* Lsr2 has been characterized but much of this research may not be applicable to *Streptomyces* as most reported phenotypic changes affect genus-specific traits such as cell wall lipid composition (Chen *et al.*, 2006), hypermotility (Arora *et al.*, 2008), horizontal gene transfer within biofilms (Nguyen *et al.*, 2010), multi-drug resistance (Colangeli *et al.*, 2007) and formation of biofilms (Arora *et al.*, 2008; Chen *et al.*, 2006). The genes in the two *Streptomyces* Lsr2-controlled regulons are likely to be different from those in the *Mycobacterium* Lsr2-controlled regulon and their effects on the biology of the organism will surely be different. Also it was not known which of the two *Streptomyces* paralogues was most similar to *Mycobacterium* Lsr2.

The characteristics of Lsr2 have been partially determined and found to be similar to those of *E. coli* H-NS. Lsr2 was found to bridge AT-rich tracts (Chen *et al.*, 2008) and form dimers which are converted to compacted nucleoprotein structures following proteolytic removal of the first three residues (Summers *et al.*, 2012).

This chapter explores the phenotypic effects of genetically deleting the *lsr2* genes from *S. coelicolor*, with particular respect to growth rate, development and antibiotic production. It also investigates the distribution of the different forms throughout the Actinobacteria.

Aims:

- 1) Generate single knockouts of both *lsr2* genes in M145 and complement them *in trans*.
- 2) Generate strains derived from M145 over-expressing each *lsr2* gene.
- 3) Investigate their phenotypes on a variety of media.
- 4) Investigate the phylogenetic distribution of *lsr2* homologues among the Actinomycetes.
- 5) Use this information to infer their respective roles and potential overlap, in particular whether they show partial developmental specialization like HupA/HupS.

5.2 Strains

Single knockouts of both *Lsr2* genes (*sco3375* and *sco4076*) were created in M145 using the REDIRECT[®] protocol and verified as before (section 2.8 and section 4.2). Three independent clones of $\Delta 3375$ ($\Delta 3375.11$, $\Delta 3375.19$ and $\Delta 3375.42$) and three independent clones of $\Delta 4076$ ($\Delta 4076.G1$, $\Delta 4076.G3$ and $\Delta 4076.W10$) were tested and found to behave similarly. Unless otherwise specified, the clones presented in this chapter are $\Delta 3375.42$ and $\Delta 4076.G3$.

Complementation of the $\Delta 3375$ strain was carried out by transforming clone $\Delta 3375.42$ with plasmid pIJ11405 which contained the *sco3375* gene under its native promoter in the vector pMS82, creating the complemented strain B1004. Complementation of the $\Delta 4076$ strain was carried out by transforming clone $\Delta 4076.G3$ with plasmid pIJ11406 which contained the *sco4076* gene under its native promoter in the vector pMS82, creating the complemented strain B1008.

Over-expression strains containing NAPs in the multi-copy plasmid pIJ86 were previously found to give highly unstable phenotypes, possibly due to variations in copy number (gene dosage effects), so the integrative vector pIJ10257 which contains the strong constitutive promoter *ermE**p was used instead. pIJ11407 contained *sco3375* and was conjugated into M145 to create over-expression strain B1005, of which two clones are shown here (B1005.2 and B1005.21). pIJ11408 contained *sco4076* and was conjugated into M145 to create over-expression strain B1007, of which two clones are shown here (B1007.1 and B1007.14).

Wild-type (M145) strains harbouring empty versions of plasmids pMS82 and pIJ10257 were found to grow largely as M145, but produced somewhat more actinorhodin on R5+ and DNA media. These are not shown here.

5.3 Phenotypes screened

The same selection of media were chosen for the initial screen as were used for the HU mutants (Section 4.3): SFM, R5+, DNA and TSB:YEME34%. Cryo-SEM (see section 4.11) was used to examine the morphology of spores sampled from seven-day-old colonies. Spore stress tolerance was assessed by heat-shocking aliquots of spore suspension of similar dilute concentration at 60°C in a water bath for increasing amounts of time, then plating them onto SFM and calculating the proportion of CFUs which had survived at each time point. Some photographs of agar plates in this chapter have a yellowish or bluish cast due to an error in setting the white balance on the camera which could not be removed.

5.4 Summary of phenotypes observed relative to M145 (WT)

$\Delta 3375$

- Delayed aerial mycelium formation on SFM
- Delayed aerial mycelium formation and actinorhodin production on R5+
- Delayed and reduced production of undecylprodigiosin and actinorhodin in liquid
- Moderate spore chain formation defect

B1005

- Delayed development and actinorhodin production on R5+
- Lower final optical density in TSB:YEME34%
- Slow growth and delayed production of undecylprodigiosin and actinorhodin in liquid

$\Delta 4076$

- Slightly delayed development and actinorhodin secretion on R5+
- Early and increased production of actinorhodin on DNA in high density region
- Delayed production of undecylprodigiosin and actinorhodin in liquid
- Moderate sensitivity of spores to heat stress
- Mild spore chain formation defect

B1007

- Slightly delayed development and actinorhodin secretion on R5+
- Low final optical density in TSB:YEME34%
- Loss of production of undecylprodigiosin and actinorhodin in liquid
- Early and increased production of actinorhodin on DNA from low density regions

5.5 Phenotype of $\Delta 3375$ on agar media

The $\Delta 3375$ mutant showed a developmental delay on SFM in comparison to M145 (Figure 5.1). On day 1 M145 had erected white aerial mycelium while $\Delta 3375$ had produced only vegetative mycelium. However it was not known whether the defect was in early growth or the erection of aerial mycelium. By day 3 the mutant had “caught up” and produced grey spores over the densely-plated region, although less-densely plated streaks were still white. This phenotype was fully complemented.

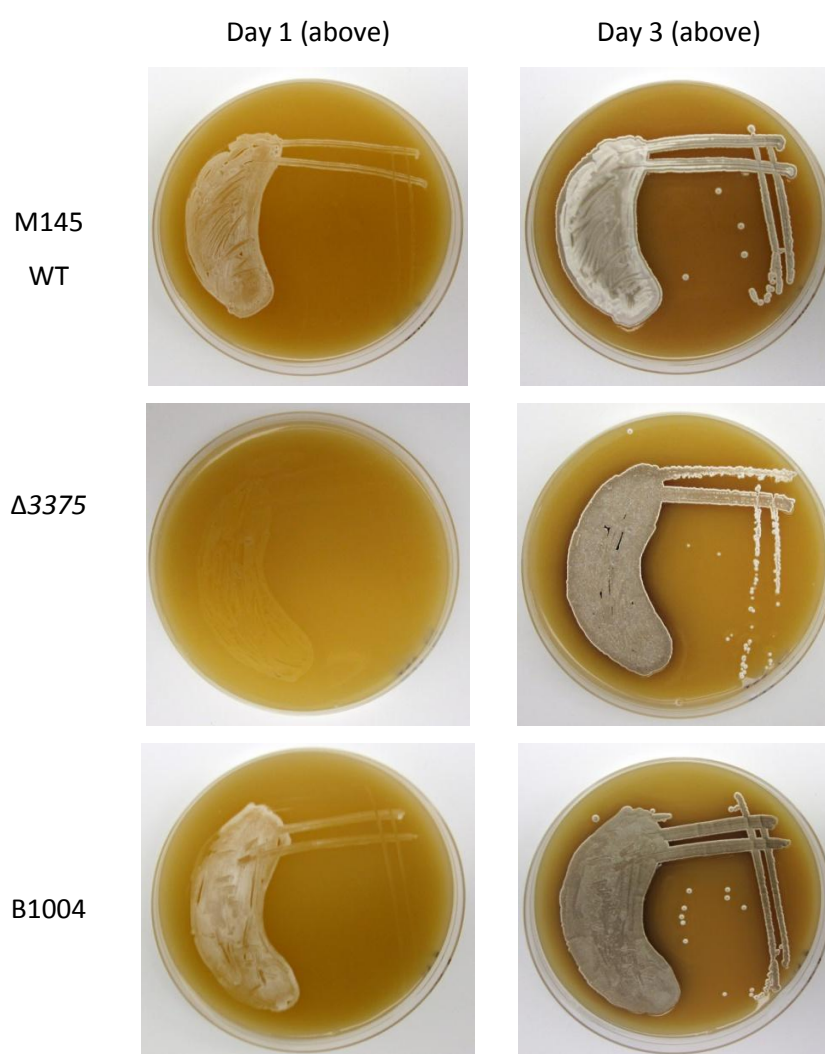


Figure 5.1 Delayed aerial mycelium formation of $\Delta 3375$ on SFM plates. Strain B1004 is a $\Delta 3375$ strain harbouring the complementation plasmid pMS82-*sco3375*.

$\Delta 3375$ behaved inconsistently on DNA plates. There were strong differences between clones of the mutant, none of which could apparently be complemented by plasmid pIJ11405.

$\Delta 3375$ also showed a developmental delay on R5+ (Figure 5.2). On day 2 M145 had begun to secrete actinorhodin into the medium and had a partial covering of white aerial mycelium while $\Delta 3375$ had less extensive aerial mycelium, did not secrete any actinorhodin and had grown more weakly in less densely-plated regions such as the streaks and single colonies. The pattern of changes observed was similar between different independent clones of the mutant but the severity of the phenotype was variable. By day 5 development of the mutant was still delayed: M145 had a fuller coverage of aerial mycelium while the mutant had patches where aerial mycelium had not yet formed. Also on day 5 M145 was secreting more actinorhodin into the medium. The phenotypic changes in growth and actinorhodin production were fully complemented.

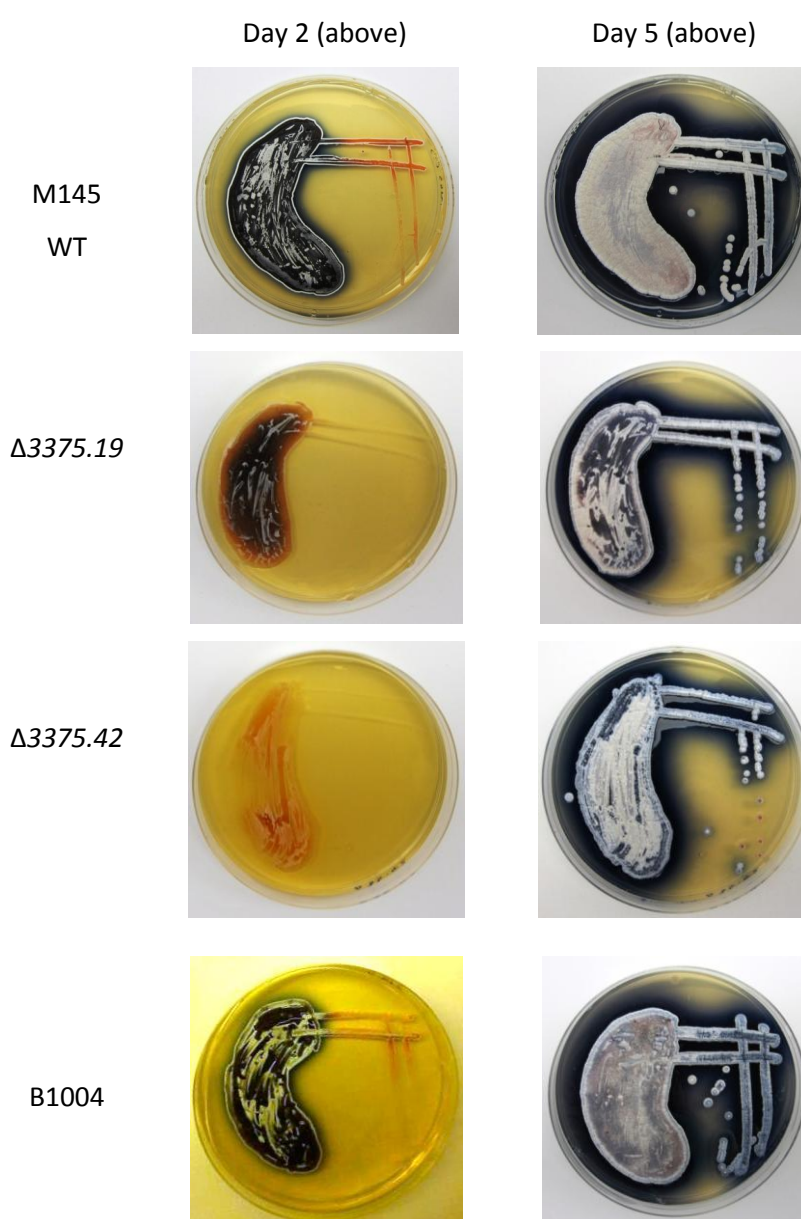


Figure 5.2 Delayed aerial mycelium and actinorhodin secretion of $\Delta 3375$ on R5+ plates. Two independent clones of $\Delta 3375$ are shown ($\Delta 3375.19$ and $\Delta 3375.42$). Strain B1004 is a $\Delta 3375$ strain harbouring the complementation plasmid pMS82-*sco3375*.

5.6 Phenotype of the SCO3375 over-expression strain B1005 on agar media

It was expected that the SCO3375 over-expression strain would show the opposite phenotype to the mutant, however this was not found to be the case.

B1005 was indistinguishable from M145 on SFM and DNA. B1005 formed aerial mycelium later than M145 on R5+ (Figure 5.3) and showed delayed and reduced levels of actinorhodin production. Again, the phenotype varied somewhat between clones: the defects were more severe in B1005 than in $\Delta 3375$, with B1005.21 still being almost devoid of aerial mycelium by day 5. Again, B1005 eventually “caught up” with M145 and aerial hyphae and eventually pigmented spores were seen if the plates were incubated for more than 7 days. This was not due to the presence of the vector itself as an empty vector strain harbouring pIJ10257 alone showed a moderate increase in actinorhodin production.

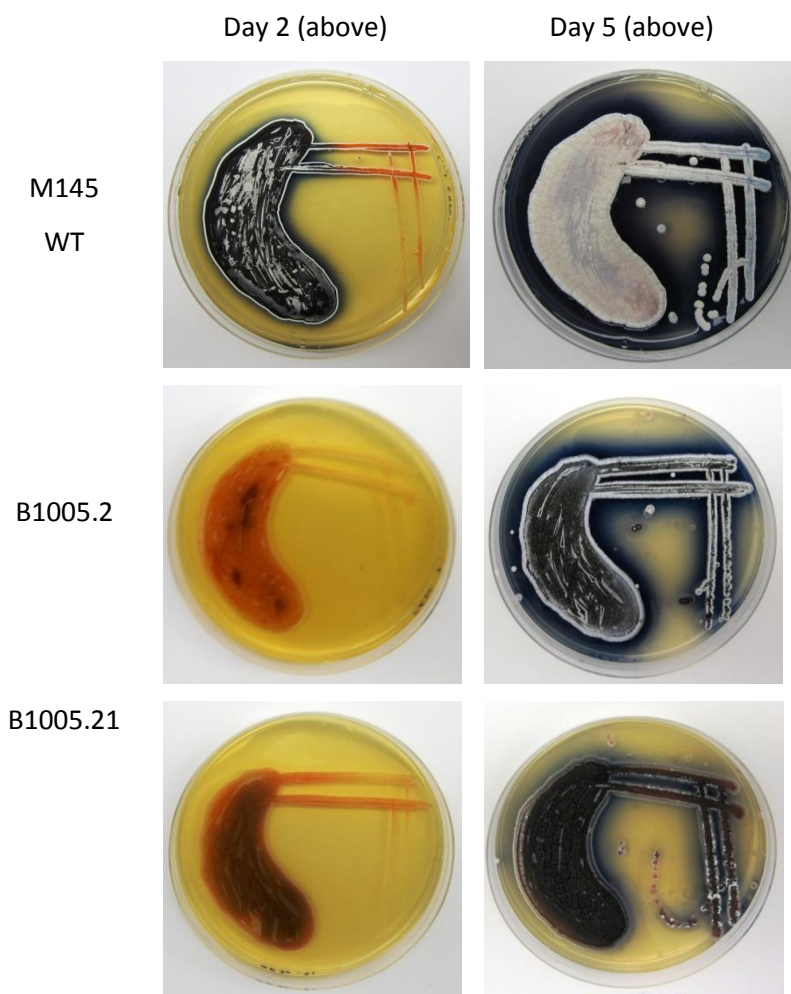


Figure 5.3 Delayed aerial mycelium and actinorhodin secretion of the SCO3375 over-expression strain B1005 on R5+ plates. Two independent clones of the SCO3375 over-expression strain are shown (B1005.2 and B1005.21).

5.7 Phenotype of $\Delta 4076$ on agar media

$\Delta 4076$ was indistinguishable from M145 on SFM. $\Delta 4076$ showed a mild developmental delay on R5+ (Figure 5.4). By day 2, $\Delta 4076$ had produced a comparable amount of growth to M145 but had a lesser coverage of aerial mycelium and had not yet begun to secrete actinorhodin into the medium. However the densely-plated area appeared dark so intracellular actinorhodin was probably still being produced at a similar level to M145. By day 5 one clone ($\Delta 4076.G1$) closely resembled M145 while the other ($\Delta 4076.G3$) was still marginally delayed. By day 6 no differences were seen.



Figure 5.4 Delayed aerial mycelium and reduced actinorhodin secretion of $\Delta 4076$ on R5+ plates. Two independent clones of $\Delta 4076$ are shown ($\Delta 4076.G1$ and $\Delta 4076.G3$). Strain B1008 is a $\Delta 4076$ strain harbouring the complementation plasmid pMS82-*sco4076*.

On DNA medium, the $\Delta 4076$ mutant produced far more actinorhodin than M145 (Figure 5.5). On day 2 the mycelium of the $\Delta 4076$ mutant was less yellow than that of M145 and actinorhodin was being produced within the middle of the lawn region (the lawn edges and streaks remained pale). By day 5 the lawn region was completely blue and actinorhodin was secreted from all regions apart from the single colonies, which remained pale yellow. By day 13 actinorhodin was being secreted into the medium from all regions. However differentiation was similar to M145, with aerial mycelium only being erected around the edges of the mycelium. This phenotype was only partially complemented as B1008 still produced more actinorhodin than M145.

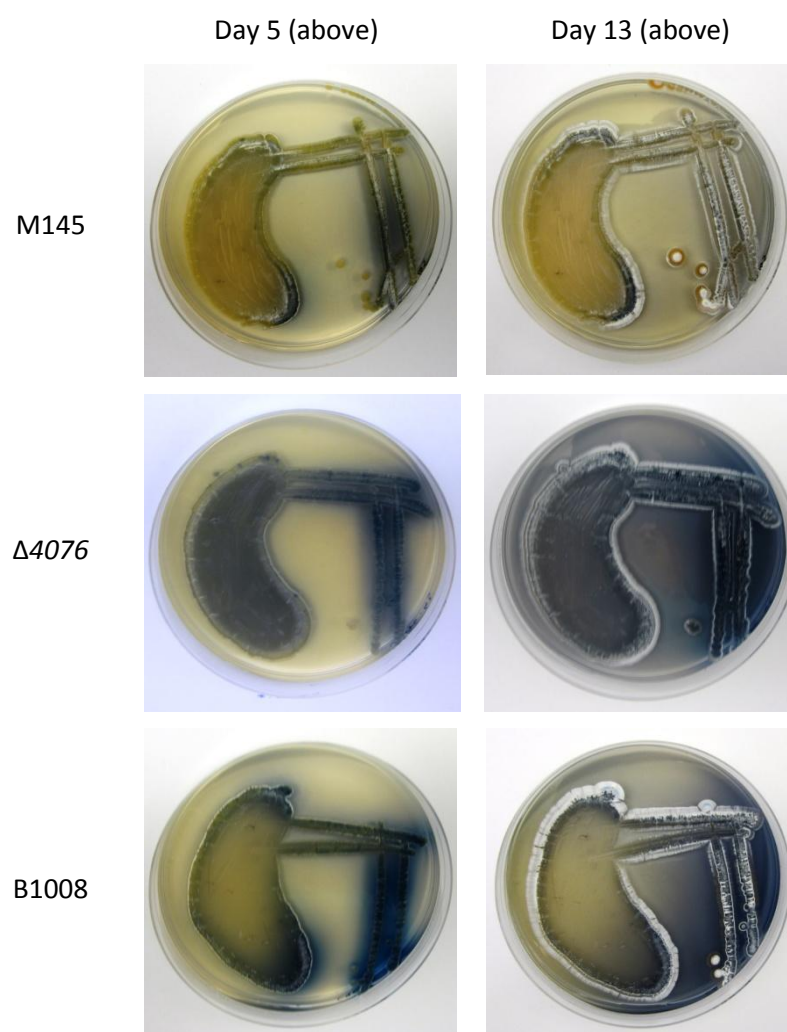


Figure 5.5 Over-production of actinorhodin from densely-plated regions of $\Delta 4076$ on DNA plates. Strain B1008 is a $\Delta 4076$ strain harbouring the complementation plasmid pMS82-*sco4076*.

5.8 Phenotype of the SCO4076 over-expression strain B1007 on agar media

It was originally expected that B1007 would have the opposite phenotype to $\Delta 4076$. However, like the $\Delta 3375$, B1005 and $\Delta 4076$ strains, on R5+ there was again a delay in development with reduced actinorhodin production (Figure 5.6). On day 2 growth was weaker in low density regions but also somewhat weaker in the densely-plated regions, particularly around the margins. By day 5 B1007 was still delayed relative to M145, particularly the clone B1007.14. By day 13 both clones of B1007 had fully differentiated and resembled M145.

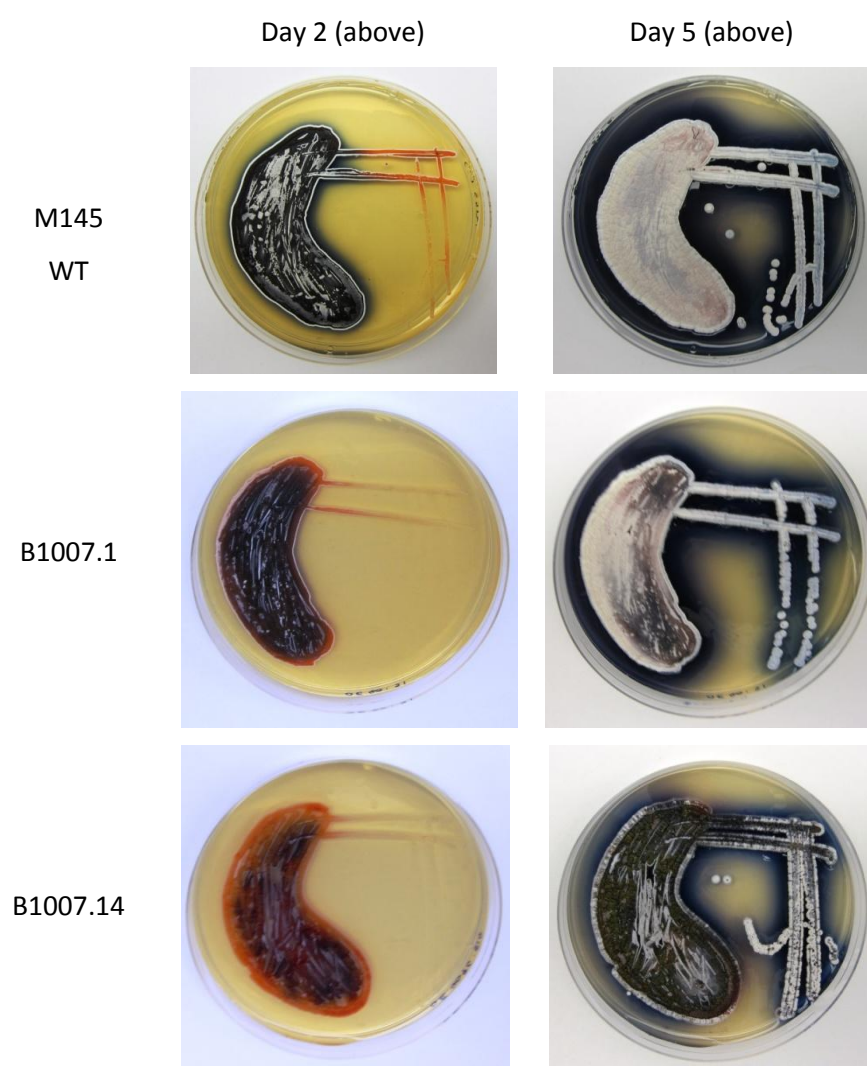


Figure 5.6 Delayed aerial mycelium and actinorhodin secretion of the SCO4076 over-expression strain B1007 on R5+ plates. Two independent clones of the strain are shown (B1007.1 and B1007.14).

On DNA medium, B1007 was paler than M145 and over-produced actinorhodin from the streaks from day 4 onwards (Figure 5.7). Again, B1007.14 had a stronger phenotype than B1007.1.

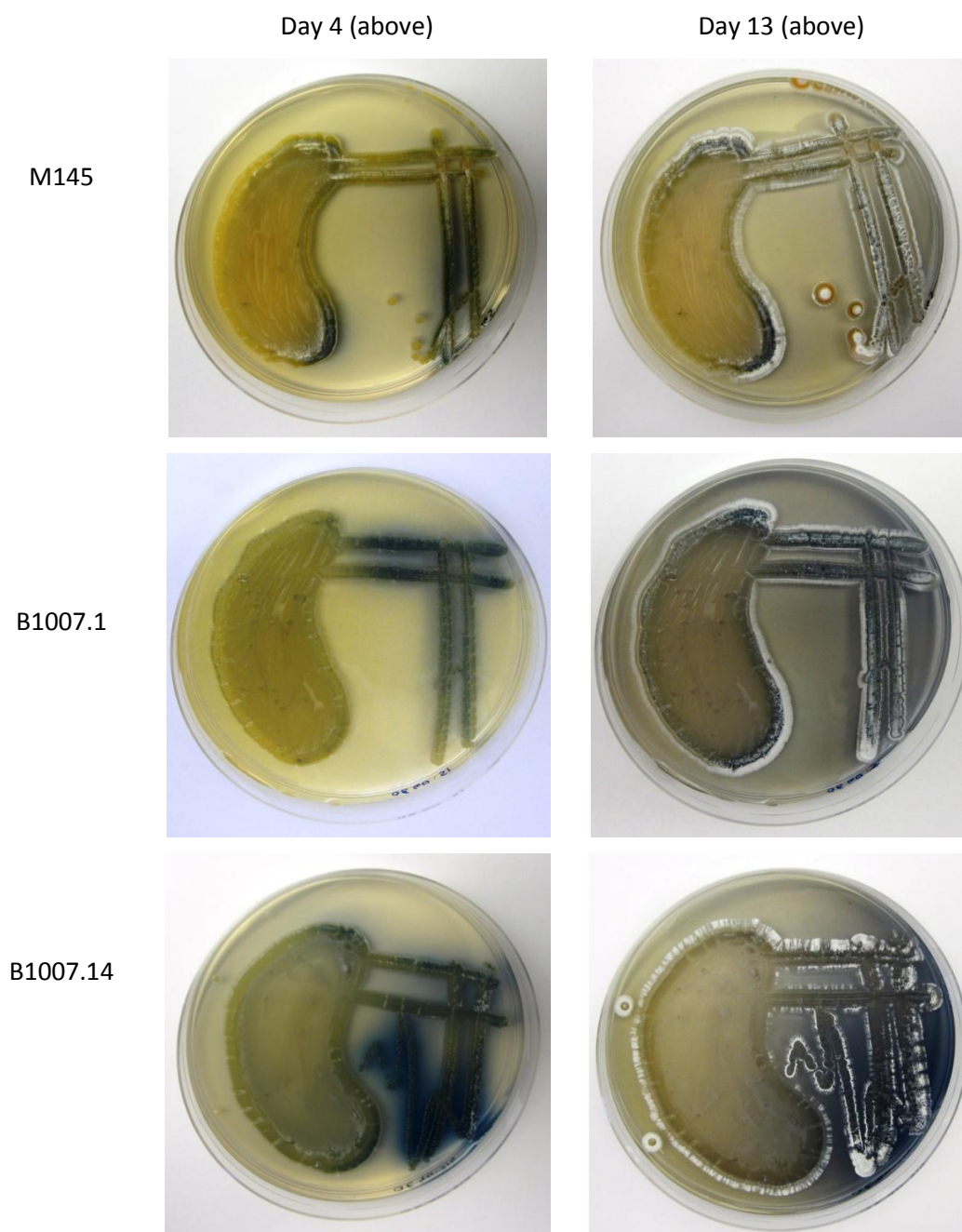


Figure 5.7 Over-production of actinorhodin from less densely-plated regions of the SCO4076 over-expression strain B1007 on DNA plates. Two independent clones of the strain are shown (B1007.1 and B1007.14).

5.9 Liquid growth phenotypes of the $\Delta 3375$ and SCO3375 over-expression strains

Growth curves were performed for the Lsr2 knockout and over-expression strains as before (sections 4.6 and 2.3). For clarity, each graph shows the same M145 data compared against one mutant and its matching over-expression strain. $\Delta 3375$ had a similar growth rate and final density to M145 (Figure 5.8, above) but was delayed in onset of production of undecylprodigiosin (by around 40 hours; Figure 5.8, below left) and actinorhodin (by around 20 hours; Figure 5.8, below right). Once underway, production of both increased at a slower rate and did not reach the same concentration as M145 by the end of the experiment at 143 hours. B1005 initially grew at a similar rate to M145 but too slow at approximately 40 hours (Figure 5.8, above). It was not clear whether its final density was lower than that of M145 as there was excessive variability between flasks during the last two time points. Undecylprodigiosin production was delayed by around 40 hours (Figure 5.8, below left) but the concentration caught up to M145 by approximately 120 hours. Actinorhodin was produced later and in smaller quantities (below right).

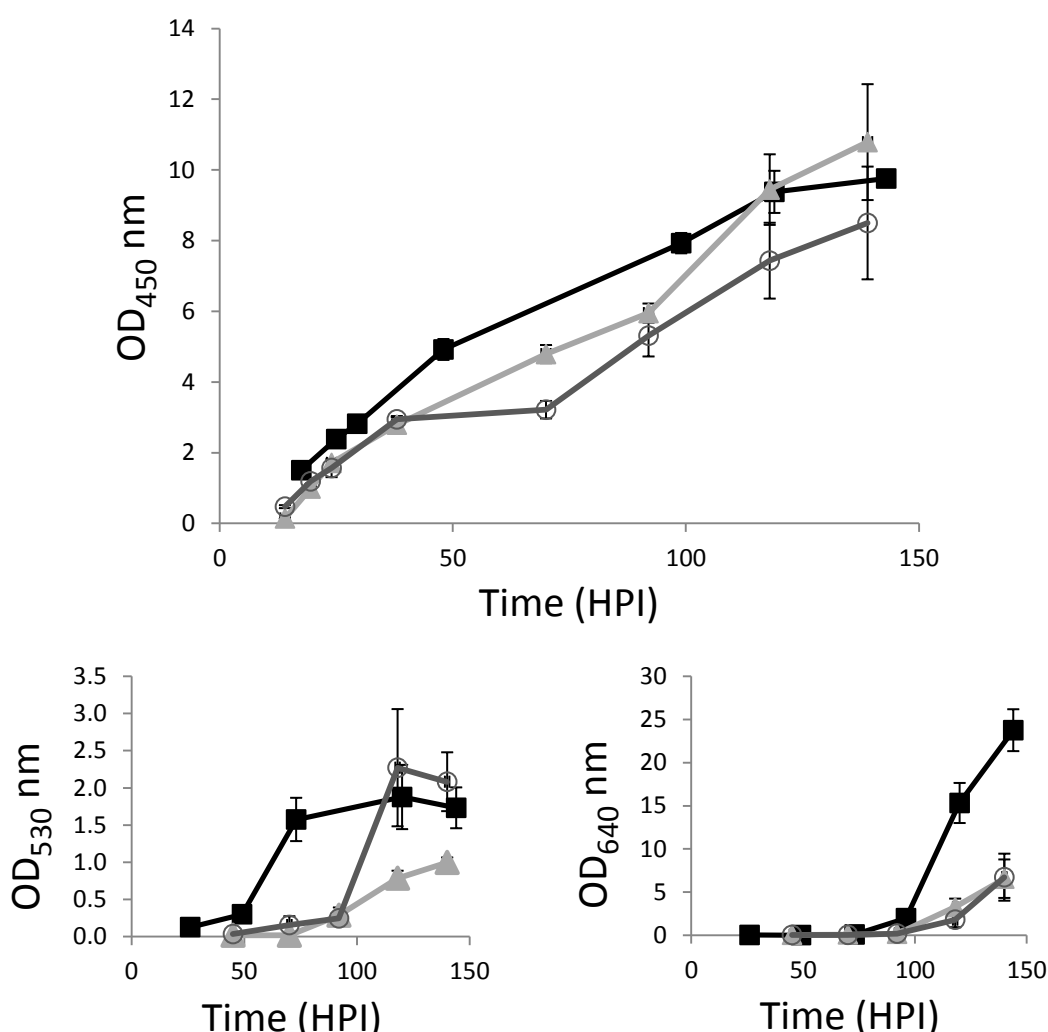


Figure 5.8 Growth (above), undecylprodigiosin production (below left) and actinorhodin production (below right) of $\Delta 3375$ and B1005 as compared to M145 (WT). Error bars represent the SEM of three flasks. M145: closed squares; $\Delta 3375$: closed triangles; B1005: open circles.

5.10 Liquid growth phenotypes of the $\Delta 4076$ and SCO4076 over-expression strains

The $\Delta 4076$ mutant grew at a similar rate to M145 and ended at a similar final density (Figure 5.9, above). Undecylprodigiosin production was delayed by approximately 40 hours but produced at similar levels to M145 by 120 hours (Figure 5.9, below left). Actinorhodin production was delayed by approximately 20 hours and was produced at lower levels than M145 by the end of the experiment at 120 hours (Figure 5.9, below right). B1007 initially grew at the same rate as M145 but began to slow at around approximately 50 hours, with the final density being only around half that of M145 (Figure 5.9, above). It did not produce undecylprodigiosin (Figure 5.9, below left) or actinorhodin (Figure 5.9, below right) in detectable quantities at any time point.

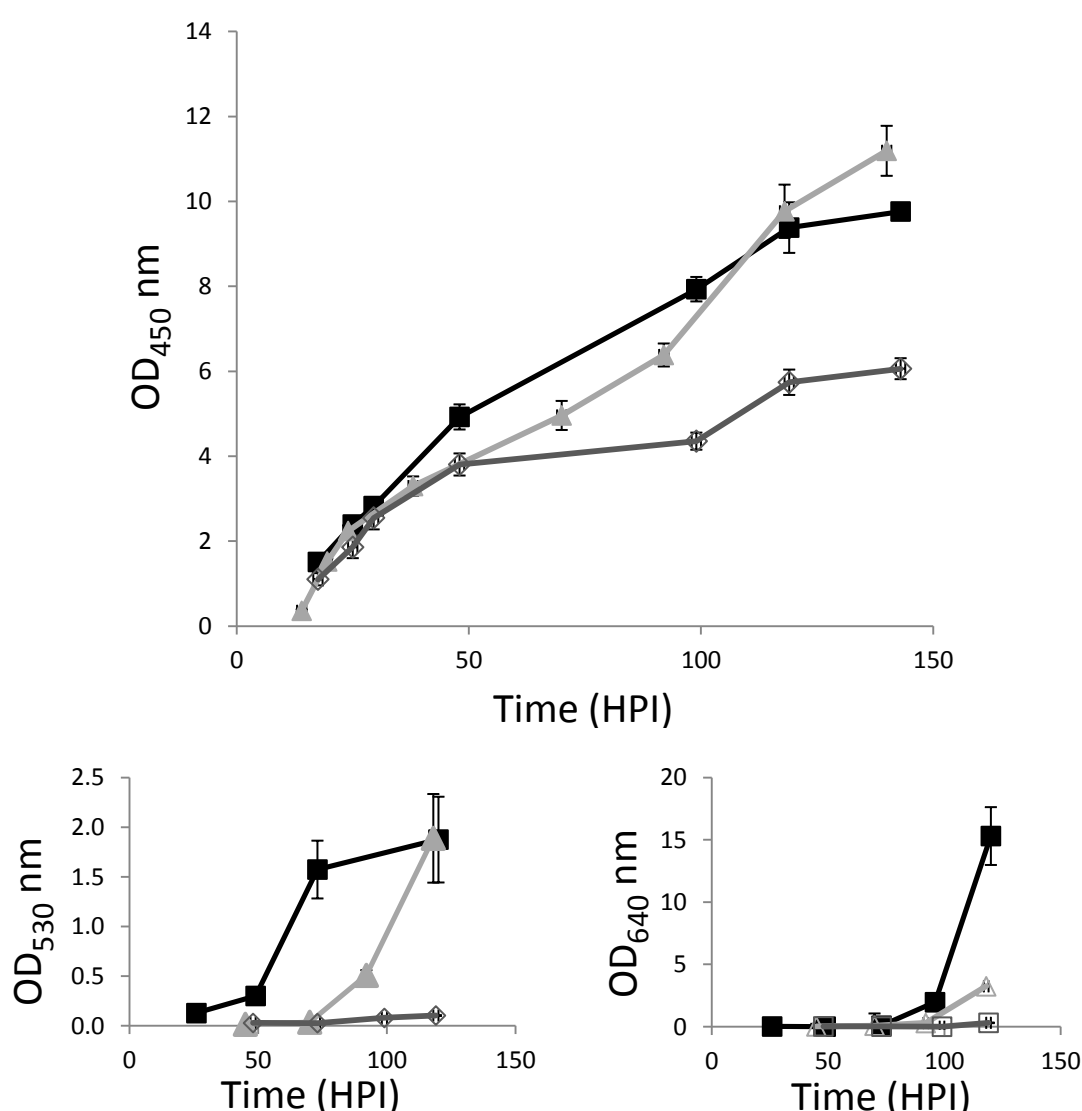


Figure 5.9 Growth (above), undecylprodigiosin production (below left) and actinorhodin production (below right) of $\Delta 4076$ and B1007 as compared to M145 (WT). Error bars represent the SEM of three flasks. M145: closed squares; $\Delta 4076$: closed triangles; B1007: open circles.

5.11 Spore heat tolerance of $\Delta 4076$

As $\Delta 3375$ was found to have an early growth defect and SCO4076 was not very abundant in liquid cultures it was hypothesised that SCO4076 could be specific to later stages of development, therefore the tolerance of $\Delta 4076$ spores towards heat stress was tested by incubating identical aliquots of spores at 60°C for increasing amounts of time and counting the number of surviving CFUs relative to unheated aliquots. The experiment was performed twice, with each time point representing the mean of four aliquots. As controls, the same experiment was performed twice on M145 and on once each on both of the HU mutants because $\Delta hupS$ spores are known to be highly susceptible to heat (Salerno *et al.*, 2009) while $\Delta hupA$ spores are not.

After 60 minutes at 60°C, around 50 - 60% of M145 and $\Delta hupA$ spores were still viable (Figure 5.10). After only 15 minutes the viability of $\Delta hupS$ had dropped to around 10%, which is comparable to the rate previously observed for this genotype (Salerno *et al.*, 2009). After 60 minutes the viability of $\Delta 4076$ fell to 15% and 33% respectively in two independent repeats of the experiment. Therefore $\Delta 4076$ is partially susceptible to heat stress but not to as great an extent as $\Delta hupS$. A single attempt of the same experiment carried out on $\Delta 3375$ suggested that it may have a very mild defect in heat tolerance, less than that of $\Delta 4076$, but addition experiments would be required to verify this as a genuine phenotype.

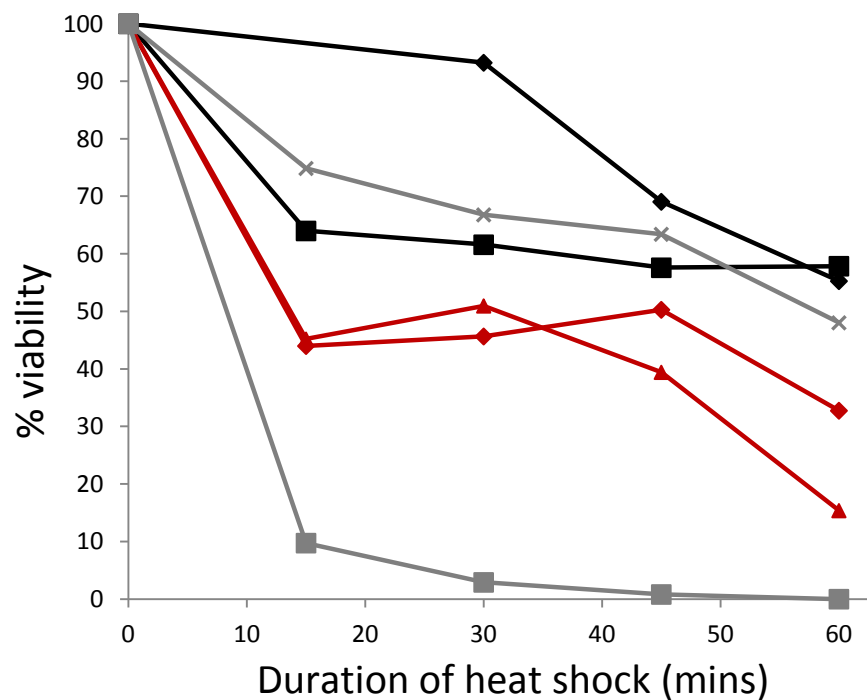


Figure 5.10 Heat sensitivity of spores of M145, $\Delta hupA$, $\Delta hupS$ and $\Delta 4076$. Each line represents one independent experiment and each point represents the mean of four aliquots, normalised against the mean of four untreated aliquots of the same genotype. M145: black diamonds and black squares; $\Delta hupA$: grey crosses; $\Delta hupS$: grey squares; $\Delta 4076$: red triangles and red diamonds.

5.12 Spore morphology of $\Delta 3375$ and $\Delta 4076$.

As $\Delta 4076$ was found to be important for spore heat tolerance, the morphology of the spores was examined by cryo-SEM. The appearance of M145 (WT) spore chains is shown in Figure 4.17 in the previous chapter. Many spore chains contained a majority of normally-formed spores interspersed with a smaller proportion of abnormal spores (Figure 5.11). Some of these appeared to be around double the size of a standard spore (Figure 5.11, red arrows), as though a single septum had failed to form, while some were smaller (Figure 5.11, blue arrows). A moderate proportion of hyphae failed to septate at all (Figure 5.11, white arrows). This phenotype resembled the changes seen in the $\Delta hupS$ mutant but in this case the entirely non-septated hyphae were of a normal diameter rather than thickened.

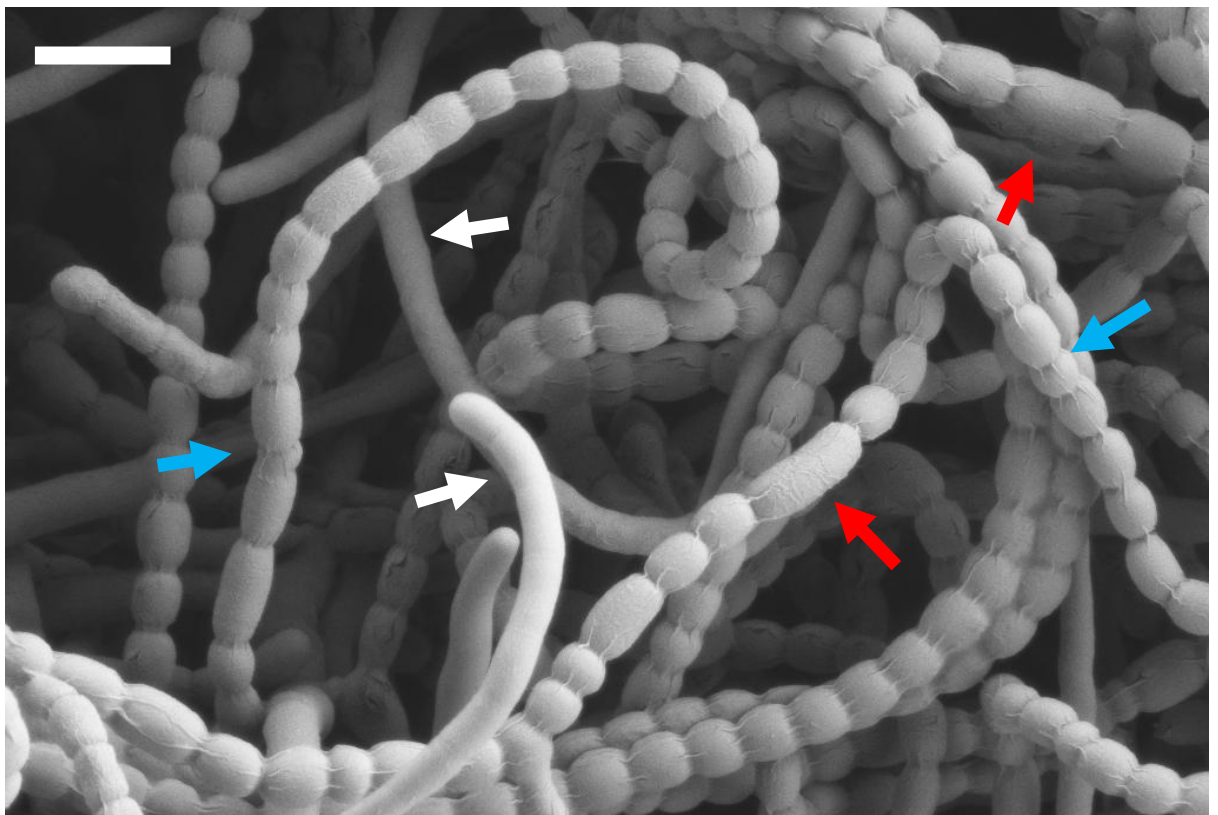


Figure 5.11 Scanning electron micrograph of $\Delta 4076$ grown on SFM for 7 days at 30°C, showing a mixture of normal spore chains, unusually small spores (blue arrows), individual enlarged spores (red arrows) and aerial hyphae which have entirely failed to septate (white arrows). Magnification is x 12,000 and scale bar = 2 μ M.

As $\Delta 3375$ showed a delay in erecting aerial hyphae on the sporulation medium SFM, the morphology of its spores were also examined using SEM. This genotype showed highly abnormal spore chains with multiple types of defect (Figure 5.12). As seen for $\Delta hupS$ and $\Delta 4076$, some hyphae appeared smooth and completely devoid of septation (Figure 5.12, white arrows). Some spores were smaller than usual (Figure 5.12, red arrows) while some were considerably larger (Figure 5.12, red arrows), as if one or more septa had failed to form. There was also an

abnormality that had not previously been observed: some hyphae had a form intermediate between an undivided sporogenic hypha and a fully-divided spore chain, with regular ridges dividing them into portions of approximately the same size as spores (Figure 5.12, green arrows). Perhaps these hyphae have developed as far as positioning septa but were unable to mature.

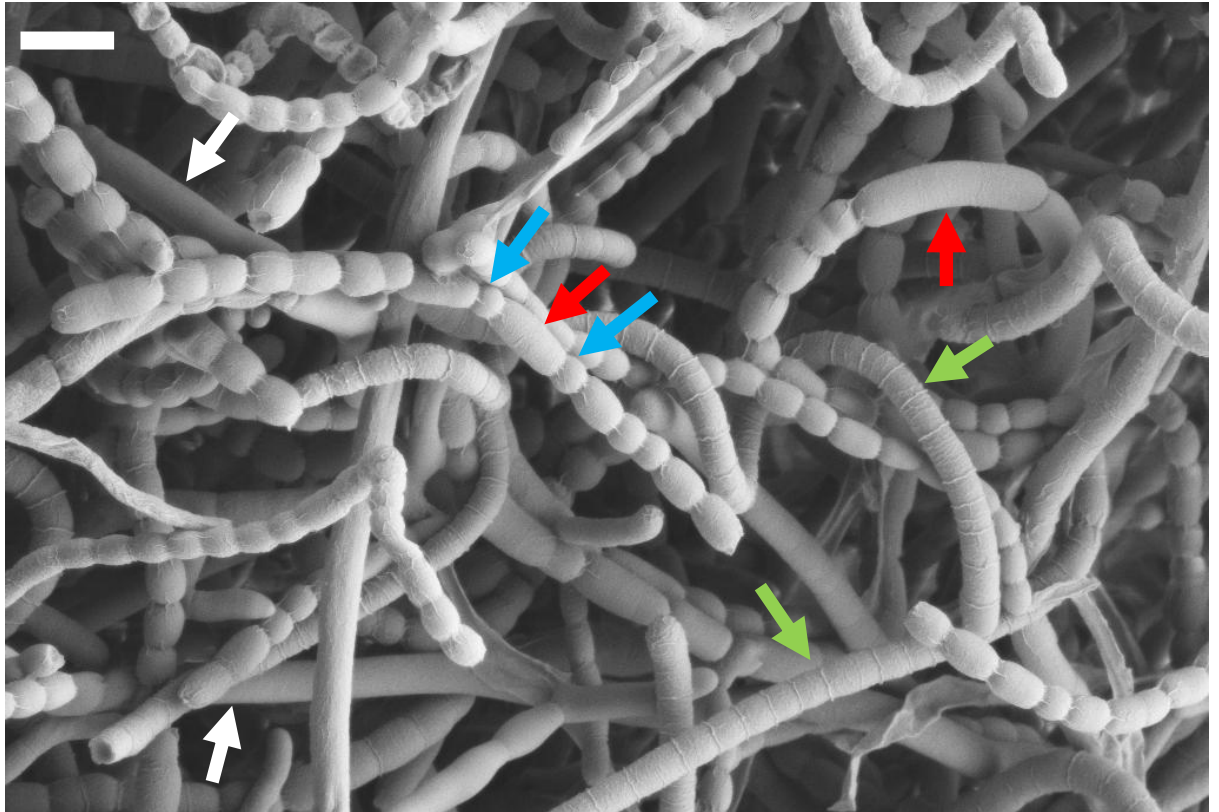


Figure 5.12 Scanning electron micrograph of $\Delta 3375$ grown on SFM for 7 days at 30°C, showing a mixture of normal spore chains, unusually small spores (blue arrows), individual enlarged spores (red arrows), aerial hyphae which have entirely failed to septate (white arrows) and hyphae which had formed evenly-spaced surface ridges but had not developed to form mature spore chains (green arrows). Scale bar = 2 μ M.

It was somewhat surprising that spore chain defects were seen for both $\Delta 4076$ and $\Delta 3375$, in particular that the phenotype was more severe for $\Delta 3375$ which we originally hypothesised to be involved in vegetative growth. However the defects were not as severe as those observed for mutants lacking chaplins (Elliot, 2003) or EsxAB (Roman 2010) and did not affect every spore chain.

5.13 Phylogenetic analysis of Lsr2 distribution among the Actinomycetes

An unrooted neighbour-joining tree was constructed using ClustalW and iTOL for Lsr2 amino acid sequences in the same way as was done for HU (section 4.13). Sequences were retrieved from NCBI by using the amino acid sequence of SCO3375 as a tBLASTn query against individual complete genomes. The same set of species was searched as for HU but no Gram negative or Firmicutes species had homologues of Lsr2. As with HU, real hits showed a high level of conservation over the full length of the sequence and were easily distinguished from erroneous hits. Additionally, Lsr2 homologues from 19 *Mycobacterium* phages were retrieved from NCBI using the terms “Lsr2 phage” in order to look for evidence of horizontal gene transfer. All sequences were trimmed back to 105 amino acids prior to alignment.

Questions to be answered included: 1) When did SCO3375 and SCO4076 diverge? 2) Do they represent two distinct classes of Lsr2 (like HupA and HupS)? 3) Which *Streptomyces* Lsr2 is orthologous to *Mycobacterium* Lsr2? 4) Is it common to single or multiple paralogues per species? 5) Is there any evidence of HGT?

The number of Lsr2 paralogues per species was found to range from one to five, with the majority of species having two or three. As well as having the most copies of HU (5 paralogues), *A. mediterranei* also had the most copies of Lsr2 (5 paralogues). All sequences contained the minor-groove-binding “RGR” motif in their C-terminal DNA-binding domains apart from *Kineococcus radiotolerans* paralogues 2 and 3, which had the motif “TGR” instead. These two sequences also had very unusual domain structures: paralogue 2 has a 14aa N-terminal extension (although the supposed protease site was in the normal position (Summers *et al.*, 2012), so the processed protein would presumably be the normal size) and an approximately 70aa C-terminal extension which would likely affect its DNA-binding properties. Paralogue 3 had an approximately 50aa deletion covering the majority of the oligomerisation domain and would presumably have greatly altered oligomerisation properties. In contrast, the other Lsr2 from *K. radiodurans*, paralogue 1, had a normal domain structure and RGR motif.

The gene tree showed that the Lsr2 proteins did not show a clean division into SCO3375-like and SCO4076-like classes (Figure 5.13). Rather, there was a very complex pattern of distributions which did not closely mirror the species tree. All *Streptomyces* species (plus *Kitasatospora*) clearly had orthologues of both SCO3375 and SCO4076, however the most similar sequences to SCO3375 were found in *Arthrobacter*, *Kineococcus* and *Propionibacterium* while the most similar sequences to SCO4076 were found in *Leifsonia*, *Pseudonocardia* and *Thermobifida*. The relatively recent

appearance (and therefore lower sequence diversity) of Lsr2 relative to HU may partially account for this lack of a clear pattern representing vertical inheritance. The existence of multiple paralogues per lineage is not likely to have arisen by vertical inheritance from a common ancestor with multiple paralogues. While HU showed one major Actinobacterial duplication (into HupA-like and HupS-like classes) with a small number of genus-specific radiations (e.g. *Kineococcus* / *Amycolatopsis*), the *Lsr2* genes showed a stronger propensity to duplicate themselves. Genus-specific duplication events were seen for *Amycolatopsis*, *Kineococcus*, *Gordonia*, *Microbacterium*, *Kitasatospora*, *Nocardia*, *Streptosporangium*, *Thermobifida*, *Salinispora*, *Tsukamurella* and *Pseudonocardia*. It has previously been reported that the *Mycobacterium Lsr2* gene is situated adjacent to the ATPase subunit of the Clp protease complex (Summers *et al.*, 2012). *sco3375* is situated two genes from a similar ATPase (SCO3373) while *sco4076* is not, suggesting that if there is an orthologous relationship it is between *Mycobacterium Lsr2* and *sco3375*. However the amino acid sequence of *Mycobacterium Lsr2* did not cluster with either SCO3375 or SCO4076 so it is not clear to what extent data obtained from *Mycobacterium* could be applied to *Streptomyces*.

The complex pattern of distribution also meant it was not easy to search for clear instances of HGT. One potential instance was seen for *Actinosynnema*, whose genome contained two paralogues. One of these clustered with sequences from *Pseudonocardia*/ *Mycobacterium* and was presumably vertically inherited while the other clustered with sequences from *Mycobacterium* phages and may perhaps have been horizontally acquired.

The threonine residue of SCO3375 which can be phosphorylated (Parker *et al.*, 2010) was found to be conserved in all *Streptomyces* and *Kitasatospora* SCO3375 homologues, while all SCO4076 homologues from these species have a proline at that position, which is not considered a target for phosphorylation. Just fewer than half the chromosomal sequences surveyed had either a threonine or a serine at that position, and many that did not were from a species containing another paralogue which did, suggesting that phosphorylation may be an evolutionarily conserved method of regulation in these proteins. For example both *K. radiotolerans* proteins with unusual domain structures (paralogues 2 and 3) had a histidine at this position (which could potentially be phosphorylated by a different kinase) while the normal version (paralogue 1) has a threonine. However the phage sequences almost invariably had lysine at that position, perhaps because they function differently or perhaps because reliance on a host kinase is evolutionarily unfavourable.

Figure 5.13 Unrooted neighbour-joining tree of Lsr2 amino acid sequences from Actinomycetes and bacteriophages.

Figure 5.13 Unrooted neighbour-joining tree of Lsr2 amino acid sequences from Actinomycetes and bacteriophages.

[this will be an A3 page that folds out, as it doesn't fit easily onto an A4 page with these margins.]

Surprisingly, some of the Lsr2 sequences had short sections of sequence rich in lysine/alanine/proline which were reminiscent of the HupS tail (Figure 5.14). There were 1-3 motifs which each typically consisted of two basic residues separated by 2 - 4 other residues. These were seen on the C-terminal ends of Lsr2 proteins from *Amycolatopsis* and some *Mycobacterium* phages. Oddly, the repeats were also seen towards the middle of one phage sequence (gp *Mycobacterium* phage Thibault).

<i>Amycolatopsis mediterranei</i> (2)	MAQKVSVEILDDIDGSTAAQTVQFSLDGVTFEIDLSDRNASALRHELARY IGAGRRVGGGRKVRVATGQSTTASDRERNQIRAWANANGFEVSEGRRL SSEVVAAYERVHDADTESSVPS <u>PRKRAPRRK</u> VAA
<i>Amycolatopsis mediterranei</i> (3)	MAQKVLVEILDDIDGSTAAQTVQFGLDGVTYEIDLSDENASALRDELARY IGAGRRIGGRKVRVATGQSTTTSTTDRERNQQIRAWANANGYEVSEGR RLSSEVIAAYEQAQVEEAEAPAAPP <u>RKRAPRKKVAAAKK</u>
<i>Amycolatopsis mediterranei</i> (4)	MAQRVQVQMVDLDLGSEASQTVPFSLDGVTYEIDLSEENASALRDELG RYVAAARRIGGRKVRLATGQSLSGPSGAGTDRERNRQIREWAQANGYE VAERGRLSSEIIAGFEADQAAAAEPAE <u>AKPARKRATRKKK</u> S
<i>Amycolatopsis mediterranei</i> (5)	MAQKVLVEILDDIDGSAAAQTVQFGLDGVTYEIDLSDDNATALRDELAR FIGAGRRIGGRVRVATGQSTATTASDRQRNQIRAWANANGYEVSEGR GRLSSEVVSAYEQAQVAEAATPAPT <u>PRKRSRKKTA</u> A
gp33 <i>Mycobacterium</i> phage LeBron	MGKRVTVTLFDDLEPELEASTERNFAVDGVEYHLDLSDKNAKAFDKDLA KWLEVATRVGRAERVRRRTNRITHSGGEPGLPLAEIRQWARANGFDISD KGRVSAEVRRAWKAATSEAEPE <u>KKPAAKKAAPRK</u> DAPEFSAAQS
gp51 <i>Mycobacterium</i> phage Thibault	MIKTVLVDDITGDEGAKTRRFSIEGEAYEIDLVDSTYEELEKALAGFIES <u>ARKVHGSKKAAPKPKAK</u> VSPDKADQLSAIRDWARRSGKKVSDRGRIPRDI VNAFEEAHPQFSHAG
gp33 <i>Mycobacterium</i> phage UPIE	MGKRVTVTLFDDLEPELEASTERNFAVDGVEYHLDLSDKNAKAFDKDLA KWLEVATRVGRAERVRRRTNRITHSGGEPGLPLAEIRQWARANGFDISD KGRVSAEVRRAWKAATSEAEPE <u>KKPAAKKAAPRK</u> DAPEFSAAQS
gp36 <i>Mycobacterium</i> phage Rumpelstiltskin	MGKRVTVTLFDDYEPELEAEVERNFSIEGVAYHLDLSEKNAKQFDKDLEK WLGVATRVGKESTRRRSSSKNTYSGGEPGLPLADVRKWAQLNGHEVSD RGRVSAQIVRDWRAAGSPTGDEI <u>KAAAPAKKAAPAKKAAPAQKK</u> DDP QFSAAGES

Figure 5.14 Lsr2 homologues containing lysine-/alanine-/proline-rich repeats.

5.14 Discussion

It was originally conjectured that SCO3375 and SCO4076 may show partial developmental specialization towards vegetative growth and differentiated tissues respectively, in a similar manner to HupA and HupS, because SCO3375 was highly abundant in liquid culture nucleoids throughout growth while SCO4076 was not detected at high abundance at any time point. However the changes in phenotype observed for their mutant strains did not support this hypothesis. The phenotypes of the Δ l_{sr}2 mutants were relatively mild and showed a greater degree of overlap than expected, perhaps due to redundancy between these proteins. For example both showed reduced antibiotic production in liquid culture (despite a normal growth rate) and both showed defects in spore chain morphology. Unfortunately time constraints did not permit the construction of a double mutant. We cannot be sure that there are no further H-NS-like proteins in this species, as *Mycobacterium* has another gene, *hns*, which is not related by sequence to Lsr2 or H-NS.

Some phenotypic changes occurred in one mutant only. The developmental delay seen for Δ 3375 on SFM suggests a role in early growth or erection of aerial mycelium and is consistent with a large spike in expression at 12 hours in microarray data generated from cultures grown on agar media (not shown). It would be interesting to determine whether these effects were caused by a delay in germination in this mutant. SCO4076 had a role in actinorhodin production on DNA while SCO3375 did not show a consistent phenotype on this medium. Interestingly, Δ 4076 produced excessive actinorhodin from regions of high plating density while B1007 produced excessive actinorhodin from regions of low plating density and single colonies. This density-dependent effect could indicate changes in quorum-sensing signalling or receiving, but it could alternatively be caused by changes in the way individual hyphae alter their immediate surroundings e.g. using up nutrients faster.

It was expected that mutant strains would also show the opposite phenotypic changes to their cognate over-expression strains. However this was only found to be the case for actinorhodin production by Δ 4076/B1007 on DNA medium. All four strains showed delayed and/or reduced antibiotic production in liquid media to varying degrees. In particular the phenotypes of all four strains on R5+ were strikingly similar, with delays in development and actinorhodin secretion (of varying severity). Similar changes were seen for both HU mutants on R5+ and in TSB:YEME45% so it is possible that these changes occur frequently when there is a disturbance to the structure of the nucleoid.

A phylogenetic survey of Lsr2 proteins within the Actinobacteria was less informative than the survey performed for the HU proteins, most likely because these are relatively young sequences with less sequence diversity (i.e. a potentially weaker evolutionary signal). However it was shown that genus-specific duplications are very common and that one of the two paralogues present in *Actinosynnema* may have been horizontally acquired as it most resembles Lsr2 sequences from *Mycobacterium* phages. It is possible that *Mycobacterium* Lsr2 is orthologous to SCO3375 and not SCO4076 due to its genomic context beside an ATPase subunit of Clp protease, although the gene tree did not strongly support this relationship. Surprisingly, many of the sequences examined contained a short C-terminal lysine-rich region of similar amino acid composition to the HupS tail.

The conservation of a phosphorylatable serine or threonine within the DNA-binding domain of at least one paralogue of Lsr2 from the majority of Actinobacteria is particularly intriguing. *S. coelicolor* spores treated with protein kinase inhibitors show delayed germ tube emergence (Paleckova *et al.*, 2007) which could be part of the same pathway as the mechanism behind the growth delay seen for $\Delta 3375$. It would be very interesting to look at the occurrence of serine/threonine phosphorylation within different species of Actinobacteria at different stages of growth or under different conditions. Also it would be useful to know what effects phosphorylation has on the biochemical function of Lsr2, for example on affinity for various DNA substrates. In particular it would be worth investigating its role in *M. tuberculosis*, which encodes a single copy of Lsr2, to look for important regulatory steps in virulence which could be therapeutically targeted by kinase or phosphatase inhibitors. Indeed, one screen of kinase inhibitors showed that around 5% showed activity against *M. tuberculosis* (in a single-dose growth assay) but the targets involved were not known (Reynolds *et al.*, 2012).

It was surprising that the number of paralogues varied so much between species. It would be very interesting to investigate how the different versions interact with one another at the biochemical level, in particular whether SCO3375 and SCO4076 are able to form heterodimers as this may account for the commonalities in their mutant phenotypes. The high diversity of paralogue complements and the possibly widespread role of phosphorylation may serve to increase the diversity of higher-order structures which can be formed by these proteins.

Summary:

- 1) Both paralogues of *Lsr2* (*sco3375* and *sco4076*) were deleted and over-expressed in *S. coelicolor* M145. The mutants were successfully complemented *in trans* with the vector pMS82 containing the genes under their native promoters.
- 2) On agar media all four strains ($\Delta 3375$, $\Delta 4076$, B1005, B1007) showed a similar developmental delay and reduced actinorhodin production on R5+. $\Delta 3375$ also showed a 1-day delay in forming aerial mycelium on SFM which the other strains did not share. In liquid media both over-expression strains had a reduced final density and all four had a normal growth rate with reduced or delayed antibiotic production. $\Delta 4076$ spores were less resistant to heat stress and both $\Delta 3375$ and $\Delta 4076$ showed moderate defects in spore chain formation.
- 3) Phylogenetic analysis showed that most Actinomycetes have multiple copies of *Lsr2* and that many of these were recent duplications. One possible instance of HGT was identified where one of the two *Actinosynnema* *Lsr2* paralogues clustered with sequences from *Mycobacterium* phages rather than close relatives of *Actinosynnema*. Lysine-rich repeats which resembled the HupS tail were found on the C-terminus (DNA-binding domain) of several phage homologues and four of the five sequences from *A. mediterranei*.

Chapter 6 – SCO5592, a putative sRNA-binding protein

6.1 Introduction

In addition to NAP-mediated regulation at the level of transcription, gene expression can also be regulated at the post-transcriptional level by NAPs such as HU and StpA (section 1.2) and by global RNA chaperones such as Hfq and CsrA/RsmA (section 1.3). There are common features to both types of regulator including small size, recognition of structures (rather than consensus sequences), the propensity to form higher-order structures in association with nucleic acids and highly pleiotropic mutant phenotypes.

Gene expression can be controlled at the post-transcriptional level through the action of sRNAs, whose importance and ubiquity are becoming increasingly well known. sRNAs are small (50 – 500 nt), untranslated RNA species which anneal to cognate mRNAs and either prevent translation by targeting them for enzymatic degradation or by modifying the secondary structure of the RNA to make the ribosomal binding site either more or less accessible to ribosomes. sRNAs have been identified in *S. coelicolor* and several species from the related Actinobacterial genera *Mycobacterium* (Arnvig and Young, 2009; DiChiara *et al.*, 2010) and *Corynebacterium* (Zemanova *et al.*, 2008) and may be expected to be present throughout the Actinobacteria. Swiercz *et al.* (2008) used a combined bioinformatics and experimental approach to identify nine sRNAs in *S. coelicolor*, of which three were highly dependent on growth conditions and several were expressed at specific developmental stages. These were mostly *trans*-encoded (i.e. produced from intergenic regions) but one was *cis*-encoded (i.e. antisense to the gene it was produced from). A previous study reported 24 sRNAs in this species (Panek *et al.*, 2008), of which two were also identified in Swiercz *et al.* (2008), but these were only detected by microarray and not validated by Northern blot. A recent study used high-throughput sequencing to identify a further 63 putative sRNAs of which 11 were confirmed by Northern blot (Vockenhuber *et al.*, 2011). The identities of twelve sRNAs were similarly identified by Northern blot analysis of total RNA samples extracted from *Streptomyces griseus*, of which seven were found to be growth-phase-dependent (Tezuka *et al.*, 2009). Surprisingly, genetic deletion revealed that none of these sRNAs had a detectable mutant phenotype. To date only one sRNA has been studied in detail in *S. coelicolor*: scr5239 over-expression caused slow growth on solid rich medium, loss of expression of the agarose gene *dagA* (with concomitant loss of agar utilisation) and a reduction in actinorhodin production (Vockenhuber and Suess, 2012). It is anticipated that more of the identified sRNAs will

play an important role in regulating gene expression in *S. coelicolor* and that some of them will be relevant to control of secondary metabolism. The protein or proteins which mediate sRNA regulation in the Actinomycetes are unknown but the presence of *trans*-encoded sRNAs strongly suggests that one or more proteins with sRNA-binding activity exist. In many species the mediating protein is Hfq but this could not be identified in the *S. coelicolor* genome by BLAST analysis (Sun *et al.*, 2002), by structure-based JPRED searches of small proteins (Swiercz *et al.*, 2008), or by HMMER searches (personal communication, Juan-Pablo Gomez Escribano) so most likely it not present. Clear homologues of the CsrA/RsmA-family of proteins were also unidentifiable by BLAST searches performed by myself (not shown). Alternatively it is possible that an entirely new class of protein could be involved. If a global RNA chaperone could be identified in *Streptomyces* it would be expected to have an important role in controlling secondary metabolism and may be a potential target for intervention. If an orthologue of such a protein were widespread among the Actinomycetes then this could have implications for drug discovery, particularly in the activation of cryptic clusters. It may also have a role in the regulation of pathogenicity in closely-related species such as *Mycobacterium*, as has been shown for Hfq in many species (Christiansen *et al.*, 2004; Hansen and Kaper, 2009; Liu *et al.*, 2011).

E. coli isolated nucleoids were found to be rich in Hfq in section 3.3 of this work and in experiments of Ohniwa *et al.* (2011) so it was logical to ask whether any of the candidate NAPs detected in *S. coelicolor* nucleoids in chapter 3 were actually part of the missing sRNA chaperone machinery. The protein would not necessarily need to be a homologue of Hfq or CsrA/RsmA. This represents a novel approach to an important question.

This chapter investigates the possibility that one of the candidate proteins identified in chapter 3 might be a global sRNA-binding protein which acts in an similar manner to Hfq or CsrA/RsmA.

Aims:

- 1) Identification of the most promising candidate RNA chaperones.
- 2) Attempted complementation of an *S. enterica* Δhfq mutant with the candidates.
- 3) *In silico* analysis of the SCO5592 amino acid sequence.
- 4) Generation of a $\Delta 5592$ mutant and phenotypic analysis.
- 5) Purification of His-5592 and measurement of its oligomerisation state.

6.2 Candidates arising from chapter 3

The proteomic analysis of the nucleoid described in chapter 3 generated a list of 25 proteins predicted to be both highly abundant and able to bind DNA or RNA (Figure 3.9). Many of these could be assigned to a particular class of transcription factors (e.g. MerR) or were already known to be NAPs (e.g. HupA/sIHF) and so were discounted as global RNA chaperones while others were not deemed sufficiently abundant. The three candidates which stood out were SCO3767, SCO5592 and SCO5783.

SCO3767 (tellurium-resistance protein; 17 kDa) is a member of a class of tellurium-resistance proteins (TerB). In the proteomic analysis it was found to be highly abundant at all time points but particularly during early and late vegetative phase, decreasing slightly during stationary phase. An NMR structure generated for *Klebsiella pneumoniae* TerB showed it to have one highly basic and one highly acidic face (Chiang *et al.*, 2008), which is a feature common to many RNA-binding proteins (Figure 6.1), but the ability of this class of protein to bind nucleic acids has never been investigated.

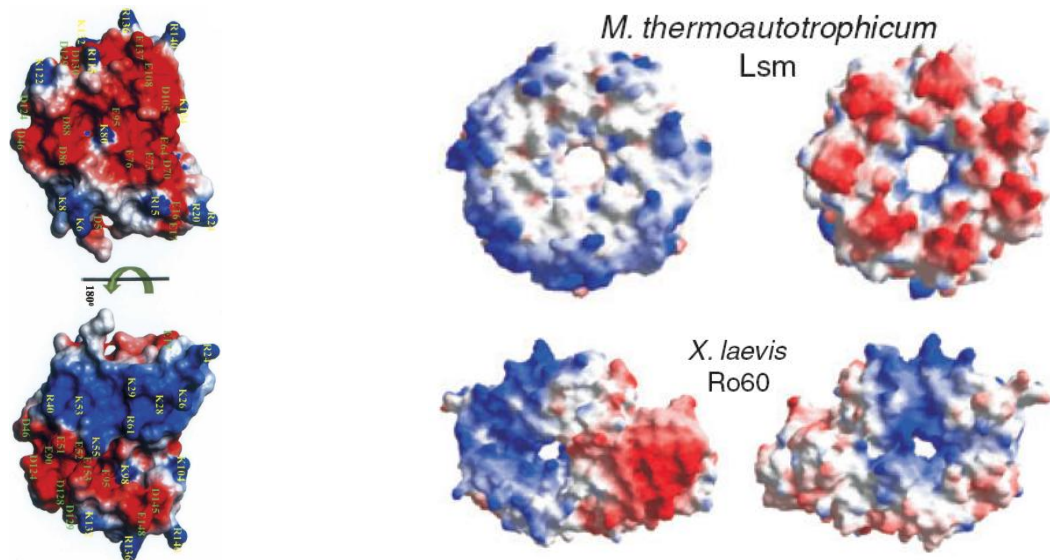


Figure 6.1 Left: NMR structure of the tellurium-resistance protein TerB from Chiang *et al.* (2008). Right: crystal structures of the RNA-binding proteins Lsm (above) and Ro60 (below), showing doughnut-like structures with highly charged surface patches. From Pruijn *et al.* (2005).

SCO5592 is a small (8.9 kDa) and highly conserved protein which was found to be particularly abundant during late vegetative growth. It is annotated as an RNA-binding protein on the basis of similarity to proteins containing a KH domain, although it does not closely resemble any other characterised protein. SCO5783 (17 kDa) was highly abundant during early and especially late vegetative growth.

6.3 Microsynteny of the *hfq* superoperon

The genomic context of *hfq* is highly conserved between bacterial species, with the gene usually being situated either immediately downstream of *miaA* (a tRNA transferase) or immediately upstream of *hflX* (a GTPase which binds the 50S ribosomal subunit), or frequently both (Figure 6.2; information retrieved from NCBI RefSeq annotated genomes using the UCSC Microbial Genome Browser available at <http://microbes.ucsc.edu/>). *mutL* (a gene involved in DNA mismatch repair) is often found upstream of *miaA*, while *hflC* (a protease subunit) and *hflK* (a protein kinase) are often found downstream of *hflX*.

Species	Genomic context of Hfq
<i>E. coli</i>	Downstream of <i>mutL/miaA</i> , upstream of <i>hflX</i>
<i>Salmonella enterica</i>	Downstream of <i>mutL/miaA</i> , upstream of <i>hflX/hflK/hflC</i>
<i>Moraxella</i>	Downstream of <i>miaA</i> , upstream of <i>kpsF</i> (epimerase/isomerase)
<i>Burkholderia multivorans</i>	Downstream of <i>engA</i> , upstream of <i>hflX/hflK/hflC</i>
<i>Vibrio fischeri</i>	Downstream of <i>mutL/miaA</i> , upstream of <i>hflX</i> then <i>hflK/hflC</i>
<i>Borrelia burgdorferi</i>	Downstream of <i>mutS</i> , upstream of <i>nusA/rbfA</i>
<i>Acinetobacter baumannii</i>	Downstream of <i>mutL/miaA</i> , upstream of several enzymes
<i>A. tumefaciens</i>	Downstream of nitrogen metabolism genes, upstream of <i>hflX</i>

Figure 6.2 The genomic context of the *hfq* gene is frequently conserved between bacterial species. It is commonly found downstream of *mutL/miaA* and upstream of *hflX*. Information retrieved from NCBI RefSeq annotated genomes using the UCSC Microbial Genome Browser.

The genes adjacent to *miaA* and *hflX* in *S. coelicolor* were identified. In *E. coli* and *Salmonella enterica*, *hfq* is the only gene separating *miaA* and *hflX*. In *S. coelicolor* the latter two genes are still closely linked but are instead separated by four genes (*sco5792*, *sco5793*, *sco5794* and *sco5795*), none of which are annotated as RNA-binding proteins (Figure 6.3). *sco5792* is a 17.5 kDa uncharacterised protein tentatively annotated as containing three possible transmembrane domains; *sco5793* is an epimerase; *sco5794* and *sco5795* are respectively a kinase and a putative protease, which could conceivably correspond to *hflC* and *hflK* (but this could not be confirmed by BLASTp searches). As the function of *sco5792* could not be determined from sequence searches, it was added to the list of possible candidates, although it was not found to be as abundant in nucleoid samples as the candidate brought forwards from chapter 3. The gene cluster containing *sco5792* lies just downstream of a large number of genes associated with ribosomes and translation and is situated only eight genes upstream of *sco5783*, which was identified as a putative NAP from the proteomic study.

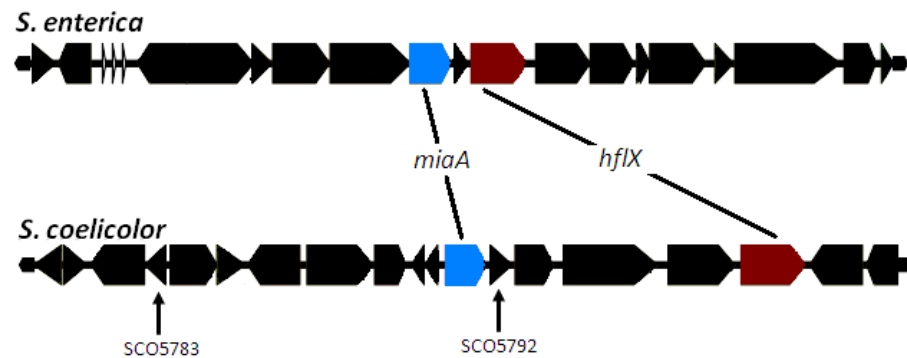


Figure 6.3 Genomic context of the linked genes *miaA* and *hflX* in *S. enterica* and *S. coelicolor*. *hflX* is the only gene to lie between *miaA* and *hflX* in *S. enterica* while in *S. coelicolor* they are separated by four genes: *sco5792* (unknown function), *sco5793* (epimerase), *sco5794* (kinase) and *sco5795* (putative protease).

6.4 Screening of candidates for Hfq-like activity by attempting to complement an *S. enterica* Δhfq mutant with *S. coelicolor* genes.

Sometimes proteins which appear to be unrelated can functionally substitute for one another. In some cases the proteins are homologues but have diverged sufficiently far in sequence as to be unrecognisable by BLAST searches, although they still share the same protein fold. For example Hfq homologues which complement an *E. coli* Δhfq mutant have been found only recently in the cyanobacteria and in the Archaeal species *Methanococcus janaschii* due to their atypical sequences (Boggild *et al.*, 2009; Nielsen *et al.*, 2007). In other cases the proteins have evolved to have the same activity independently, for example an *E. coli* Δhns mutant can be complemented by many genes, from diverse organisms, that are unrelated to H-NS on the basis of amino acid sequence and protein fold but have been found to be capable of bridging AT-rich tracts of DNA to form higher-order nucleoprotein complexes (Dame *et al.*, 2005; Gordon *et al.*, 2011). A second example of this is the complementation of an *E. coli* $\Delta hupA$ mutant by such unexpected proteins as eukaryotic HMG-box proteins (Sasaki *et al.*, 2003) or even the 34 kDa C-terminal domain of DNA gyrase A from *Borrelia* (Knight and Samuels, 1999).

In order to test the hypothesis that one of the four candidates (*sco5592*, *sco3767*, *sco5792*, *sco5783*) could encode a protein with an Hfq-like activity, these proteins were over-expressed in an *S. enterica* Δhfq strain which was previously shown to have a phenotype of lower final optical density than the parental strain (SL1344). Restoration of the wild-type growth rate by one of the constructs would strongly suggest that the corresponding *S. coelicolor* gene had a similar function and could be prioritized as a candidate for the missing Hfq activity. Each gene was amplified using a forward primer that included the strong ribosomal binding site “AGGAGGA” encoded 8 nt

upstream of the start codon. These inserts were cloned into the low-copy-number expression vector pWKS30 (Wang and Kushner, 1991) using the *Asp*718 and *Hind*III restriction sites, under the control of the *lac* promoter. Constructs were confirmed by Sanger sequencing then transformed into the *S. enterica* Δhfq strain JVS-0255 (Sittka *et al.*, 2007). As suitable controls, two strains derived from the same mutant were used: one strain contained the empty vector pWKS30 alone and one strain was complemented *in trans* with *S. enterica* *hfq* amplified from *S. enterica* genomic DNA and inserted into the vector pWKS30 in the same way as the *S. coelicolor* genes. *Salmonella* strains were kindly provided by Jörg Vogel (Universität Würzburg, Würzburg, Germany) and the vector pWKS30 was kindly provided by Arthur Thompson (Institute for Food Research, Norwich, UK). Growth of each of the strain were compared over a 10 hour period in liquid cultures grown in 96-well plates as detailed in section 2.3.2 using an Eon microplate reader (BioTek). JVS-0255 (Δhfq) was seen to grow at a similar rate to WT during exponential phase but achieved a lower final optical density (Figure 6.4). All strains harbouring pWKS30 with or without an insert grew noticeably slower than JVS-0255 and finished at a marginally lower final density, showing that the plasmid itself had an inhibitory effect on growth. Of the five genes tested, the only one to have a positive effect on growth was *S. enterica* *hfq*.

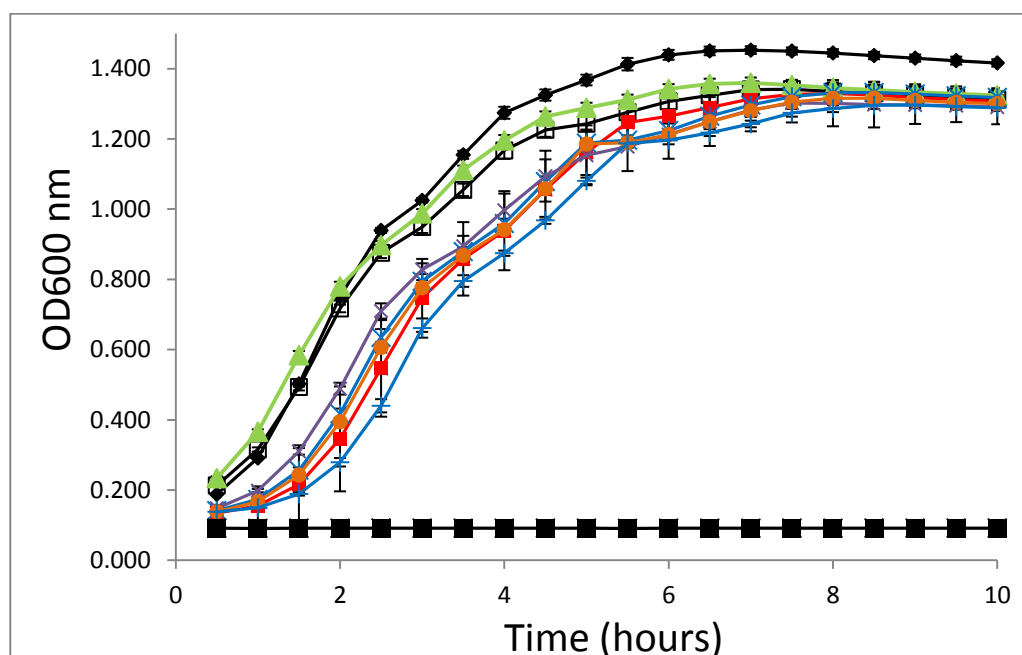


Figure 6.4 Expression of candidate genes *sco5592*, *sco5783*, *sco5792* and *sco3767* in the vector pWKS30 did not complement the reduced growth phenotype of a *Salmonella* Δhfq mutant while expression of *Salmonella* *hfq* in the same vector did increase growth. WT (SL1344): closed black diamonds; Δhfq (JVS-0255): open squares; Δhfq pWKS30 empty vector: closed red squares; Δhfq pWKS30-*Se-hfq*: closed green triangles; Δhfq pWKS30-*sco5592*: purple diagonal crosses; Δhfq pWKS30-*sco5792*: blue asterisks; Δhfq pWKS30-*sco3767*: closed orange circles; Δhfq pWKS30-*sco5783*: blue vertical crosses; uninoculated L medium: closed black squares.

Therefore none of the *S. coelicolor* candidates were shown to complement an Δhfq mutant when expressed in trans, although the *S. enterica hfq* gene did. There are several possible explanations for this: that the properties of the protein used to facilitate sRNA-mRNA interactions in Actinomycetes is very different from those of Hfq (e.g. a different RNA motif is recognised); that the *S. coelicolor* genes were not expressed well from this vector; that the Hfq-like activity is carried out by a protein other than these four candidates. Only a single replicate of this experiment was performed therefore it would be beneficial to carry out more replicates to confirm the result. It would also be valuable to construct a wild-type strain harbouring the empty vector pWKS30 as a control because the magnitude of the effect on growth of the plasmid was significant (surprisingly, it was even greater than the effect of the Δhfq mutation).

Of the three candidates selected from the proteomic data, *sco3767* belongs to a class of proteins not known to bind RNA and *sco5783* is part of a two-gene operon with a possible transmembrane transport protein. *sco5592* is the most credible candidate for a novel global RNA chaperone as it is situated amongst translation-related genes (section 6.5) and is annotated as containing an RNA-binding motif (section 6.6), therefore it was selected for further investigation as a putative novel global RNA chaperone.

6.5 Microsynteny of *sco5592*

In order to gain clues as to the function of SCO5592, its genomic context of the gene encoding it was examined. The four genes surrounding *sco5592* – *rpsP*, *rimM*, *trmD* and *rpsL* – form an evolutionarily highly conserved region of microsynteny, with the genes found in the same order and orientation in other species of *Streptomyces* and in the distantly-related species *E. coli* and *Bacillus*. Unusually, in *Streptomyces* the cluster is interrupted between *rpsP* and *rimM* by *sco5592* (Figure 6.5). This location suggests that SCO5592 is involved in processes governing translation.

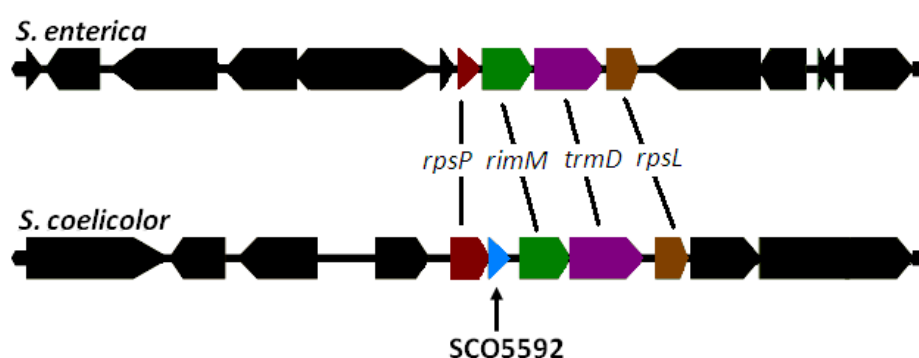


Figure 6.5 The highly conserved operon containing the translation-related genes *rpsP*, *rimM*, *trmD* and *rpsL* is interrupted by the uncharacterised gene *sco5592* in *Streptomyces* species.

6.6 Sequence motif analysis of SCO5592

Homology searches show that SCO5592 has similarity to other small proteins which bind RNA via a KH domain. This is a small and degenerate motif which is found in many RNA-binding proteins with diverse functions, which commonly combine multiple KH domains with other RNA-binding domains.

Multiple sequence alignment in ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) of the amino acid sequence of SCO5592 against other KH domain-containing proteins supports the notion that SCO5592 contains a KH motif. Importantly, the protein contains two conserved and functionally important glycines separated by one basic and one non-basic residue formed the core of this motif, with several nearby aliphatic residues also being conserved (Figure 6.6).

SCO5592	RVLEVRVHPDDLGVIGRNGRTARALRTVVG
Xl-hnRNP1	VELRILLQSKNAGAVIGGGKMICALRTDYN
Xl-hnRNP2	ITTQVTIPKDLAGSIIGGGQRIKQIRHRSG
EC-S3	KSIRVTIHTARPGIVIGKKGEDVEKLRKVVA
EC-PNP	RIHTIKINPDKIKDVIGGGSVIRALTEETG
Gg-vigilin	RTKDLIIIEQKFHRTIIGQKGERIREIREKPP
EC-CsrA	VGETLMIGDEVTVTLGVKGNQVRIGVNAPK
EC-NusA2	HTMDIAVEAGNLAQAIGRNGQNVRLASQLSG

Figure 6.6 Alignment of the amino acid sequence of SCO5592 with KH domains sequences from other proteins. Conserved glycines are shaded in grey; conserved aliphatic residues are shaded in cyan; conserved basic residues are shaded in pink.

Additionally, predictions of the secondary structure of SCO5592 were made using the software SCRATCH Protein Predictor SPP4.5 (available from <http://scratch.proteomics.ics.uci.edu/>). KH domains have a β - α - α - β structure, with the two conserved glycines forming “caps” at the inner ends of a pair of alpha helices (Grishin, 2001). The C-terminal KH domain of SCO5592 appears to have a similar structure, although the first alpha helix was not strongly predicted by this program. The N-terminal extension was predicted to have an α - β structure which is typical of a type II KH domain (Grishin, 2001).

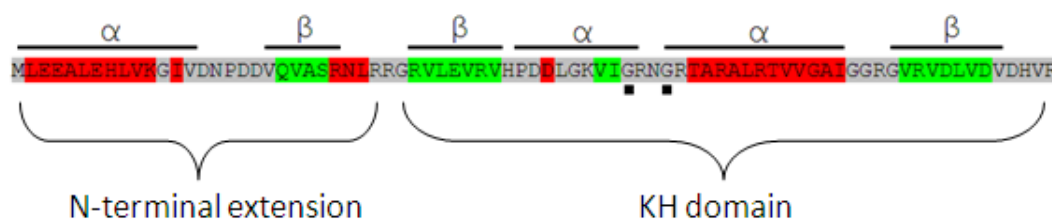


Figure 6.7 Graphical representation of the secondary structure predicted for SCO5592 by the SCRATCH Protein Predictor. Alpha helices are shaded in red; beta sheets are shaded in green. The two conserved glycines representing the core of the KH domain are marked by black squares.

Like SCO5592, the *E. coli* global RNA chaperone CsrA contains a single KH domain. CsrA binds as a homodimer to the 5' UTRs of targets mRNA transcripts to modulate their stability but can alternatively be sequestered by the sRNA CsrB to which it also binds (Mercante *et al.*, 2009). To determine whether SCO5592 is likely to be a member of this class, its amino acid sequence was aligned against those of CsrA and its homologue RsmA from *E. carotovora*. The sequences were not strikingly similar and the N-terminal extension of SCO5592 made it considerably larger than CsrA/RsmA (Figure 6.8), suggesting that they may not be very closely related. Also the reciprocal BLAST search used in chapter 3 to detect orthologues reported SCO5592 as being absent from *E. coli*.

```

CsrA                      MLILTR-RVGETLMIGDEVITVLGVKGNQVRIGVNAPKEVSVHREEIYQRIQAEKSQQSSY 61
RsmA                      MLILTR-RVGETLMIGDEVITVLGVKGNQVRIGVNAPKEVSVHREEIYQRIQAEKTTQISY 61
SCO5592  MLEEALHLVKGIVDNPDDVQVASRNLRGRVLEVRVHPDDLQ-KVIGRNGRTAR-ALRTVVGAIIGRGVRVDLVDVDHVR----
              * * * * * : *:: .*: * :*. .* .:: . * : :::: :

```

Figure 6.8 Alignment of the amino acid sequences of SCO5592 (*S. coelicolor*) with the sRNA-binding KH-domain proteins CsrA (from *E. coli*) and RsmA (from *E. carotovora*), showing that SCO5592 has an N-terminal extension and relatively low sequence similarity to these regulators.

While SCO5592 was not a clear orthologue of CsrA, it is possible that it interacts with mRNAs and sRNAs in a similar fashion via its type II KH domain. As the N-terminal extension is not basic it is tempting to speculate that it is involved in protein-protein interactions, for example in forming higher-order oligomers rather than a part of the RNA-binding interface or in interactions with signalling proteins in a similar manner to the eukaryotic KH domain protein Sam68 (Lukong and Richard, 2003).

6.7 $\Delta 5592$ mutant phenotype

A $\Delta 5592$ mutant was generated and gene replacement was confirmed by PCR (see section 2.8.5). Two independent clones were tested ($\Delta 5592.Q$ and $\Delta 5592.1$) and both are shown here. It was unfortunately not possible to generate complemented and over-expression strains due to time constraints.

The phenotypic consequences of *sco5592* deletion were examined by comparing the properties of the mutant strain with the parental M145 strain when grown on the agar media SFM, R5+ and DNA and in the liquid medium TSB:YEME34%. See section 4.3 for further explanation of the media selected.

In liquid culture the two clones of $\Delta 5592$ grew at a similar rate as M145 (WT) until around 50 hours, at which time its growth terminated while M145 continued to grow (Figure 6.9). By 150 hours the final density of $\Delta 5592$ was only around two-thirds that of M145. The lower final density may indicate that the mutant has defects in tolerating the stresses experienced towards the end of vegetative growth (as is the case for Δhfq mutant) or that $\Delta 5592$ is unable to catabolise carbon sources associated with stationary phase such as maltose (Novotna *et al.*, 2003).

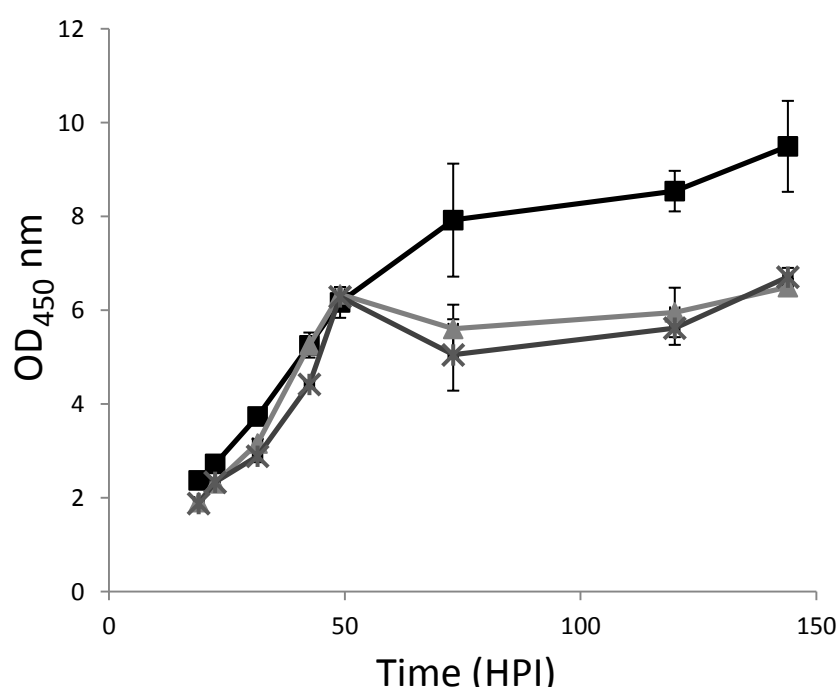


Figure 6.9 Growth curve in liquid TSB:YEME34% of M145 and $\Delta 5592$. M145: closed squares; $\Delta 5592.Q$: closed triangles; $\Delta 5592.1$: crosses. $\Delta 5592.Q$ and $\Delta 5592.1$ are two independent clones of $\Delta 5592$.

Production of antibiotics by $\Delta 5592$ was also reduced in liquid culture in comparison to M145. The onset of production of undecylprodigiosin was delayed by around 20 hours compared to M145 (Figure 6.10, above). This delay was not simply a result of the observed changes in growth rate because the difference in undecylprodigiosin concentration is seen earlier than the difference in OD₄₃₀, by around 10 hours. A similar pattern was seen for actinorhodin, again with a delay of approximately 20 hours (Figure 6.10, below).

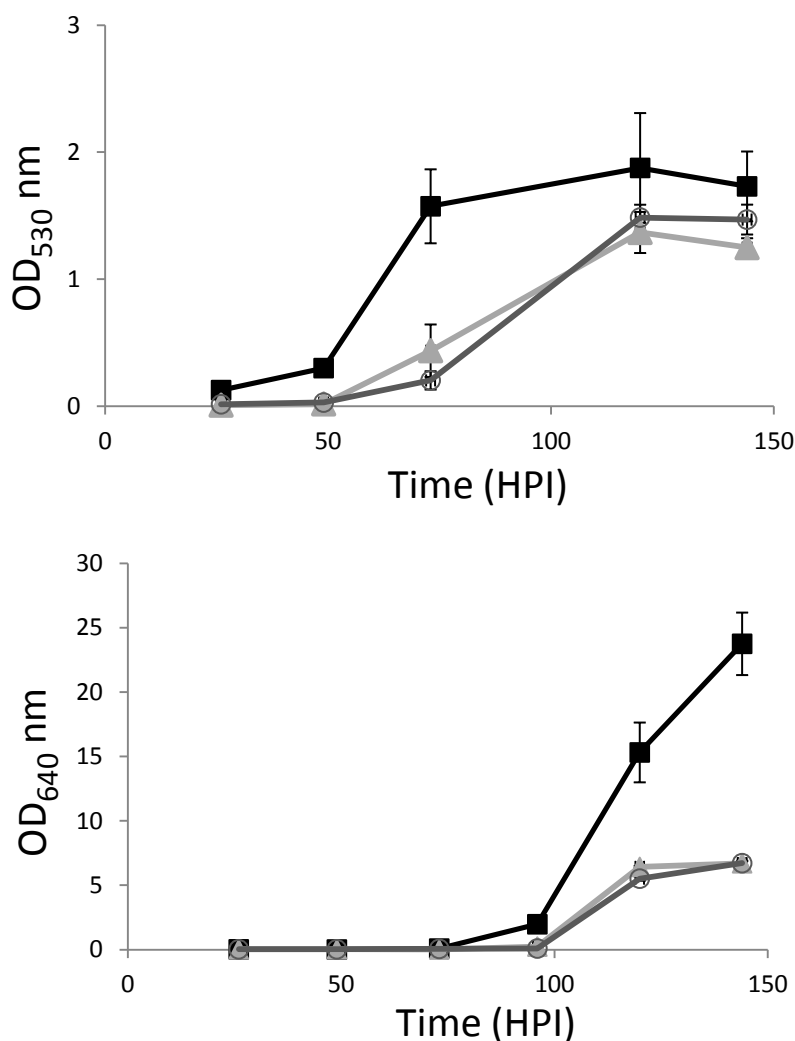


Figure 6.10 Production of undecylprodigiosin (above) and actinorhodin (below) in liquid TSB:YEME34% of M145 and $\Delta 5592$. M145: closed squares; $\Delta 5592.Q$: closed triangles; $\Delta 5592.1$: open circles. $\Delta 5592.Q$ and $\Delta 5592.1$ are independent clones of $\Delta 5592$.

There was little difference seen between the growth and morphological differentiation phenotypes of $\Delta 5592$ and M145 on SFM (Figure 6.11), the only dissimilarity being that the spores of $\Delta 5592$ had a darker, brownish shade that was difficult to capture in photographs, especially at earlier time points. $\Delta 5592$ also produced less actinorhodin than M145 on its upper surface and in the form of guttates (exuded droplets rich in secondary metabolites).



Figure 6.11 $\Delta 5592$ produced browner spores on SFM plates and secreted less actinorhodin from the surface of the colonies. $\Delta 5592.Q$ and $\Delta 5592.1$ are independent clones of $\Delta 5592$.

On R5+ plates, $\Delta 5592$ took somewhat longer than M145 to erect aerial mycelium and fully sporulate. By day 3, dark grey spores were seen across the surfaces of M145 single colonies as well as the confluent lawn. In contrast, $\Delta 5592$ growth and actinorhodin production in the less-dense regions was delayed. Single colonies were most severely affected and appeared bright red. By day 5 $\Delta 5592$ resembled M145 in terms of growth and sporulation but less actinorhodin was produced, particularly from the densely-plated part of the plate.

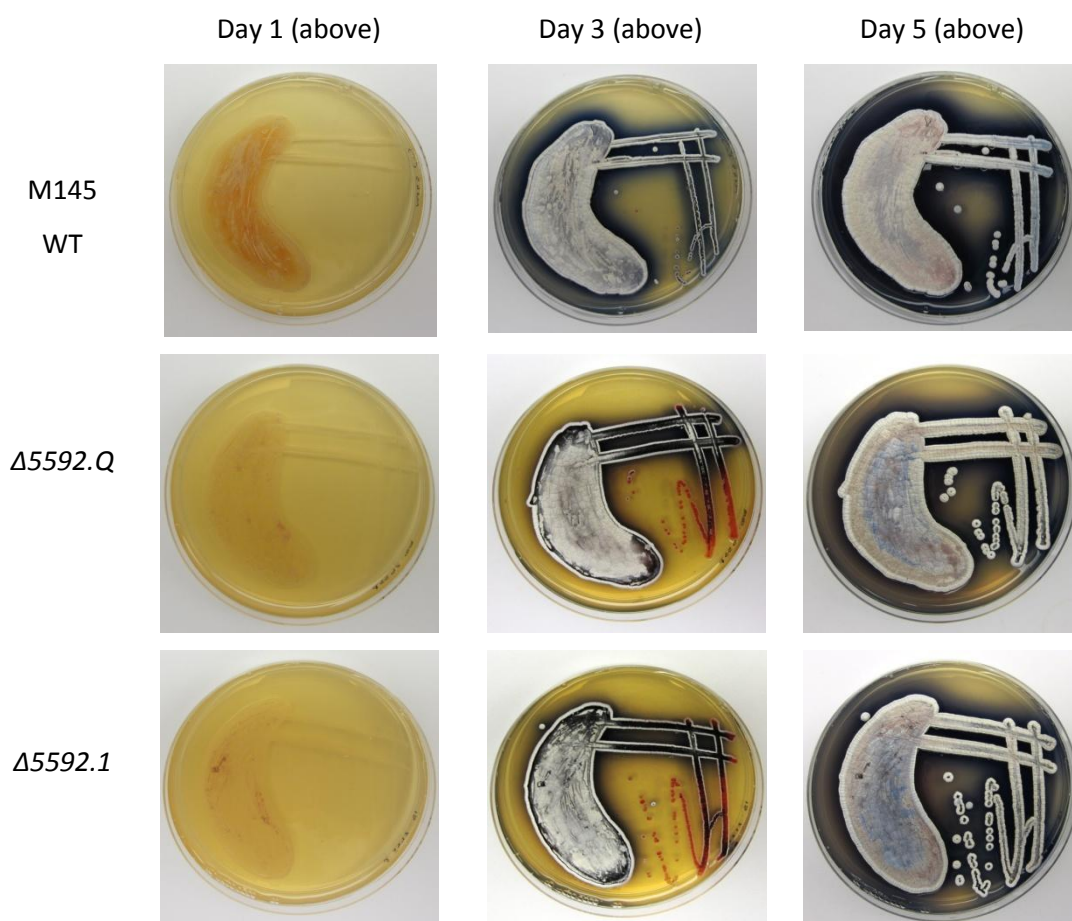


Figure 6.12 $\Delta 5592$ grew more slowly and produced less actinorhodin on R5+ plates. $\Delta 5592.Q$ and $\Delta 5592.1$ are independent clones of $\Delta 5592$.

On DNA medium, $\Delta 5592$ grew at a similar rate to M145 and had produced a similar amount of actinorhodin by day 3. However the extent of aerial mycelium formation was much greater and covered much of the confluent lawn by day, while M145 showed only minimal aerial mycelium at the edges of the growth area.

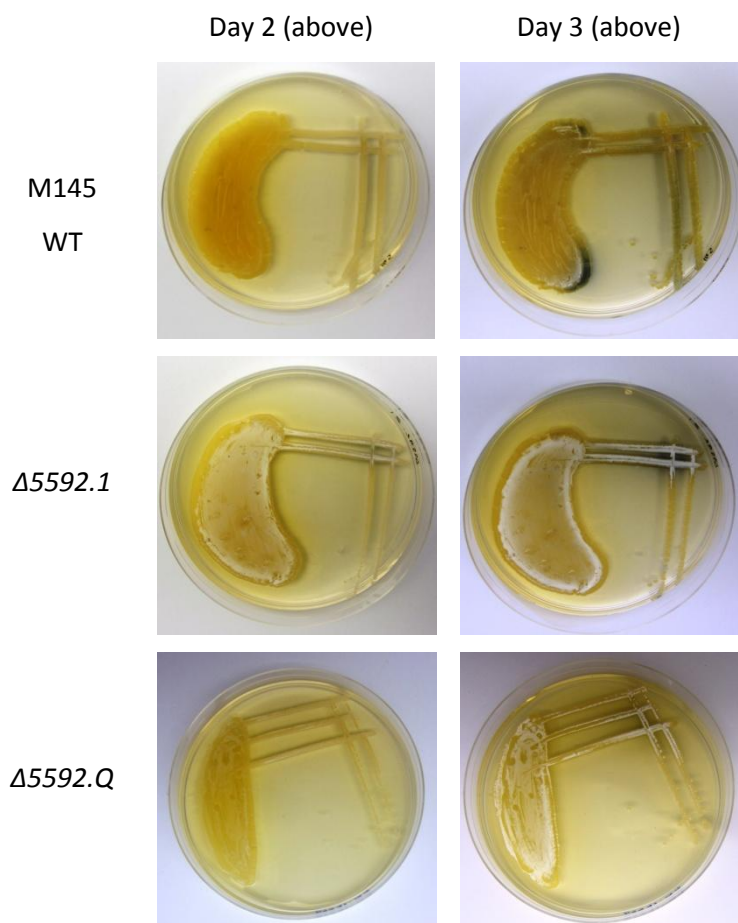


Figure 6.13 $\Delta 5592$ sporulated earlier and more fully on DNA plates. $\Delta 5592.Q$ and $\Delta 5592.1$ are independent clones of $\Delta 5592$.

In summary, $\Delta 5592$ showed subtle differences to the parental strain: it grew to a lower final density in rich liquid medium, with delayed undecylprodigiosin and delayed, reduced actinorhodin production. The deletion mutant grew relatively normally on agar media with some minor differences: there was a slight delay in the onset of morphological differentiation on R5+ and the extent of aerial mycelium formation was much greater on DNA medium. A minor reduction in actinorhodin production was also seen on the agar media SFM and R5+. A complemented strain would be highly desirable for this mutant, particularly as it is part of an operon and insertion of the apramycin cassette may have caused polar effects on the expression of downstream genes. Complementation would be judged on the two clearest phenotypes shown here: low final density in TSB:YEME34% and increased extent of aerial mycelium formation on DNA medium.

6.8 Purification and tag removal of recombinant His-5592

A recombinant His-tagged version of SCO5592 was produced in *E. coli* and purified using nickel affinity chromatography as detailed in section 2.9.5. It was anticipated that this protein could be used for *in vitro* studies to determine the binding and oligomerisation properties of SCO5592. To achieve this the SCO5592 gene was cloned into vector pET15b (Novagen) between the *Nde*I and *Bam*HI sites, adding the sequence “MGSSHHHHHHSSGLVPAGSH” to the N-terminal of the resulting protein, and the insert was confirmed by Sanger sequencing. The resulting construct was transformed by electroporation into BL21(DE3) pLysS freshly for each round of purification. His-5592 was found to be produced in greatest quantities when induced with 0.1 mM IPTG at an OD₆₀₀ of 0.6, although induction was good under all conditions tested and it was mostly present in the cell lysate supernatant (i.e. it was soluble).

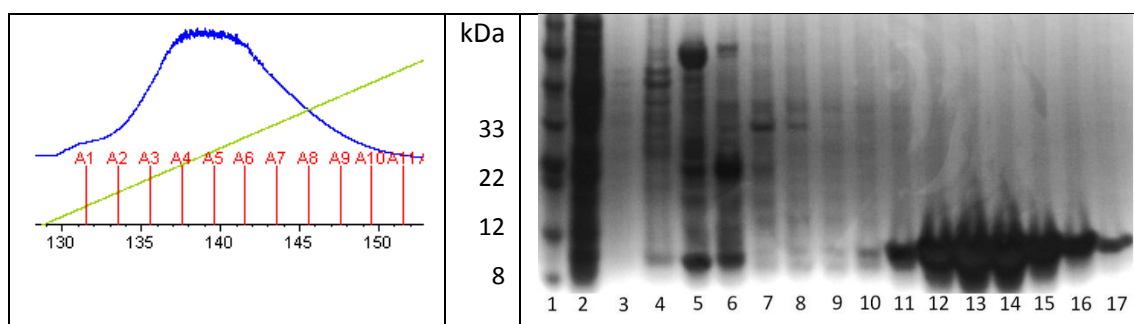


Figure 6.14 Left: typical elution profile of His-5592 from a nickel affinity chromatography column. Fractions collected are labelled in red (A1 – A11). Blue = absorbance at 280 nm; green = concentration of imidazole; scale = volume eluted (ml). Right: SDS-PAGE of fractions collected during purification. Lane 1 = Expedeon RunBlue Prestain markers (Broad range); lane 2 = clarified lysate; lane 3 = pre-loading column wash; lanes 4 - 6 = flow-through fractions; lanes 7 - 9 = wash fractions; lanes 10 - 17 = fractions A2 - A8 containing His-5592.

His-5592 was purified by nickel affinity chromatography on a 1 ml HisTrap HP column (GE Healthcare) using an ÄKTA Purifier FPLC system (Amersham Biosciences), with the buffers and method described in sections 2.4 and 2.9.5 respectively. Briefly, the IPTG-treated bacterial pellet from 1 litre of culture was harvested by centrifugation, frozen at -80°C and then thawed in His buffer A (25 mM imidazole) containing proteinase inhibitors and 10 µg/ml lysozyme, then the cells were lysed using a French press at 1,000 psi. The resulting lysate was clarified by centrifugation and syringe filtration (0.22 µm) before being loaded onto the column, which was then washed with His buffer A until the trace flattened. His-5592 was eluted using an imidazole gradient (25 mM - 500 mM), with the bulk of the protein eluting at around 125 - 250 mM imidazole (Figure 6.14, left). Eluted protein was >95% pure (Figure 6.14, right) and recovered at a good concentration (typically 5-10 mg/ml). However it was found to precipitate heavily during

overnight dialysis so, following elution, protein was immediately exchanged into imidazole-free cross-linking buffer or recombinant protein storage buffer (recipes in section 2.4) using a PD-10 Desalting column (Amersham Biosciences) equilibrated with 30 ml of the same buffer.

As the tertiary structure of SCO5592 was not known, the possible effects of the tag on binding activity were also unknown. Attempts were made to remove the His tag by cleavage with Novagen thrombin (see section 2.9.6) but the enzyme was found to unexpectedly cleave SCO5592 at an unknown internal site, making it impossible to obtain a pure solution of de-tagged protein (Figure 6.15). On-column cleavage (see section 2.9.6) for 2 hours was attempted but was also unsuccessful as no protein was released into the buffer, although it could be eluted with 500 mM imidazole, suggesting that the conformation of column-bound protein was sterically unfavourable for cleavage.

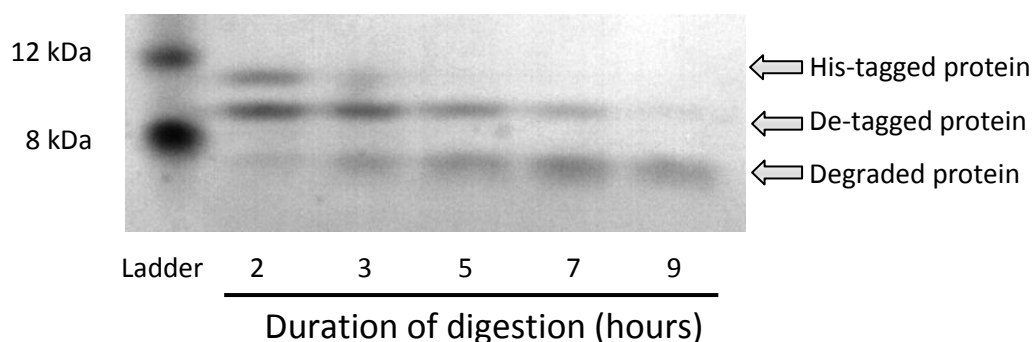


Figure 6.15 Cleavage of recombinant His-5592 in solution with thrombin (Novagen) effectively removed the His tag but also degraded the protein by cleavage at an unknown internal site.

6.9 Measuring the oligomerisation state of His-5592

Different KH domain proteins are known to form oligomers of different sizes via protein-protein interactions between the KH domains themselves (Chen *et al.*, 1997). For example CsrA forms dimers which are able to further oligomerise: when two dimers bind within the 5' UTR of transcripts at sites 10 – 63 nt apart these then interact to form a tetrameric “bridge” (Mercante *et al.*, 2009). In contrast the eukaryotic single KH domain protein Sam68 forms trimers (Chen *et al.*, 1997). Experiments were carried out to determine whether SCO5592 forms oligomers and, if so, to estimate the number of subunits per oligomer.

Chemical cross-linking of protein combined with SDS-PAGE was used as has been done previously in studies on Hfq and Sam68 proteins (Argaman *et al.*, 2012; Chen *et al.*, 1997). This approach has the advantage of capturing intermediate species from which it is sometimes possible to deduce information about the arrangement of monomers within the oligomer, for example whether hexamers are cyclic or a dimer-of-trimers (Azem *et al.*, 2010; Hajdu *et al.*, 1976).

His-tagged SCO5592 was prepared and immediately desalted using a PD-10 column and eluted into crosslinking buffer, then stored on ice and used for cross-linking experiments within two hours. Aliquots of this protein were incubated for 30 minutes at room temperature with various concentrations of either formaldehyde or glutaraldehyde. Formaldehyde is the mono-functional aldehyde crosslinking agent commonly used for ChIP studies while glutaraldehyde is a larger, bi-functional crosslinking agent more commonly used for protein-protein interaction studies. With formaldehyde cross-linking, a very small proportion of the protein (< 5%) was seen in a band approximately three times the weight of the monomer. The intensity of this band increased slightly as the concentration of formaldehyde was raised to a final value of 1%. With glutaraldehyde cross-linking a different pattern was seen: the intensity of the 29 kDa band increased and three additional bands of higher mass were visible (Figure 6.16) at approximately 35 kDa, 48 kDa and 60 kDa, suggesting the largest oligomer in solution may possess more than four subunits (which has not previously been reported for a KH domain). This also showed that the 29 kDa band is more likely to be a dimeric form which migrated slowly than to represent the trimeric form (Figure 6.17). Unfortunately both crosslinking agents caused the protein to precipitate heavily a few minute after addition so that it was not possible to get the reaction to run to completion, therefore the total number of monomers in the complex was not determined and data about the arrangement of subunits could not be obtained.

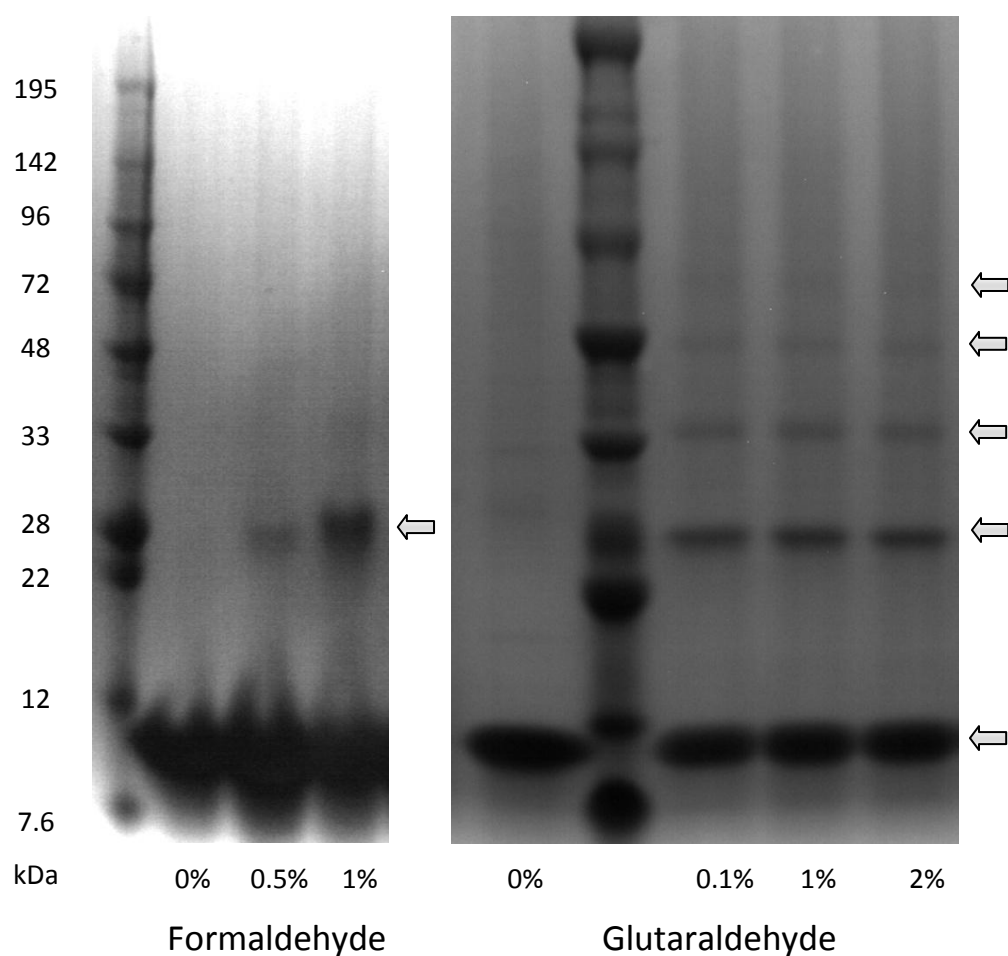


Figure 6.16 Cross-linking agents were able to capture some oligomeric states of recombinant His-5592 in solution. SDS-PAGE gels of purified recombinant His-5592 treated with 0 – 1% formaldehyde (left) or 0 – 2% glutaraldehyde (right). Bands representing different cross-linked oligomeric states of His-5592 are marked with white arrows.

Oligomer number	Predicted mass of oligomer (kDa)	Distance migrated in Figure 6.16 (kDa)
1	10.8	10.5
2	21.5	29
3	32.3	35
4	43	48
5	53.8	60

Figure 6.17 Predicted and observed band masses (kDa) for oligomers of His-5592.

A second approach, gel filtration (size exclusion chromatography), was used to measure the oligomer size of His-5592 in solution. This works on the principal that smaller particles travel a longer path through a column packed with porous beads, so that the largest particles are eluted

first. An Äkta Xpress system was used to purify His-tagged SCO5592 by nickel affinity chromatography and subsequently perform gel filtration using a HiLoad Superdex 75 (26/60) preparative grade column (GE Healthcare). Protein from the supernatant of the cell lysate from 500 ml culture was purified on a 5 ml HisTrap HP nickel column using the same His buffers as before. When His-5592 began to elute, a 5 ml aliquot of the eluate (corresponding to 1.6 % of the bed volume of the column) was passed onto the gel filtration column which had been equilibrated in gel filtration buffer (20 mM HEPES pH 7.5 / 150 mM NaCl). The gel filtration column was operated at 4°C at a flow rate of 0.5 ml/min and the effluent was recorded at 280 nm. A single protein-containing peak was detected at 116.87 ml (Figure 6.18). The approximate molecular mass of this oligomer was calculated to be 120,401 Da using the formula “ $\log MW = (-1.241 \times V_e) / (110.7 + 6.3908)$ ” (where V_e represents the elution volume of SCO5592), based on a calibration curve generated from Bio-Rad gel filtration standards. While the structure (and therefore the hydrodynamic radius) of SCO5592 is not known, it is likely to be relatively globular so an oligomer migrating at 120 kDa would likely represent a complex containing 10 - 12 monomers. This is a surprisingly high number of subunits for a single KH domain protein, which usually form dimers, trimers or tetramers.

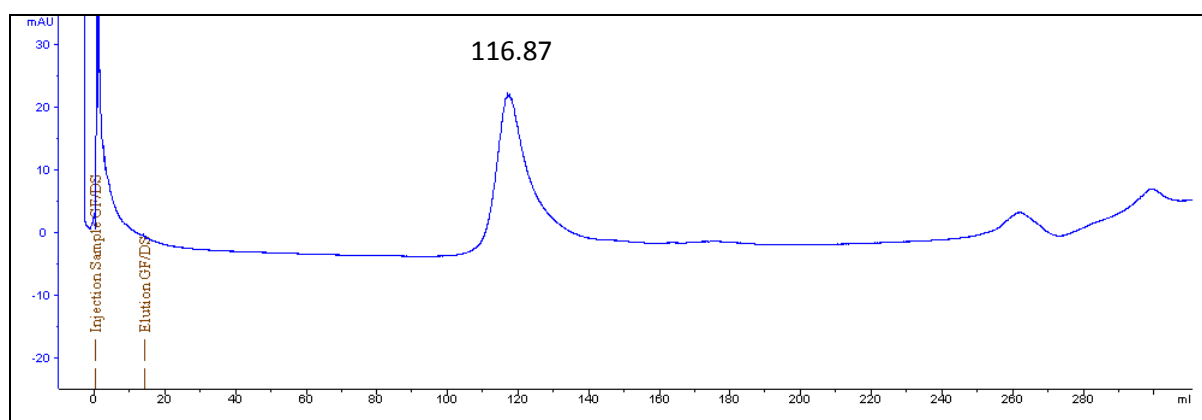


Figure 6.18 Gel filtration profile of His-5592 showing UV absorbance (mAU) over volume eluted (ml). One strong peak eluted at 116.87 ml, corresponding to an oligomer of 10 - 12 subunits.

While the standards used here covered the range 1.35 kDa – 670 kDa), the Superdex 75 26/60 column is designed primarily to measure particles of less than 70 kDa so this experiment would need to be repeated on a larger column such as the HiLoad 16/600 Superdex 200 preparative grade column (GE Healthcare) to be considered accurate, ideally in several experiments using different quantities of protein to determine whether there are any concentration-dependent effects on oligomerisation. It would also be beneficial to run the protein on a native gel (i.e. under non-denaturing conditions) as a second line of evidence to confirm the size of the multimer.

6.10 Discussion

Of the four candidates for an Hfq-like protein identified by the proteomics experiments described in chapter 3 (*sco5592*, *sco5783*, *sco3767*) and by linkage to *miaA/hflX* (*sco5792*), none were able to complement the growth defect of an *S. enterica* Δhfq strain and therefore it is unlikely that any of them are an Hfq of atypical sequence or a direct functional equivalent of Hfq. These results support the notion that Hfq is indeed absent in Actinobacteria.

SCO5592 was selected as a candidate for a sRNA-interacting protein of a non-Hfq-like mechanism due to its genomic context of its gene among many genes involved in translation and its possession of a single KH domain. This domain is shared by many RNA-binding proteins in both prokaryotes and eukaryotes including the *E. coli* global RNA chaperone CsrA, which is slightly smaller than SCO5592 as it does not have an N-terminal extension.

The construction of a $\Delta 5592$ strain showed that this gene is not essential and that the mutant strain had a relatively mild phenotype. $\Delta 5592$ grew to a lower final density in liquid culture (although its growth rate prior to 50 hours was normal), with a reduction in production of undecylprodigiosin and actinorhodin. Further investigation will reveal whether this was due to poor tolerance of stressful conditions or the loss of ability to utilize specific nutrients such as maltose during stationary phase. Several media-specific phenotypes were seen on agar media: actinorhodin production was reduced on SFM; a mild developmental delay was seen for low-density regions on R5+; aerial mycelium was erected earlier and with a greater coverage than M145 on DNA. However the $\Delta 5592$ mutant was not complemented, therefore phenotypic changes due to polar effects on downstream genes cannot be ruled out. It is possible that SCO5592 is involved in responses to stresses which were not assayed here such as temperature shock, oxidative stress or various nutrient limitations.

Based on cross-linking and gel filtration experiments, it is likely that recombinant SCO5592 forms oligomers of around 10 – 12 monomers in the absence of RNA although further experiments are required to confirm this result. This is a larger oligomer than has previously been reported for either KH domain proteins (which form dimers, trimers or tetramers) or Hfq proteins (which generally form hexamers) but is strikingly similar to the oligomers formed by the 8.3 kDa *Bacillus* TRAP protein (*trp* RNA-binding attenuation protein), which controls transcriptional attenuation of the *trp* operon. In *B. subtilis* and *Bacillus stearothermophilus* TRAP forms a ring of 11 subunits (Antson *et al.*, 1994) while in *Bacillus halodurans* TRAP form a ring of 12 subunits (Chen *et al.*, 2011). It is possible that SCO5592 oligomerizes and binds to RNA in a similar manner although they do not share sequence identity.

While SCO5592 is certainly an intriguing protein with a reasonable chance of proving to be a global RNA chaperone, several crucial experiments remain to be done. Most importantly, the ability of SCO5592 to interact with RNA species from *S. coelicolor* must be demonstrated. This could be achieved by co-immunoprecipitation of RNAs with FLAG-tagged SCO5592 using anti-FLAG antibodies followed by identification by a cloning or sequencing approach. These sRNAs could then be synthesized and tested *in vitro* for interaction with purified SCO5592, for example by EMSA or by RNaseE protection assays. If SCO5592 does indeed interact with multiple mRNAs and/or sRNAs it would be highly interesting to determine the structure of the SCO5592-sRNA complex to investigate the mechanism of binding and whether the N-terminal extension plays a role in oligomerisation.

Future work on this protein will be greatly facilitated by the *in silico* sequence analysis, mutant strain construction and purification protocol presented in this chapter.

Summary:

- 1) The four most promising candidates for a global RNA chaperone were selected on the basis of abundance in the nucleoid (from chapter 3) and on linkage to *miaA/hflX*. None of these complemented an *S. enterica* Δhfq mutant. SCO5592 was selected as the candidate taken forwards for further experiments as it is annotated as having an RNA-binding domain.
- 2) SCO5592 contains a single class II KH domain and interrupts a highly conserved gene cluster containing translation-related genes.
- 3) Genetic deletion of SCO5592 was found to cause several media-specific changes on agar media, including excessive sporulation on DNA. Maximum growth rate in liquid culture (TSB:YEME34%) was normal while final density was reduced.
- 4) His-tagged SCO5592 was purified from *E. coli* by nickel affinity chromatography. It was tentatively shown to form large oligomers in solution.

Chapter 7 – DNase-seq as a tool to probe the structure of the *S. coelicolor* nucleoid

7.1 Introduction

To complement the “bottom-up” approaches taken in previous chapters, a “top-down” experiment to explore the overall structure of the nucleoid was undertaken. A robust assay for probing of structure of the nucleoid at a genome-wide level would firstly allow us to define the structural elements of the *S. coelicolor* nucleoid and secondly would allow us to monitor the effects on these features of perturbations such as NAP deletion and chemical epigenetic treatments (section 1.4.4).

A number of methods are already available for investigating genome-wide occupancy by proteins, which can be correlated with other datasets such as expression. ChIP-seq can be used to locate the binding sites of a particular protein over the whole genome. However this has several drawbacks: only one protein is measured at a time and results are limited by the availability of high-quality antibodies or the addition of epitope tags. An ingenious alternative method measured total protein occupancy across the genome by fixing the nucleoid with formaldehyde and separating densely-occupied DNA and sparsely-occupied DNA by phenol-chloroform extraction. A high degree of occupancy was found to correlate positively with binding of RNAP and to a lesser extent Fis, but not IHF of H-NS (Grainger *et al.*, 2006).

Structural changes can also be measured by techniques which do not rely directly on measuring protein occupancy, for example the use of gyrase-mediated fragmentation to estimate topological loop size (see section 1.1.3) or the use of the fluorescent intercalating agent psoralen to measure global levels of supercoiling (Bermudez *et al.*, 2010). To measure 3D structure the distances between pairs of loci can be measured by microscopy (Niki *et al.*, 2000), recombination assays (Valens *et al.*, 2004) or by 3C-based techniques (see section 1.1.2) but the former two give limited resolution while the latter has been applied genome-wide to *Salmonella* and *C. crescentus* (Umbarger *et al.*, 2011) but is technically challenging.

We wished to expand the toolkit available for nucleoid research by developing a robust protocol for bacterial DNase-seq, whereby the genome is fragmented using a non-specific endonuclease and the cut sites are mapped by high-throughput sequencing. Structural features are identified by

mapping the reads to the genome and looking for regions which are enriched or depleted in cut sites. The patterns seen in eukaryotic experiments are typically very pronounced, with hypersensitive sites (e.g. in promoters) being at least 10-fold more digestible than denser nucleosome-bound regions (Sabo *et al.*, 2006) and are readily detectable by clustering algorithms (Crawford *et al.*, 2006). This type of work is more challenging in bacterial genomes due to their much lower protein occupancy and debate over the types and sizes of features (e.g. topological loops) likely to be found, but a strong positive correlation between expression and DNase I digestibility has previously been found in *S. coelicolor* (McArthur and Bibb, 2006). Eukaryotic-like hypersensitive sites were not detected in their study so it is possible that digestibility is a property of the gene as a whole, or it could be that the sites are less distinct or simply not seen under those conditions.

Improvements in sequencing technology have made it possible to map nuclease cut sites to single base pair resolution across very large regions. In humans this is typically a chromosome arm or isochore but in bacteria this can be applied to the entire genome. DNase-seq would be a valuable addition to the range of available techniques because it gives information on the structural effects on accessibility rather than simply determining whether it is occupied by protein. It could be used in conjunction with other techniques such as ChIP-seq, microscopy, 3C, bottom-up experiments and computer modelling to answer important biological questions.

This chapter describes attempts to develop an assay for probing nucleoid structure based on partially digesting the nucleoid with an endonuclease, then mapping the cut sites using high-throughput sequencing. The method was attempted using both the non-specific endonuclease DNase I and the sequence-specific frequent cutting restriction enzyme *HaeIII*.

Objectives:

- 1) Adapt the DNase I fragmentation method of McArthur and Bibb into a high-throughput sequencing assay.
- 2) Determine whether DNase I sensitivity and expression are correlated across the genome.
- 3) Look for structural features such as macrodomains.
- 4) Use this method as an assay to monitor changes in these structures caused by factors such as growth phase, genetic deletion of NAPs and chemical perturbations.

7.2 System and approach

All experiments were carried out using *S. coelicolor* M145 (WT) grown to stationary phase (74 hours; mycelium appears blue-purple) in the rich liquid medium 50:50 TSB:YEME34% (Kieser *et al.*, 2000). It was anticipated that this time point would show the clearest difference between decondensed secondary metabolic clusters (for example the actinorhodin cluster) and condensed silent regions, although it was anticipated that some level of noise would be generated by heterogeneity of physiological state between those hyphae on the surface of pellets and those in the pellet interior. Additionally, fewer gene dosage effects around the origin of replication were expected at this time point as growth had ceased and most nucleoids would not have been actively dividing.

DNase I (37 kDa) was selected as the optimal endonuclease due to its favourable substrate DNA preferences: DNase I can cut at any site (complete digestion gives di-, tri- and tetra-nucleotides) but has a small preference for a purine-pyrimidine pair (Sambrook *et al.*, 2001) - AT, GT, GC or AC - which was not expected to create any systematic bias in this experiment. Micrococcal nuclease is often used in eukaryotic research due to its smaller size (17 kDa) but it has a 30-fold preference for AT-rich DNA over GC-rich DNA which can strongly influence the digestion pattern (Horz *et al.*, 1983) and was therefore deemed unsuitable. In McArthur and Bibb (2006), digestion was carried out by DNase I in the presence of magnesium ions. However this favours production of single-stranded cuts which would require extensive end-repair prior to library construction and may shift the apparent site of the cut (Figure 7.1) so manganese chloride, which favours double-stranded breaks or at most a 1-2 bp overhang (Sambrook *et al.*, 2001), was substituted here (Figure 1.7.1).

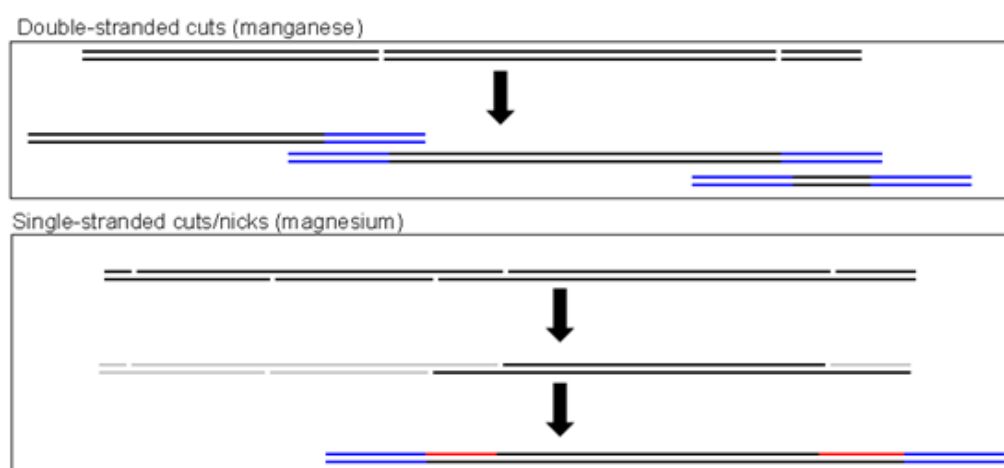


Figure 7.1 DNase I digestion in the presence of manganese ions favours the creation of double-stranded cuts (above) to which sequencing adaptors can be ligated directly, whereas the presence of magnesium ions favours single-stranded nicks (below) which require extensive end-repair prior to incorporation into the sequencing library. Black = genomic DNA; red = end-repaired regions; blue = sequencing adaptors.

A limiting DNase I digestion was expected to capture the most biological information as excessive digestion may begin to fragment even strongly protected regions and/or digest the most digestible regions into fragments too small to be incorporated into the sequencing library.

While standard methods are available for tasks such as read alignment and peak calling, it was anticipated that novel techniques would need to be developed to define nucleoid structures.

Following a visual inspection of the cutting patterns, preliminary exploration of the data were expected to include comparisons in cutting rate of the following: intergenic versus intragenic sequences; active versus silent genes (by comparison with microarrays data from the same culture); core genome versus arm domains (see section 1.1.7). If hypersensitive sites are present in *S. coelicolor* DNA, they were also expected to be detected by visual inspection of the cut patterns or by clustering algorithms in a similar fashion to experiments performed on eukaryotic chromatin (Crawford *et al.*, 2006). Their size and fold-enrichment could then be defined more rigorously and some correlation with biological function could be deduced.

Macrodomains are visible in transcriptomic data as fluctuating undulations on the 250 kb scale; they are expected to be visible here because expression is known to correlate with digestibility. However these patterns are typically noisy and cannot reliably be detected from one information source alone – the best method would be to combine data from DNase-seq, transcriptomics and other measures such as evolutionary conservation. In particular we wished to know whether *S. coelicolor* had more macrodomains or bigger macrodomains than *E. coli* as its genome is significantly larger.

While DNase I shows a modest preference for supercoiled over relaxed DNA, the cutting pattern generated in this assay was not expected to be able to reveal topological domains because the act of cutting the DNA would quickly release supercoiling. If there are equivalent structures to the H-NS-bound patches thought to delineate topological domains in *E. coli* (Noom *et al.*, 2007), these would likely be resistant to DNase I. 1 kb patches of highly nuclease-resistant DNA separated by 2-10 kb of more accessible DNA could be indicative of this type of process.

7.3 Preparation of DNase I-digested mycelium

Genome fragmentation was carried out using the method published in McArthur and Bibb (2006) with two modifications. Manganese chloride was substituted for magnesium chloride in the DNase I digestion buffer (see section 7.2). Also the duration of pre-digestion was increased from 60 to 90 minutes. Solutions used are listed in section 2.4.

Two biological replicates (MnX and MnY) were prepared in the presence of manganese chloride, as well as one biological replicate (Mg) prepared in the presence of magnesium chloride as a comparator to facilitate comparisons with previous work done by McArthur and Bibb (2006). Practice replicates (not shown here) which were used to optimise the methods showed that the degree of digestion achieved by the same amount of enzyme varied somewhat between biological replicates, presumably due to small differences in dispersal between different batches of mycelium and perhaps due to unquantifiable differences during handling (e.g. variable losses during ethanol precipitation). Within each biological replicate, samples were named according to amount of enzyme used per ml of original culture volume, e.g. a 1 ml aliquot of lysate representing 10 ml culture digested with 1 μ l enzyme (10 U enzyme) would be named 1 U.

M145 spores were germinated synchronously as described in section 2.3.4 and grown in 50 ml of 50:50 TSB:YEME34% (rich liquid medium) for 74 hours until they were in stationary phase and were producing both actinorhodin and undecylprodigiosin. 5 mM CaCl_2 , 100 U RNase I (Promega) and 800 U DNase I (Roche) were added to the culture, then incubation was continued at 30°C for 90 minutes as a pre-digestion step to remove extracellular nucleic acids which are present in mycelial pellets (Kim and Kim, 2004). Cells were harvested by centrifugation at 4,000 rpm for 5 min, then washed three times in TES+ buffer which contained 7.5 mM EDTA to stop the extracellular digestion by chelating the metal ions required for DNase I to function.

The mycelium was then resuspended in 5 ml TES buffer supplemented with 0.5 % each of the detergents Nonidet P40 and Triton X-100 to permeabilise the cell walls. Cells were incubated at 30°C for 15 min with occasional gentle inversion then washed in 50 ml DNase I digestion buffer and finally resuspended in 5 ml DNase I digestion buffer. The mycelium was split into 6 aliquots which were digested with either 0, 10, 30, 50, 80 or 200 U Roche DNase I for 3 min at 37°C. The reaction was stopped by the addition of 1 ml DNase I STOP buffer which contained 10 mM EDTA to inhibit DNase I. STOP buffer also contained 5% SDS to lyse cells and 10 mg/ml proteinase K (Roche) to remove the proteins from the DNA. Incubation continued at 55°C overnight. DNA was harvested in the same manner as HMW DNA (section 2.5.3), with the following purification steps: phenol-chloroform wash, chloroform-only wash, isopropanol precipitation, ethanol wash, 2-hour

RNase I-digestion, phenol-chloroform wash, chloroform-only wash, isopropanol precipitation, ethanol wash. DNA was finally resuspended in 100 µl EB buffer at a high concentration of around 1 - 5 mg/ml as estimated by the Nanodrop ND-1000 and stored at 4°C.

Gel electrophoresis was used to estimate the degree of fragmentation which had occurred, but the appearance of the gel was found to be highly dependent on factors such as DNA concentration, quantity of DNA loaded per well, gel percentage, voltage and distance migrated. In order to achieve sufficient solubility of DNA for reproducible gel electrophoresis it was necessary to further dissolve the samples by diluting the recovered DNA 1/10 with water (to a concentration of around 100 - 500 ng/µl) and incubating at 37°C for one hour. Without this step excessive vertical streaking and abnormal migration were seen. Figure 7.2 shows one biological replicate (MnX) where the relative proportion of low molecular weight (LMW) and medium molecular weight (MMW) fragments is clearly seen to increase as the quantity of enzyme used increased, but the apparent absolute size of the fragments appeared very different depending on the amount loaded. When a large quantity of DNA (15 - 20 µg) is loaded and run for a short distance (15 min at 80 V), a clear pattern of increasing digestion is seen: less material runs in the HMW band and the apparent average fragment size is reduced. However if less material is loaded (e.g. 2 - 5 µg), or if a lower percentage gel is used (e.g. 0.7 %), or even if the same gel is run for longer, then the pattern becomes less distinct and the proportion of DNA migrating as HMW increases.

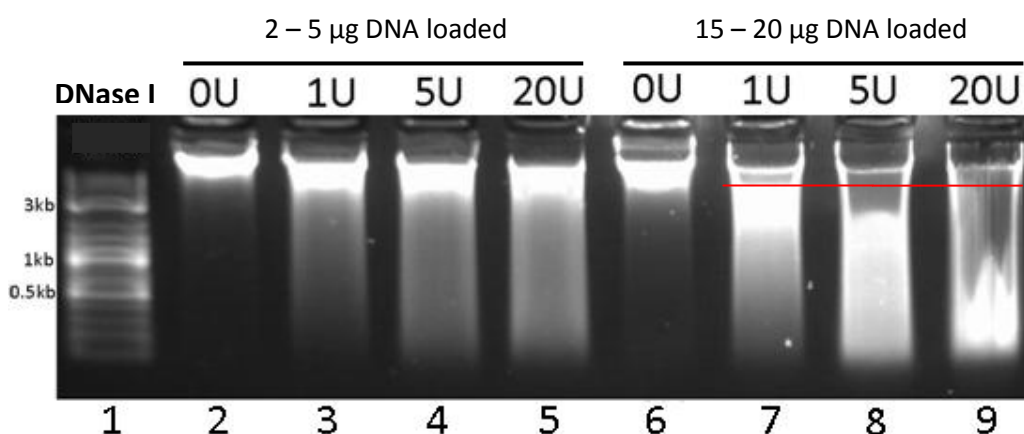


Figure 7.2 DNase I-digested mycelium showed greater fragmentation as the amount of DNase I increased, with digested samples having a smear of low-molecular-weight fragments. The appearance of the smear of fragments depended on the quantity of DNA loaded. Lane 1: NEB 2-log ladder; lanes 2 - 5: 2 - 5 µg DNA loaded (0U, 1U, 5U, 20 U DNase I used for digestion); lanes 6 - 9: 15 - 20 µg DNA loaded (0U, 1U, 5U, 20 U DNase I used for digestion). The red line indicates the cut-off point used for size fractionation by gel extraction. Biological replicate MnX is shown here. Vertical streaking in lane 9 was caused by small amounts of poorly-soluble DNA in that sample.

The smears of fragments were purified from the gel using the Qiaquick gel extraction kit. HMW DNA containing very few cuts was excluded (Figure 7.2, red line) in order to improve the signal to noise ratio, based on the premise that the smear of fragments released by digestion will be enriched for those whose ends have both been cut by the nuclease. DNA was eluted in 40 µl EB, giving a concentration of around 25 ng/µl, and stored at -20°C. Undigested DNA and total gDNA standards were all HMW and therefore could not be gel extracted, but were diluted to approximately 50 ng/µl and allowed to fully dissolve for at least two hours before use to maximise solubility.

7.4 qPCR experimental design

Differences in DNase I digestibility between high-expression and low-expression genes were previously found to be detectable by qPCR (McArthur and Bibb, 2006), so it was decided to use this type of assay to perform a quality control step on the DNase I-digested samples prior to committing them to sequencing. In this case, rather than measuring copy number loss relative to a standard (*rbcH* in McArthur and Bibb, 2006), gene copy numbers were compared directly between *afsK* and *sti1* in the DNase experiment and *afsK*, *sti1*, *hrdB* and *actII-ORF4* in the *HaeIII*-seq experiment. Approximately a 2-fold difference in copy number between *afsK* and *sti1* was taken as a benchmark of satisfactory digestion for DNase I treatment but the expected difference for *HaeIII* treatment could not be predicted in advance. DNase-seq primers (Figure 7.3) amplified regions of around 120 nt in the promoter regions of two genes: *afsK* (*sco4423*; protein kinase; constitutively low expression) and *sti1* (*sco0762*; protease inhibitor precursor; constitutively high expression). *HaeIII*-seq primers (Figure 7.4) amplified regions of 125 - 150nt in the coding regions of *afsK* and *sti1*, plus *hrdB* (*sco5820*; the primary vegetative sigma factor) and *actII-ORF4* (*sco5085*; a developmentally-regulated pathway-specific actinorhodin activator). Assay optimisation was carried out as detailed in section 2.7.2 and calculations were performed as detailed in section 2.7.4 - 2.7.6 using M145 gDNA standards.

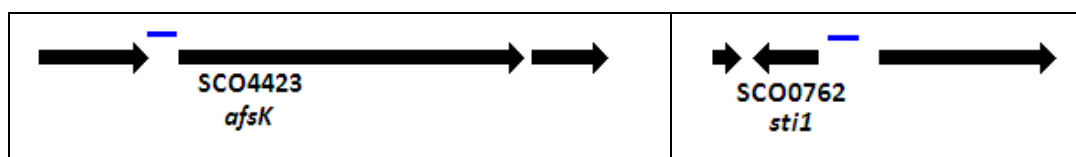


Figure 7.3 qPCR primers for DNase-seq in the promoter regions of the low-expression gene *afsK* (*sco4423*) and the high-expression gene *sti1* (*sco0762*). Amplicons (marked by blue lines) measured 125 nt and 121 nt respectively.

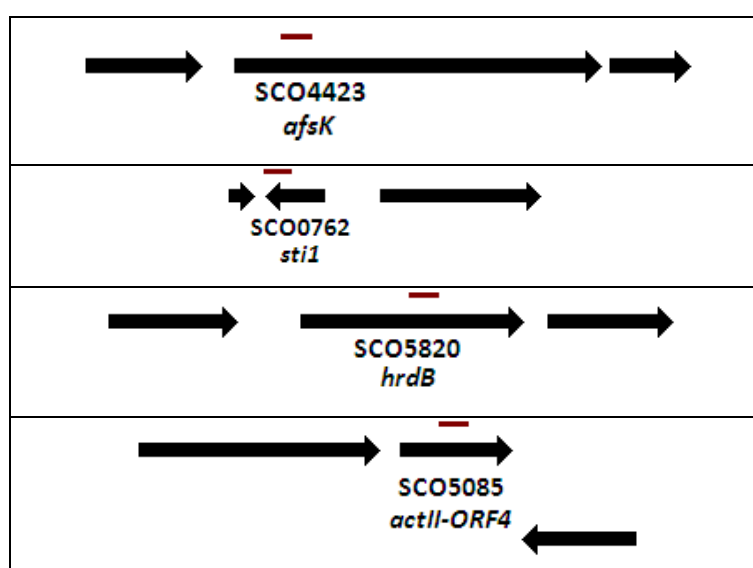


Figure 7.4 qPCR primers for *HaeIII*-seq in the coding regions of the low-expression gene *afsK* (*sco4423*), the high-expression gene *sti1* (*sco0762*), the vegetative sigma factor *hrdB* (*sco5820*) and the actinorhodin pathway transcriptional activator *actII-ORF4* (*sco5085*). Amplicons (marked by dark red lines) measured 147 nt, 138 nt, 126 nt and 126 nt respectively.

7.5 qPCR validation of DNase I-digested samples

In order to quantify the degree of differential digestion that had taken place, the copy numbers of two genes known to have different digestion rates were measured using qPCR (described in section 2.7 and 7.4). *sti1* (*sco0762*) is highly expressed throughout development while *afsK* (*sco4423*) is expressed at low levels throughout development, therefore the copy number ratio of *afsK* : *sti1* should increase as digestion increases.

A mock-digested control and a range of DNase I-digested samples were tested for each of three biological replicates: MnX and MnY were identical replicates prepared in the presence of manganese chloride while MgX was a separate biological replicate prepared in the presence of magnesium chloride.

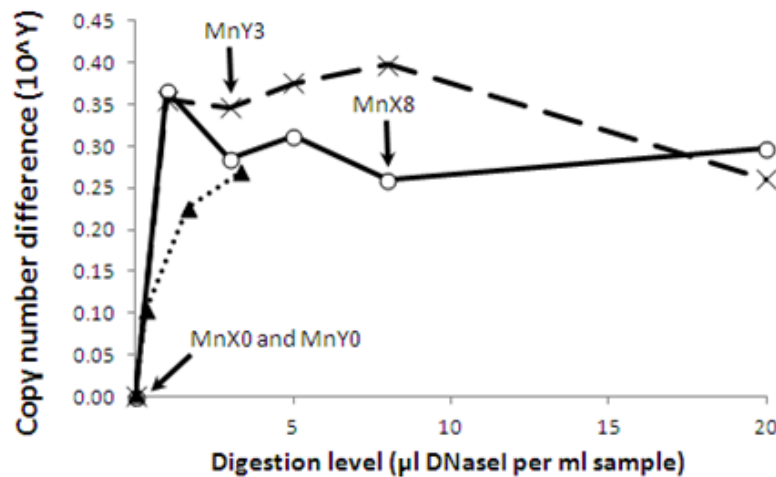


Figure 7.5 DNase I digestion was able to generate differential digestion between two genes of different expression level. qPCR showed that the copy number difference between the high expression gene *sti1* and the low expression gene *afsK* increased with digestion level to around 2-fold before stabilising. MnX (open circles) and MnY (crosses) are separate biological replicates of the experiment performed in the presence of manganese chloride while MgX (closed triangles) is a replicate of the experiment performed in the presence of magnesium chloride. Samples submitted for Illumina sequencing are indicated with arrows: MnX0 and MnY0 are mock-treated samples while MnX8 and MnY3 are samples treated with 3 µl and 8 µl DNase I (30 U and 80 U) respectively per ml of culture volume treated.

The copy numbers of *afsK* and *sti1* were calculated for each sample from the standard curves constructed for each of the genes (see section 2.7.6). For each biological replicate the ratios of all samples were normalised against the ratio seen for undigested samples (which was not zero due to small differences in efficiency of the two qPCR reactions). Copy number difference increased

quickly with increasing digestion level and soon levelled off at around $10^{0.3}$ genomes, corresponding to a 2-fold difference (Figure 7.5). This confirmed that the degree of differential digestion was similar between replicates and was comparable to that seen in McArthur and Bibb (2008). A second technical replicate of the qPCR using freshly gel-extracted template DNA showed the same pattern. MnX0, MnX8, MnY0 and MnY3 were selected for sequencing.

7.6 *Sma*I digestion to reduce the average fragment size

Illumina sequencing libraries are created from fragments falling within a size window of 250 - 550 bp. However the majority of the fragments produced by DNase I are considerably larger than this, so that the average size had to be reduced. ChIP fragments are generally reduced in size by sonication but this creates non-specific breaks which would be indistinguishable from the non-specific breaks created by DNase I, so a restriction digest was chosen instead to allow downstream removal of RE-generated reads. Desirable characteristics for this enzyme were thought to be 1) an easily recognisable cut site (i.e. a six-base cutter), 2) frequent cutting to give all fragments a good chance of getting into the size window, 3) a blunt-ended cut site to facilitate library construction. The genomic DNA insert of *S. coelicolor* cosmid StF85 was digested *in silico* with a wide variety of enzymes using the StrepDB restriction tool available from <http://strepdb.streptomyces.org.uk/>. The best enzyme found was *Sma*I, which creates a blunt cut in the frequently-occurring sequence CCC|GGG.

Partial restriction digest of MnX0, MnX8, MnY0 and MnY3 were performed by incubating approximately 5 µg each DNA sample with 5 µl *Sma*I (New England Biolabs) in a 250 µl reaction for 3 hours at 37°C. 50 µl of this was run on a 1.5% gel at 50 V for 60 min (Figure 7.6) and confirmed that size reduction had taken place. The remaining 200 µl was submitted to TGAC for incorporation into the sequencing library. Undigested material appears as HMW in this gel due to the variable migration patterns of these types of sample (c.f. less heavily loaded lanes in Figure 7.4 and gels run for longer in Figure 7.15).

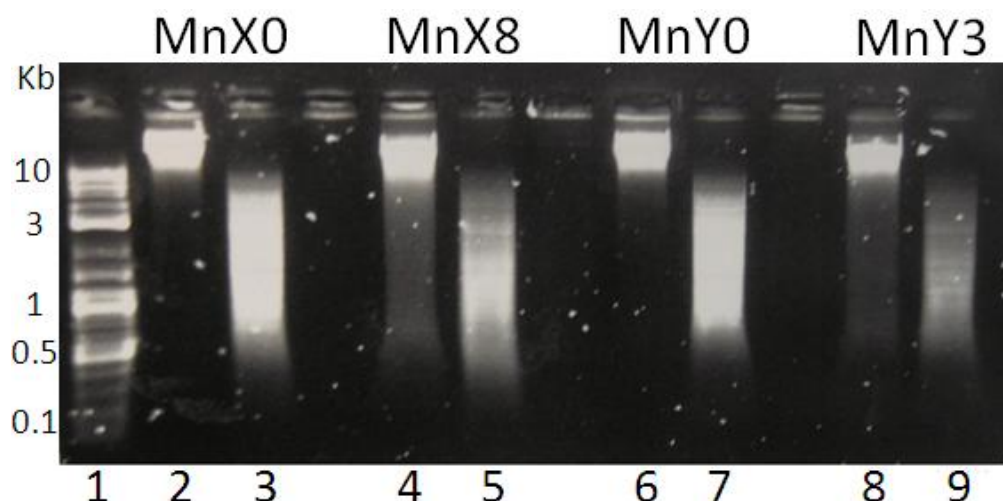


Figure 7.6 The 6-base cutter *Sma*I was used to reduce the average size of fragments in the mock-digested samples MnX0 and MnY0 and in the DNase I-digested samples MnX8 and MnY3. MnX and MnY represent separate biological replicates. Lane 1 = NEB 2-log ladder; lanes 2, 4, 6, 8 = non-*Sma*I-digested, non-gel extracted DNA; lanes 3, 5, 7, 9 = *Sma*I-digested, gel-extracted DNA.

Following *Sma*I digestion the concentration and average fragment size were measured by staff at TGAC using a Qubit® 2.0 Fluorometer (Invitrogen) and a 2100 Bioanalyzer (Agilent) respectively (Figure 7.7). These values corresponded reasonably well to the apparent sizes of the fragments shown in the agarose gel in Figure 7.6, which showed that a lot more material was now in the correct range but also that it had not been over-digested.

Sample	DNA concentration (Qubit® 2.0 Fluorometer)	Average fragment size (Agilent 2100 Bioanalyzer)
MnX0	218 ng/μl	1,631 bp
MnX8	135 ng/μl	979 bp
MnY0	394 ng/μl	1,313 bp
MnY3	184 ng/μl	1,365 bp

Figure 7.7 DNA concentration and average size of fragments used to construct the sequencing library. MnX0 and MnY0 are mock-digested samples while MnX8 and MnY3 are DNase I-digested samples shown by qPCR to be differentially digested. MnX and MnY are separate biological replicates.

7.7 Library construction and sequencing

Library construction and sequencing were carried out by staff at TGAC (Norwich Research Park) using Illumina technology. A multiplexed paired-end library was constructed, using a different adapter for each sample (MnX0 - 4; MnX8 - 5; MnY0 - 7; MnY3 - 10). These were then pooled and sequenced in one lane on an Illumina Genome Analyzer IIx platform to generate 100 bp reads.

7.8 DNase-seq read quality control and alignment

The quality of a read decreases along its length until the Phred quality score (calculated per base call) is no longer above 30. This point is known as the “Q30 limit” and represents the point at which the base calls fall below a 99.9% probability of the call being correct; reads are normally trimmed to this length prior to alignment. The reads generated in this experiment were of an unexpectedly low quality (Figure 7.8), with the Q30 limit being reached at 45 - 65 bp (read1) and 30 - 60 bp (read2) whereas it would normally be expected to fall around 90 bp. However the relatively small size of the *S. coelicolor* genome, and lack of extensive repetitive sequences, means that 30 bp is theoretically enough to uniquely align the majority of reads in this species.

Between 8.3 million and 38.5 million reads were produced per sample (Figure 7.8). The read1 and read2 sets were pooled for each, giving a total of between 16.7 million and 39.6 million reads per sample. Of these, only a very small proportion began with a sequence other than “GGG” (typically <5%), giving a set of roughly one million usable non-*Sma*I reads per sample for MnX0, MnY0 and MnY3 (Figure 7.9). The total number of reads and proportion of non-*Sma*I reads was much greater in MnX8 for unknown reasons, giving a much larger set of almost 20 million non-*Sma*I reads.

Sample filename (sample_adaptor_read1/2)	Reads	Q30 limit
X0_TGACCA_R1_001.fastq.gz	8,328,660	45 - 49
X0_TGACCA_R2_001.fastq.gz	8,328,660	30 - 34
X8_ACAGTG_R1_001.fastq.g	38,490,850	60 - 64
X8_ACAGTG_R2_001.fastq.gz	38,490,850	45 - 59
Y0_CAGATC_R1_001.fastq.gz	19,781,499	55 - 59
Y0_CAGATC_R2_001.fastq.gz	19,781,499	45 - 49
Y3_TAGCTT_R1_001.fastq.gz	16,936,857	50 - 54
Y3_TAGCTT_R2_001.fastq.gz	16,936,857	35 - 39

Figure 7.8 The number of reads and Q₃₀ limit of reads generated for samples MnX0, MnX8, MnY0 and MnY3 were unusually low.

Sample	Total reads	Number non- <i>Sma</i> I reads	% non- <i>Sma</i> I
X0	16,657,320	715,632	4
X8	76,981,700	19,912,867	26
Y0	39,562,998	1,999,182	5
Y3	33,873,714	1,134,465	3

Figure 7.9 The majority of reads were discarded as they began with the sequence “GGG”, indicating that the DNA cut they represented had been generated by *Sma*I rather than by DNase I.

Non-*Sma*I reads were aligned against the *S. coelicolor* genome by Govind Chandra using the short-read alignment tool Bowtie (Langmead *et al.*, 2009). Only around 30 - 50 % of the reads could be uniquely aligned, resulting in 368,553 (MnX0), 17,620,011 (MnX8), 793,486 (MnY0) and 634,759 (MnY3) reads ultimately being included in the alignment. Of the reads which could not be uniquely aligned, around half gave significant BLAST hits from Enterobacteria phage ϕ X174 which is run in a separate lane on the sequencing chip as a positive control as part of the base-calling step. Smaller proportions of unalignable reads produced hits against *Halomonas*, *Vibrio* and other *Streptomyces* species but were presumed to be erroneous as DNA from these species had not been used in the lab where the library construction and sequencing had taken place.

7.9 Analysis of DNase-seq cutting pattern

Data analysis was carried out with the help of Govind Chandra. Quality control and alignment of reads to the genome were carried out according to standard methods. Each aligned read was trimmed back to 1 bp, representing the site of one DNase I cut with an expected error of $\pm 1 - 2$ bp, and the distributions of the cuts were stored as BED files.

Samples were first analysed individually without normalisation to a common reference. For each sample, the cutting index (CI) for each gene in the whole genome) was calculated by dividing the number of cuts falling within its coding region by the length of its coding region, giving a value with the units in cuts per bp. A whole-genome view of CI for each gene confirmed the absence of strong gene dosage effects around the origin of replication (Figure 7.10), however a distinct pattern of digestibility was not visible at this level for either the mock-digested or DNase I-digested samples.. The range of CI values was different in each sample due to variations in the number of alignable reads, particularly for sample MnX8 which had generated far more alignable reads and therefore had particularly high CI values.

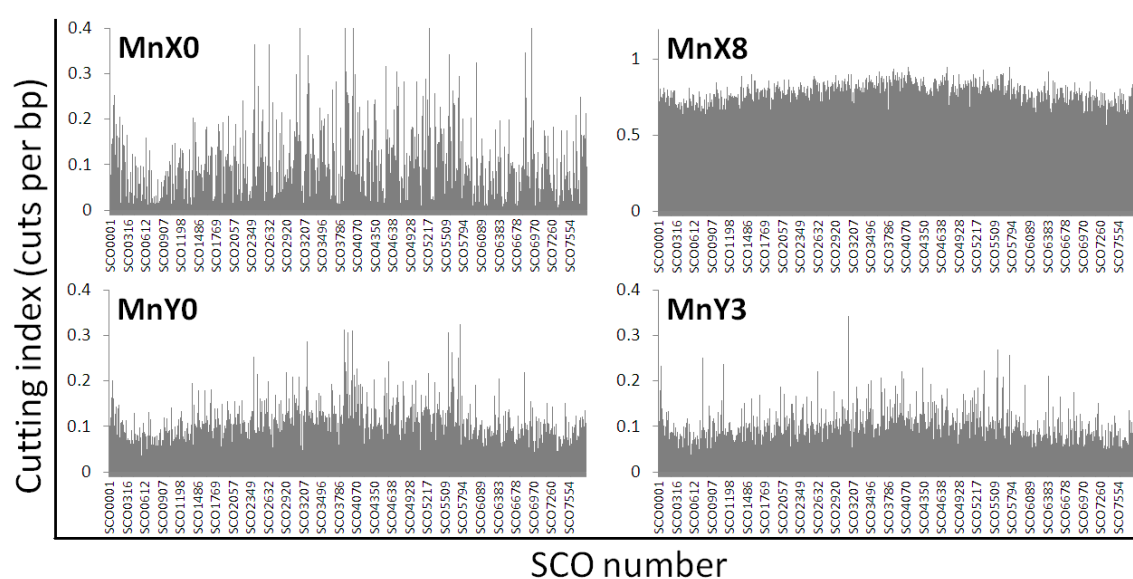


Figure 7.10 The cutting index (CI) of each gene was calculated by dividing the number of aligned reads starting within the coding region of that gene by the length of the gene (bp). The CIs for genes lying within the middle portion of the genome were not significantly higher than those lying at the ends of the genome, showing that strong copy number effects resulting from replication were not present.

CI values for twelve genes (*actII-ORF4*, *cdaR*, *redD*, *afsK*, *rlpA*, *rbnH*, *afsQ2*, *rpmG3*, *pgk*, *hrdB*, *sti1*, *fba*) were plotted against their stationary phase relative expression levels reported in McArthur and Bibb (2006). It was expected that there would be a significant positive correlation for digested samples and no correlation for undigested samples. However no significant correlation was seen in any sample (Figure 7.11), indicating that the differential digestion signal observed by qPCR had been lost during either library construction or sequencing.

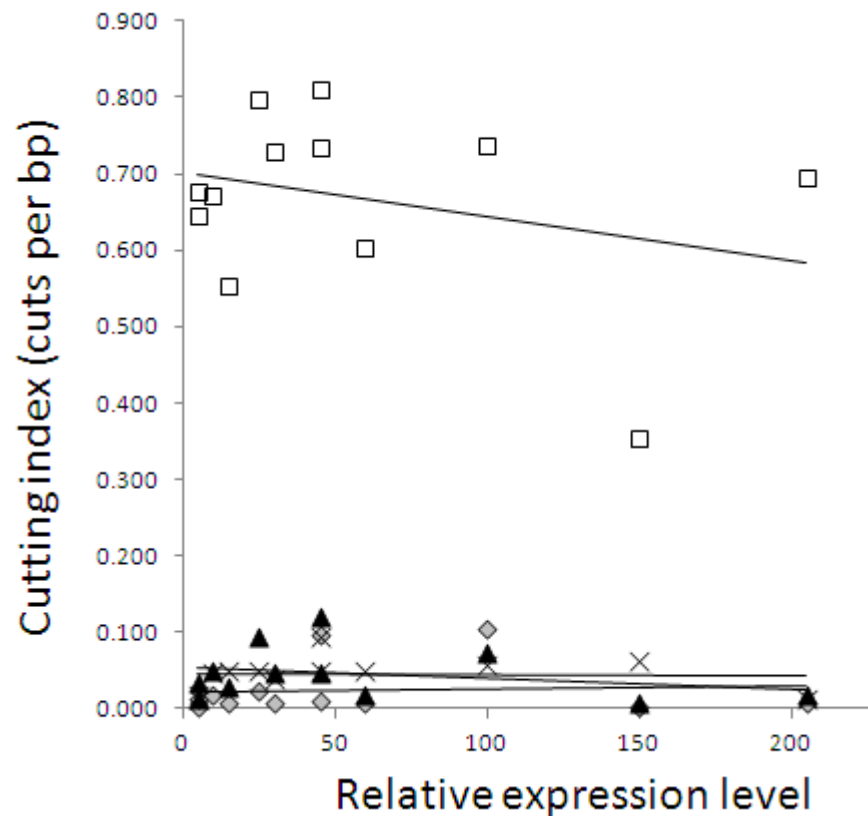
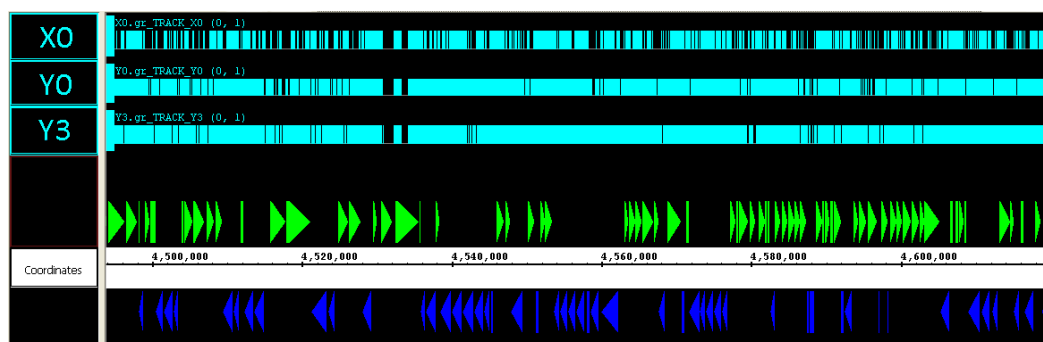


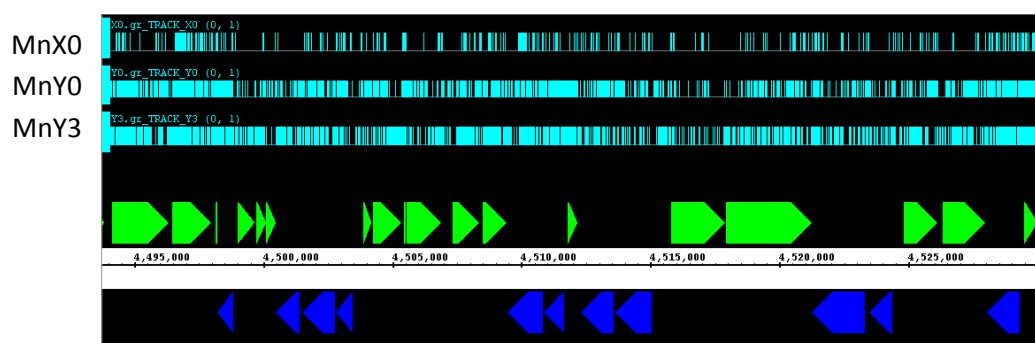
Figure 7.11 CI was not positively correlated with expression genes for the panel of twelve genes previously found to be responsive to DNase I digestion. CI values were generated in this work while relative expression levels are from McArthur and Bibb (2006), with values normalised against HrdB. MnX0: grey diamonds; MnX8: white squares; MnY0: black triangles; MnY3: crosses.

In order to investigate this discrepancy further, the cutting patterns were inspected visually by viewing the BED files as tracks in Integrated Genome Browser (IGB). The cut distributions were far from uniform across the genome in the undigested samples (MnX0 and MnY0) and they were

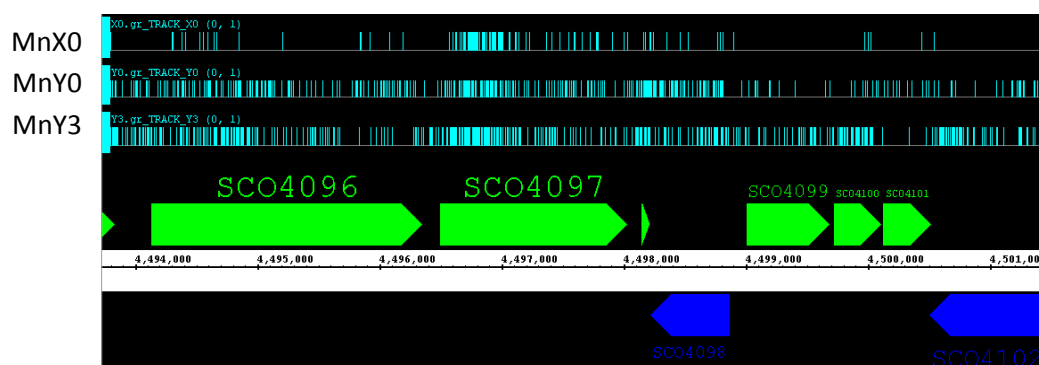
strikingly similar to the patterns seen for the digested samples. Any prominently dense or sparse regions were similar for all tracks (Figure 7.12, sample MnX8 omitted for clarity).



Approximately 100 kb in view



Approximately 30 kb in view



Approximately 5 kb in view

Figure 7.12 Integrated Genome Browser view of aligned reads showing a similar pattern of distribution between two mock-digested biological replicates (MnX0 and MnY0) and one DNase I-digested replicate (MnY3). Top track = MnX0, middle track = MnY0, bottom track = MnY3. Each vertical blue line represents the start point of one uniquely aligned read. Views displaying approximately 100 kb (top), 30 kb (middle) and 5 kb (bottom). Green and blue arrows represent the genes on each DNA strand.

The 12 genes shown in Figure 7.11 were each inspected visually, revealing that the “background” level of patterning was very strong in some places: *sti1* was covered more-or-less evenly by cuts sites (Figure 7.13, above right) while *afsK* had a region around the centre of the gene which was strikingly free of cuts (Figure 7.13, above left). This type of pattern was seen at many points around the genome and did not correspond with GC content or gene structure, hinting at a possible systematic bias in efficiency of alignment. If there was a problem in alignment it may be expected to be seen particularly strongly in the rRNA operons, which share a particular closeness in sequence. Of the six rRNA operons, three (*rrnF*, *rrnA*, *rrnE*) were found to have particularly severe read-free zones (Figure 7.13, below) within both the 16S and 23S rRNA genes although the other three operons (*rrnB*, *rrnC*, *rrnD*) appeared more normal.

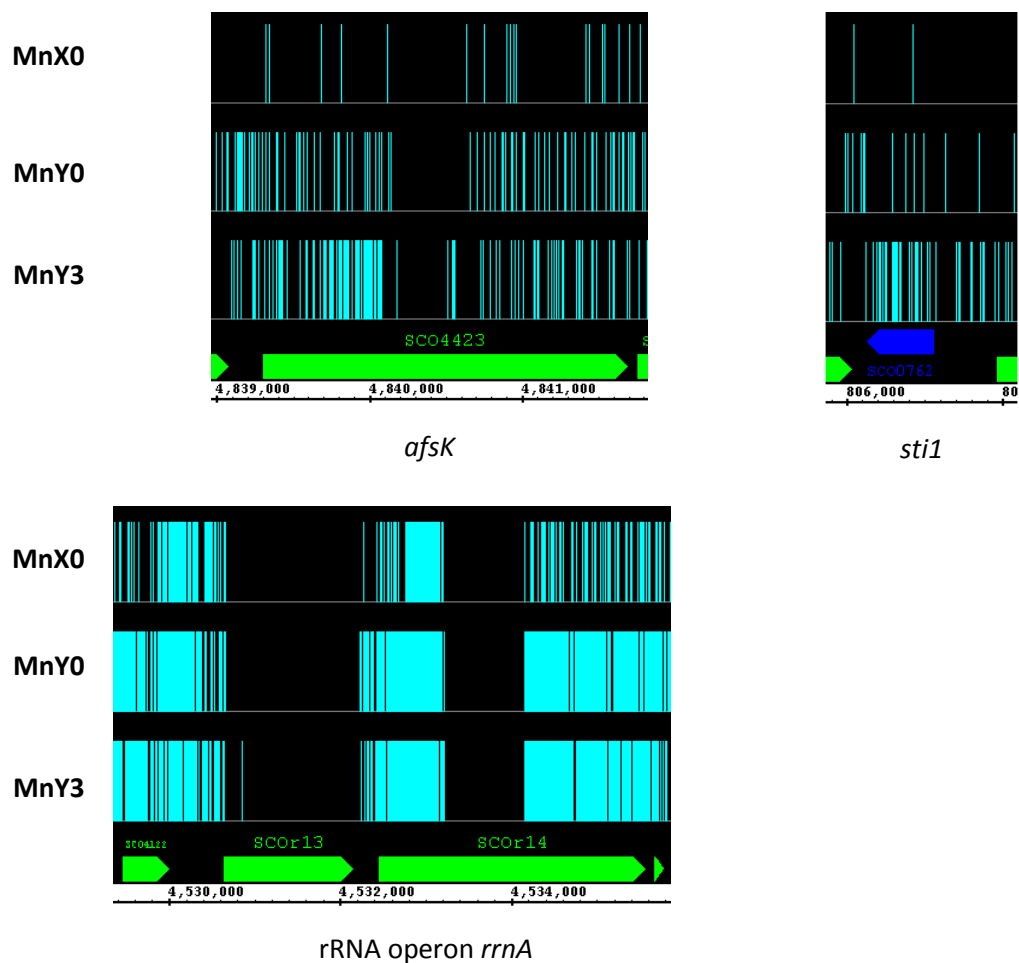


Figure 7.13 Cut sites across the genes *afsK*, *sti1* and the rRNA operon *rrnA* were distributed very unevenly even in mock-digested samples MnX0 and MnY0, with striking read-free regions in the coding regions of genes such as *afsK* and several rRNA operons. Read alignment pattern was visualised using the Integrated Genome Browser. Top track = MnX0, middle track = MnY0, bottom track = MnY3. Each vertical blue line represents the start point of one uniquely aligned read.

7.10 *HaeIII*-seq assay design

As the source of the alignment problem in the DNase-seq experiment has not yet been positively confirmed or remedied the converse experimental design was tried. Instead of creating non-specific breaks then reducing average size with a sequence-specific enzyme, an experiment was designed whereby fragments produced by an *in vivo* restriction digest using a restriction enzyme could be ligated to sequencing adaptors prior to size reduction by sonication. In this case the reads beginning with the half restriction site would be enriched over those ends generated by shearing events. Very few enzymes are suitable for this procedure, especially in such as GC-rich genome, as there is a trade-off between specificity of sequence and frequency of cutting. Most 6-base cutters have extremely few restriction sites. Even commonly-used frequent-cutters such as *EagI* and *Sau3AI* cut relatively infrequently, to the extent that many genes contain no cut sites at all and would therefore not be present in the analysis. *HaeIII* is a very frequent 4-base cutter which recognises the site CC|GG and has more than twice as many cut sites in the *S. coelicolor* genome as either *Sau3AI* or *EagI*, with nearly all genes containing at least one cut site. Complete *in silico* digestion of cosmid Ste59 gave a large number of small fragments (Figure 7.14), with almost no fragments from a complete digestion being too large for incorporation into the sequencing library (i.e. >600 bp), therefore this enzyme was selected.

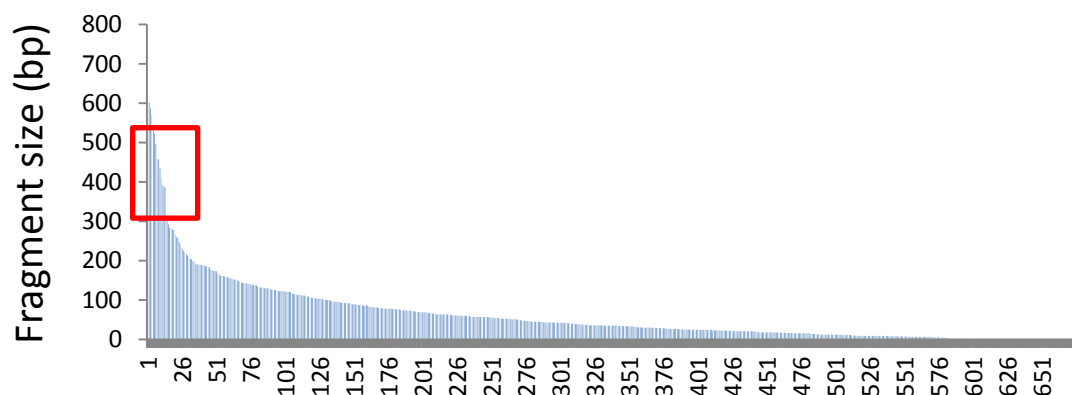


Figure 7.14 Theoretical set of fragment sizes produced by digesting the *S. coelicolor* genomic DNA-derived insert of cosmid Ste59 to completion with *HaeIII*. Fragments are ranked by size. The red box marks the fragments of a size which would fall into the window for library construction.

This method was expected to have several drawbacks compared to DNase-seq. The strength of signal in the sequencing data was likely to be lower than that generated by DNase I as there are fewer possible cutting sites. This was expected to be particularly true of the qPCR data because each amplicon spanned just two *HaeIII* sites. Another possible problem was the greater size of *HaeIII* (approximately 65 kDa) relative to DNase I (approximately 30 kDa) which could restrict access of the restriction enzyme to smaller nucleoid structures, decreasing the resolution of the approach.

7.11 Preparation of *Hae*III-digested mycelium

The same basic protocol was used as for DNase I fragmentation, with three minor modifications. Firstly the digestion time was raised from three to five minutes. Secondly the digestion was carried out in 1x NEB buffer 4 instead of DNase digestion buffer. Thirdly the samples were heated to 80°C for 20 min immediately following digestion then allowed to cool before addition of STOP buffer, in order to heat inactivate the *Hae*III but not the proteinase K. The amount of *Hae*III required was determined empirically: a single unit of *Hae*III was found to digest 1 µg naked DNA to completion within 3 minutes at 37°C (not shown), however significantly more was required to digest nucleoids to an acceptable level *in vivo*. Sample 120 U was found to be strongly digested while samples of intermediate levels of digestion (10 U and 50 U) had a smear of fragments similar to that produced by DNase I (Figure 7.15).

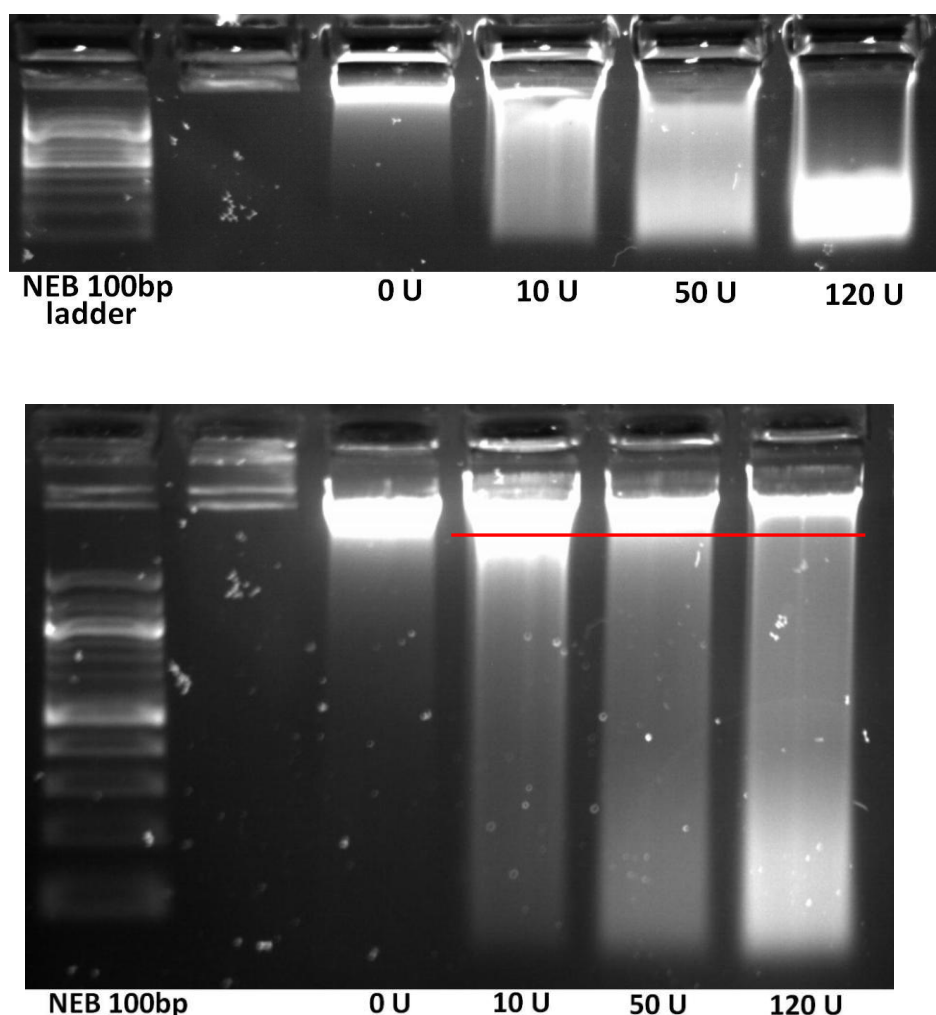


Figure 7.15 DNA fragments produced by *Hae*III digestion of permeabilised *S. coelicolor* mycelium were run on a 1% agarose gel at 80 V for 15 min (above) and 30 min (below). 0U is a mock-digested sample while 10U, 50U and 120U were digested with increasing concentrations of *Hae*III. The cut-off size used for gel extraction is marked with a red line.

As before, *HaeIII*-digested samples were resolved on a 1.5% agarose gel then extracted using the Qiagen Gel Extraction kit to remove undigested HMW DNA (Figure 7.15, right, black arrow), while the undigested sample (0 U) was used without fractionation. Digestion level was again measured by qPCR, this time against four genes: *afsK*, *sti1*, *actII-4* and *hrdB*. No significant differences in copy number were seen even at the highest level of digestion showing that, while genome fragmentation had clearly occurred, differential digestion could not be detected using this system (Figure 7.16). It is not known whether this was due to lack of sensitivity in the qPCR assay or to the inherent digestion properties of this enzyme. As differential digestion could not be demonstrated, these samples were not submitted for sequencing.

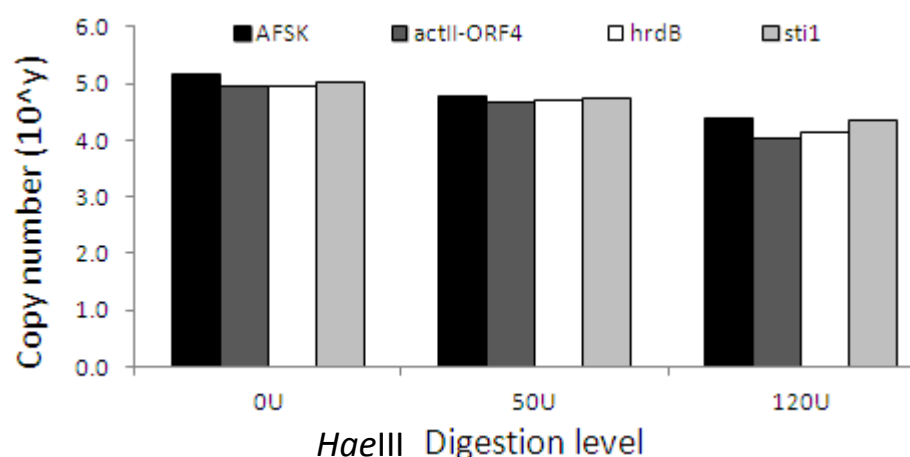


Figure 7.16 *HaeIII* digestion did not generate differential digestion between genes of different expression level. qPCR was used to measure the copy numbers of four genes in mock-digested and *HaeIII*-digested samples: the high expression gene *sti1* (light grey), the low expression gene *afsK* (black), the vegetative sigma factor *hrdB* (white) and the actinorhodin pathway transcriptional regulator *actII-ORF 4* (dark grey).

7.12 Discussion

Frustratingly, the DNase-seq assay did not produce the data expected and the source of the problem was not conclusively identified. The digestion itself presumably worked since this is a previously-successful and straightforward procedure which produced the expected degree of genome fragmentation (as visualised by gel electrophoresis) and produced approximately 2-fold differential digestion (as measured by qPCR). However no evidence of this differential digestion was seen in the sequencing data. The problem is likely to have arisen during sequencing, as the quality of the sequencing data was disappointingly low in several respects: the Q30 limit for all reads was around half that expected and a very high proportion of non-*Sma*I reads were not uniquely alignable, often giving BLAST hits to other organisms instead. Therefore the effective sequencing depth was much lower than anticipated.

The mock-digested samples (MnX0 and MnY0) were predicted to predominantly contain cuts created by *Sma*I, with only a small proportion generated by random shear, which were expected to be distributed more-or-less evenly across the genome. Mock-treated DNA was almost entirely HMW before *Sma*I digestion so it seems unlikely that degradation during handling or endogenous endonuclease activity had introduced an extra source of breaks. Unexpectedly, the cut distributions in the mock-digested samples were extremely uneven within and between genes. This, combined with the existence of conspicuous “read-free” zones in the same positions in all samples, suggested there had been a problem with alignment.

Some regions of the genome are inherently more difficult to align against (i.e. they require a read with more information to make a unique alignment), for example repetitive elements or genes with paralogues of high sequence similarity. The poor quality of the sequence data meant that reads contained less information and therefore many could not be uniquely aligned, with the effect being stronger at more problematic regions. Therefore the patterns in “ease of alignability” inherent to the genome swamped any signal generated by DNase I. This was particularly striking in some of the rRNA operons, which share high homology with one another.

Alternatively there may have been a problem with incorporation of DNase I-created ends into the library. The digested samples were expected to contain DNA breaks generated by *Sma*I and shear forces, but also have a significant proportion of cuts generated by a third source (DNase I) which would be distributed non-randomly. However this was not seen: sample MnY3 actually had a lower proportion of non-*Sma*I reads than the undigested controls. *S. coelicolor* has a very GC-rich genome, so there would have been a number of DNase-generated reads beginning “GGG” which were wrongly assigned as *Sma*I-generated and discarded, but this would only account for a tiny

proportion of the signal. It is possible that for some reason DNase I cuts were excluded from the library, for example if the end-repair step had failed to fill in the 1-2 bp overhangs.

The length and quality of reads cannot be improved in retrospect so the only solution will be to repeat the experiment and observe whether the problem recurs. If the source of the alignment problem in the DNase-seq experiment can be identified and corrected, it seems likely that this assay could still be scaled up to cover the whole genome as originally intended. In the first instance it will be necessary to make another pair of biological replicates and carry out the same experiment again, in the hope of getting better quality sequencing data. If better reads were obtained then either the signal from the differential digestion would become apparent or it would mean the signal had been lost during library construction, in which case several options for improvement are possible. Firstly, the optimal *Sma*I digestion level was guessed rather than empirically determined. It was strong enough that approximately 95% of the reads represented *Sma*I cut sites, but conversely was light enough that the average fragment size following restriction (approximately 1 kb) was still much greater than the library size window, showing that the digest was not approaching completion. Secondly, the library construction workflow could be altered, for example by ligating the sequencing adaptors onto the fragments before size reduction. If biotinylated adaptors were used then the DNase I cuts could be purified away from the cuts generated by size reduction and the effective sequencing depth would be greatly increased. However this is often an inefficient process. Thirdly, a control comprising total DNA which has been treated with the same amount of *Sma*I would be carried out, in order to quantify the differences in ease of alignment between genomic regions. This would be carried out in addition to the mock-digested control. If this assay can be fully realised as a probe of large-scale nucleoid structure, it would be of great use to the McArthur lab as it would be a relatively straightforward and unbiased way of measuring perturbations to the nucleoid caused by genetic deletion or over-expression of NAPs and chemical epigenetic treatments.

Using an alternative enzyme was not a viable alternative since *Hae*III was not found to act in the same way as DNase I. While it clearly was able to enter the cells and fragment the chromosome to a similar degree as DNase I, there was no qPCR-detectable difference in digestion level between genes. It is not known whether the enzyme was not able to perform such strong differential cutting or whether the use of qPCR amplicons spanning only two cut sites was not sensitive enough to detect it. Additionally, the quantity of enzyme required for strong fragmentation is prohibitively high for use in a routine assay.

To complement this work it would be desirable to investigate the molecular mechanism of differential cutting by DNase I in bacteria. In eukaryotes it is thought that transcriptionally silent DNA is physically protected from DNase I digestion by occlusion in dense nucleosome arrays and higher-order chromatin structures. However in bacteria the mechanism may be different, for example occlusion may be carried out by H-NS-like proteins

Summary:

- 1) *S. coelicolor* mycelium was fragmented by DNase I *in vivo*. qPCR showed that the high-expression gene *sti1* was digested at a faster rate than the low-expression gene *afsK*.
- 2) The DNase I-digested fragments were incorporated into an Illumina library and sequenced by TGAC but the quality was poor and the number of uniquely alignable reads was low.
- 3) Following alignment, no correlation was seen between digestibility and expression level for the panel of 12 genes studied in McArthur and Bibb (2006). This was probably due to variability in the “alignability” of the poor-quality sequence data against different parts of the genome. It is expected that submitting a duplicate set of samples for sequencing may result in better-quality reads which will prove more useful.
- 4) A related assay using *HaeIII* to digest the genome was developed but qPCR was unable to detect differential digestion between four genes of different expression levels, perhaps because the small number of cut sites per gene generated too little signal.

Chapter 8 – General discussion

8.1 Summary of findings

This work added to the body of knowledge about the functions and evolution of the important NAPs HU and Lsr2 in *Streptomyces*. It also generated a list of candidate NAPs, some of which are likely to represent the “missing” functions in the nucleoids of this group.

The proteomics work described in chapter 3 generated a plausible list of candidate NAPs which included many known NAPs and global regulators. The list also included a number of proteins whose great abundance and possession of a recognisable DNA-binding domain mean they have an excellent probability of being genuine NAPs or global regulators. There were also some candidates from unknown protein classes which may represent the most interesting novel NAPs and may act via currently unknown mechanisms, although some will surely turn out to be contaminants.

Surprisingly, several lines of evidence suggest that HupS plays a broader role than was originally thought. The presence of large amount of HupS protein associated with nucleoids harvested from stationary phase cultures and the strong phenotype of $\Delta hupS$ seen in liquid cultures indicate that it plays a role even in non-differentiating cultures. The complete lack of coverage of peptides within the lysine-rich tail of HupS may have been an artefact of the trypsin digest, if this process caused fragmentation of the tail into peptides too small to detect by mass spectrometry, in which case the amount of HupS present would actually have been far greater than reported (as the emPAI score is based on coverage), or alternatively it is possible that a truncated version of HupS is present instead. This would be not be inconsistent with the work of Salerno *et al.* (2009) who found that a translational fusion of EGFP to the C-terminus of HupS was not detected in vegetative cultures. This would be particularly interesting in the context of the phylogenetic analysis of HupS, which showed that strong divergence of the HU domains of HupA and HupS occurred before the acquisition of the lysine-rich tail and that many species possess only a tailed version of HupS. If one gene were producing both a short HU during vegetative phase and a longer HU during stationary phase this would represent a new means of increasing the diversity of bacterial chromatin.

As is typically the case for NAPs, the mutant phenotypes of HupA and HupS were pleiotropic. In many cases the $\Delta hupA$ and $\Delta hupS$ mutations had opposite effects on phenotype, in particular the

effects on mycelial pellet morphology, colony size, spore properties and actinorhodin production on agar media. It will be necessary to repeat the experiments of Yokoyama *et al.* (2001) in *S. coelicolor* to exclude the possibility that the $\Delta hupA$ phenotype is not simply due to a compensating upregulation of *hupS*.

The mutant phenotypes of the Lsr2 mutants were milder and showed a greater degree of overlap with each other than expected, especially given that the quantities of the two proteins detected in isolated nucleoids showed very different patterns. SCO3375 was very abundant in liquid culture nucleoids while SCO4076 was detected in low amounts during late vegetative phase and not at all at other time points.

The presence of repetitive sequences rich in the amino acids lysine, alanine and proline was much more widespread than expected. Lysine-rich tails have apparently evolved on three separate occasions on HU proteins: once on the N-terminus of *Deinococcus* HU and once each on the C-termini of HupS and *Yersinia* phage HU. There are also regions of repetitive sequence rich in lysine and alanine in the C-terminal regions of HU proteins from the Firmicutes but these have a different repeat unit length, with HupS containing a motif similar to “KKAAAKKPAKK” and Firmicutes sequences containing a motif similar to “DAVKPAVKDAVK”. There are also lysine-rich sequences of varying lengths of a HupS-like repeat on some Lsr2 proteins and at least a dozen other *S. coelicolor* DNA-binding proteins such as the sigma factor HrdB, SCO2075, SCO5750 and SCO3543 (DNA topoisomerase I). The lysine-rich tails of HupS homologues from *Mycobacterium* and from *Streptomyces* have several functions including chemical protection (Takatsuka *et al.*, 2011), increased affinity (Kumar *et al.*, 2010) and host attachment (Pethe *et al.*, 2002). It would be advantageous to know whether the repeats on other proteins have the same properties.

One anticipated complication of using techniques such as nucleoid isolation or *in vivo* DNase I fragmentation of the genome on *S. coelicolor* grown in liquid culture was that the formation of mycelial pellets would limit the surface area available for access by lysozyme or DNase I as compared with species like *E. coli* which grow as single cells. This was found to be the case, for example the yield of nucleoids recovered was hugely lower than when a similar method was applied to *E. coli*. However with careful optimisation of protocols and a liquid medium that favoured dispersed growth this problem was not insurmountable.

8.2 Relevance of this work to natural product discovery

It is clear that NAPs play an important role in the secondary metabolism of *S. coelicolor*, as the deletion and over-expression strains examined in this work all showed differences in production of actinorhodin and/or undecylprodigiosin, as has also been shown for slHF (Yang *et al.*, 2012). slHF was found to bind to the promoter of *red*, which is the pathway-specific regulator of undecylprodigiosin production (Park *et al.*, 2009), but it is not known whether the HU and Lsr2 proteins regulate antibiotic production directly or through an indirect effect on other traits. Many of the strains tested here showed alterations in growth and morphological differentiation (particularly in forming spore chains) so perhaps these were not the best NAPs to choose. Fortunately we have found candidates for new ones to try.

It may be worthwhile to investigate more sophisticated ways of genetically manipulating the nucleoid structure than simply deleting or over-expressing NAPs. For example truncated versions of the Lsr2 proteins containing only the dimerization domain could be over-expressed to weaken the ability of endogenous Lsr2 to repress large regions of genes.

It was disappointing that post-translational modifications were not detected within the proteomics datasets as these would have strengthened the basis for the chemical epigenetic approach being investigated in the McArthur lab (Moore *et al.*, 2013). However this work is ongoing and it is hoped that improvements to the method such as the addition of modification enrichment steps (e.g. anti-acetylsine antibodies) will fix this problem.

8.3 Future work

The list of candidate NAPs generated by the proteomics experiments was plausible as it included many known NAPs/global regulators. However at present they are merely candidates. To confirm that a protein is a NAP the following must be shown: that it is highly abundant; that it binds DNA with low sequence specificity; that it controls gene expression of around 5 - 10% of the genome; that it has an effect on the structure of promoters instead of (or as well as) recruiting the RNAP by protein-protein interactions. It would also be desirable to repeat these experiments on mycelium grown on a solid medium which permits morphological differentiation or under different nutrient conditions. This would expand the list of NAPs and reflect more natural growth conditions. However it is possible that mycelium grown on solid media may not be so amenable to nucleoid isolation as TSB:YEME34% liquid cultures, in which case it may be beneficial to try the

experiments on *S. venezuelae* which grows in a dispersed manner and sporulates to completion in liquid culture.

In order to fully understand the action of novel NAPs, it will be necessary to determine their structures and mode of interaction with DNA. SIHF from *S. griseus* has recently been crystallised in complex with dsDNA and some preliminary diffraction data have been collected (Nomoto *et al.*, 2012); it will be fascinating to find out whether it resembles any known DNA-binding protein or whether it represents a new protein fold. Much of the information available for HupS/Mdp1 centres on its roles in biological processes while data such as its binding sites in *Streptomyces* and the bend angle it produces are not known. It is possible that the properties of the HupS HU domain are quite distinct from those of HupA because *S. coelicolor* HupA and HupS are as divergent in sequence as *E. coli* HU α and IHF.

The molecular basis for the opposite effects of $\Delta hupA$ and $\Delta hupS$ on growth were not determined. It is conceivable that the difference is caused by ectopic production of fimbriae in the $\Delta hupA$ strain and, causing excessive attachment to hydrophobic surfaces. This could be tested by repeating the growth curves in the presence of Congo red dye, which inhibits production of fimbriae by preventing assembly of the amyloid fibrils of which they are composed. The presence of fimbriae could also be examined by electron microscopy using negative staining (de Jong *et al.*, 2009).

The molecular basis for the germination defect of morphologically-normal $\Delta hupA$ spores is similarly unknown. The fact that around 10% of the spores are able to germinate on a normal timescale to produce normal, albeit slightly small, colonies on rich SFM plates suggests that even within a population of genetically identical spores there are differences in germination behaviour. It is possible that spores have lost the ability to respond to quorum-sensing type signals produced by the first spores to germinate, in a similar manner to processes seen in germinating *Bacillus* spores (Dworkin and Shah, 2010). It is also possible that there natural variation in the quantities of storage compound such as trehalose laid down during spore formation, which could affect the ability of a spore to germinate successfully.

The basis for the spore chain defect of both *lsr2* mutants could be investigated further by microscopical analysis of events occurring during spore nucleoid segregation. Staining of DNA within forming spore chains with a fluorescent intercalating agent such as propidium iodide would shed light on the nature of the defect, for example whether the pre-spore nucleoids are failing to fully separate, failing to become condensed, or whether the proteins are required for correct

regulation of genes involved in differentiation. It may also be beneficial to visualise components of the segregation-related machinery using translational fusions of those proteins, for example to know whether FtsZ rings and SMC foci are forming properly. It would be interesting to know whether the spore chain formation defect and other overlapping features in the $\Delta 3375$ and $\Delta 4076$ phenotypes are caused by the loss of a 3375-4076 heterodimer.

In this thesis only single mutants were examined, however it is likely that double mutants of $\Delta hupA/\Delta hupS$ and $\Delta 3375/\Delta 4076$ would be highly informative. It would also be of interest to determine whether HupA/HupS or SCO3375/SCO4076 heterodimers are formed and, if so, under what conditions or stage of the life cycle. This would open up fascinating lines of research in species which possess a large number of paralogues for example *A. mediterranei*, which has five paralogues each of HU and Lsr2.

It is likely that repeating the DNase-seq assay will be fruitful, provided that good-quality sequence data is obtained on a second attempt. There will still be many challenges in designing the tools required for analysis of nucleoid structural features but this would be a very valuable assay for our lab and for the use by others.

8.4 Concluding remarks

The growing availability of knowledge about the nucleoid structure and NAPs of *S. coelicolor* make this species an excellent model for nucleoid biology within the Actinomycetes and there is potential for applying this knowledge to the process of natural product discovery and pathogenicity. In addition the importance of post-transcriptional regulation in *Streptomyces* is becoming increasingly well appreciated and it will be very exciting to learn which proteins are involved in mediating the regulatory action of sRNAs.

References

- Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A., 1999, Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid: *J Bacteriol*, v. 181, p. 6361-70.
- Allen, T.E., Price, N.D., Joyce, A.R., and Palsson, B.O., 2006, Long-range periodic patterns in microbial genomes indicate significant multi-scale chromosomal organization: *PLoS Comput Biol*, v. 2, p. e2.
- Amaro, A.M., and Jerez, C.A., 1984, Methylation of ribosomal proteins in bacteria: evidence of conserved modification of the eubacterial 50S subunit: *J Bacteriol*, v. 158, p. 84-93.
- Amit, R., Oppenheim, A.B., and Stavans, J., 2003, Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor: *Biophys J*, v. 84, p. 2467-73.
- Antson, A.A., Brzozowski, A.M., Dodson, E.J., Dauter, Z., Wilson, K.S., Kurecki, T., Otridge, J., and Gollnick, P., 1994, 11-fold symmetry of the trp RNA-binding attenuation protein (TRAP) from *Bacillus subtilis* determined by X-ray analysis: *J Mol Biol*, v. 244, p. 1-5.
- Argaman, L., Elgrably-Weiss, M., Hershko, T., Vogel, J., and Altuvia, S., 2012, RelA protein stimulates the activity of RyhB small RNA by acting on RNA-binding protein Hfq: *Proc Natl Acad Sci U S A*, v. 109, p. 4621-6.
- Arnvig, K.B., and Young, D.B., 2009, Identification of small RNAs in *Mycobacterium tuberculosis*: *Mol Microbiol*, v. 73, p. 397-408.
- Arold, S.T., Leonard, P.G., Parkinson, G.N., and Ladbury, J.E., 2010, H-NS forms a superhelical protein scaffold for DNA condensation: *Proc Natl Acad Sci U S A*, v. 107, p. 15728-32.
- Arora, K., Whiteford, D.C., Lau-Bonilla, D., Davitt, C.M., and Dahl, J.L., 2008, Inactivation of *Isr2* results in a hypermotile phenotype in *Mycobacterium smegmatis*: *J Bacteriol*, v. 190, p. 4291-300.
- Atlung, T., and Hansen, F.G., 2002, Effect of different concentrations of H-NS protein on chromosome replication and the cell cycle in *Escherichia coli*: *J Bacteriol*, v. 184, p. 1843-50.
- Atlung, T., and Ingmer, H., 1997, H-NS: a modulator of environmentally regulated gene expression: *Mol Microbiol*, v. 24, p. 7-17.
- Azam, T.A., Hiraga, S., and Ishihama, A., 2000, Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid: *Genes Cells*, v. 5, p. 613-26.
- Azam, T.A., and Ishihama, A., 1999, Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity: *J Biol Chem*, v. 274, p. 33105-13.
- Azem, A., Tsfadia, Y., Hajouj, O., Shaked, I., and Daniel, E., 2010, Cross-linking with bifunctional reagents and its application to the study of the molecular symmetry and the arrangement of subunits in hexameric protein oligomers: *Biochim Biophys Acta*, v. 1804, p. 768-80.
- Bagwell, C.E., Bhat, S., Hawkins, G.M., Smith, B.W., Biswas, T., Hoover, T.R., Saunders, E., Han, C.S., Tsodikov, O.V., and Shimkets, L.J., 2008, Survival in nuclear waste, extreme resistance, and potential applications gleaned from the genome sequence of *Kineococcus radiotolerans* SRS30216: *PLoS One*, v. 3, p. e3878.
- Bai, Q., and Somerville, R.L., 1998, Integration host factor and cyclic AMP receptor protein are required for TyrR-mediated activation of *tpl* in *Citrobacter freundii*: *J Bacteriol*, v. 180, p. 6173-86.
- Balandina, A., Claret, L., Hengge-Aronis, R., and Rouviere-Yaniv, J., 2001, The *Escherichia coli* histone-like protein HU regulates *rpoS* translation: *Mol Microbiol*, v. 39, p. 1069-79.
- Balandina, A., Kamashev, D., and Rouviere-Yaniv, J., 2002, The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids: *J Biol Chem*, v. 277, p. 27622-8.

- Balasubramanyam, K., Swaminathan, V., Ranganathan, A., and Kundu, T.K., 2003, Small molecule modulators of histone acetyltransferase p300: *J Biol Chem*, v. 278, p. 19134-40.
- Banos, R.C., Vivero, A., Aznar, S., Garcia, J., Pons, M., Madrid, C., and Juarez, A., 2009, Differential regulation of horizontally acquired and core genome genes by the bacterial modulator H-NS: *PLoS Genet*, v. 5, p. e1000513.
- Barak, R., and Eisenbach, M., 2001, Acetylation of the response regulator, CheY, is involved in bacterial chemotaxis: *Mol Microbiol*, v. 40, p. 731-43.
- Barsnes, H., Vizcaino, J.A., Eidhammer, I., and Martens, L., 2009, PRIDE Converter: making proteomics data-sharing easy: *Nat Biotechnol*, v. 27, p. 598-9.
- Basu, D., Khare, G., Singh, S., Tyagi, A., Khosla, S., and Mande, S.C., 2009, A novel nucleoid-associated protein of *Mycobacterium tuberculosis* is a sequence homolog of GroEL: *Nucleic Acids Res*, v. 37, p. 4944-54.
- Bates, A.D., and Maxwell, A., 2005, DNA topology: Oxford, Oxford University Press.
- Battesti, A., Majdalani, N., and Gottesman, S., 2011, The RpoS-mediated general stress response in *Escherichia coli*: *Annu Rev Microbiol*, v. 65, p. 189-213.
- Baumann, H., Knapp, S., Lundback, T., Ladenstein, R., and Hard, T., 1994, Solution structure and DNA-binding properties of a thermostable protein from the archaeon *Sulfolobus solfataricus*: *Nat Struct Biol*, v. 1, p. 808-19.
- Beloin, C., Jeusset, J., Revet, B., Mirambeau, G., Le Hégarat, F., and Le Cam, E., 2003, Contribution of DNA conformation and topology in right-handed DNA wrapping by the *Bacillus subtilis* LrpC protein: *J Biol Chem*, v. 278, p. 5333-42.
- Benham, C.J., 1979, Torsional stress and local denaturation in supercoiled DNA: *Proc Natl Acad Sci U S A*, v. 76, p. 3870-4.
- Bensaid, A., Almeida, A., Drlica, K., and Rouviere-Yaniv, J., 1996, Cross-talk between topoisomerase I and HU in *Escherichia coli*: *J Mol Biol*, v. 256, p. 292-300.
- Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, 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., Rabinowitsch, 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*, v. 417, p. 141-7.
- Berger, M., Farcas, A., Geertz, M., Zhelyazkova, P., Brix, K., Travers, A., and Muskhelishvili, G., 2010, Coordination of genomic structure and transcription by the main bacterial nucleoid-associated protein HU: *EMBO Rep*, v. 11, p. 59-64.
- Bermudez, I., Garcia-Martinez, J., Perez-Ortin, J.E., and Roca, J., 2010, A method for genome-wide analysis of DNA helical tension by means of psoralen-DNA photobinding: *Nucleic Acids Res*, v. 38, p. e182.
- Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.A., and Duncan, K., 2002, Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling: *Mol Microbiol*, v. 43, p. 717-31.
- Bianchi, M.E., 1994, Prokaryotic HU and eukaryotic HMG1: a kinked relationship: *Mol Microbiol*, v. 14, p. 1-5.
- Binenbaum, Z., Parola, A.H., Zaritsky, A., and Fishov, I., 1999, Transcription- and translation-dependent changes in membrane dynamics in bacteria: testing the transertion model for domain formation: *Mol Microbiol*, v. 32, p. 1173-82.
- Blasco, B., Chen, J.M., Hartkoorn, R., Sala, C., Uplekar, S., Rougemont, J., Pojer, F., and Cole, S.T., 2012, Virulence regulator EspR of *Mycobacterium tuberculosis* is a nucleoid-associated protein: *PLoS Pathog*, v. 8, p. e1002621.
- Blumenschein, G.R., Jr., Kies, M.S., Papadimitrakopoulou, V.A., Lu, C., Kumar, A.J., Ricker, J.L., Chiao, J.H., Chen, C., and Frankel, S.R., 2008, Phase II trial of the histone deacetylase

- inhibitor vorinostat (Zolinza, suberoylanilide hydroxamic acid, SAHA) in patients with recurrent and/or metastatic head and neck cancer: *Invest New Drugs*, v. 26, p. 81-7.
- Boggild, A., Overgaard, M., Valentin-Hansen, P., and Brodersen, D.E., 2009, Cyanobacteria contain a structural homologue of the Hfq protein with altered RNA-binding properties: *FEBS J*, v. 276, p. 3904-15.
- Bohn, C., Rigoulay, C., and Bouloc, P., 2007, No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*: *BMC Microbiol*, v. 7, p. 10.
- Borca, M.V., Irusta, P.M., Kutish, G.F., Carillo, C., Afonso, C.L., Burrage, A.T., Neilan, J.G., and Rock, D.L., 1996, A structural DNA binding protein of African swine fever virus with similarity to bacterial histone-like proteins: *Arch Virol*, v. 141, p. 301-13.
- Boubrik, F., and Rouviere-Yaniv, J., 1995, Increased sensitivity to gamma irradiation in bacteria lacking protein HU: *Proc Natl Acad Sci U S A*, v. 92, p. 3958-62.
- Bouffartigues, E., Buckle, M., Badaut, C., Travers, A., and Rimsky, S., 2007, H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing: *Nat Struct Mol Biol*, v. 14, p. 441-8.
- Bradley, M.D., Beach, M.B., de Koning, A.P., Pratt, T.S., and Osuna, R., 2007, Effects of Fis on *Escherichia coli* gene expression during different growth stages: *Microbiology*, v. 153, p. 2922-40.
- Brekasis, D., and Paget, M.S., 2003, A novel sensor of NADH/NAD⁺ redox poise in *Streptomyces coelicolor* A3(2): *EMBO J*, v. 22, p. 4856-65.
- Brunetti, R., Prosseda, G., Beghetto, E., Colonna, B., and Micheli, G., 2001, The looped domain organization of the nucleoid in histone-like protein defective *Escherichia coli* strains: *Biochimie*, v. 83, p. 873-82.
- Bykowski, T., and Sirko, A., 1998, Selected phenotypes of *ihf* mutants of *Escherichia coli*: *Biochimie*, v. 80, p. 987-1001.
- Bystrykh, L.V., Fernandez-Moreno, M.A., Herrema, J.K., Malpartida, F., Hopwood, D.A., and Dijkhuizen, L., 1996, Production of actinorhodin-related "blue pigments" by *Streptomyces coelicolor* A3(2): *J Bacteriol*, v. 178, p. 2238-44.
- Cameron, A.D., and Redfield, R.J., 2008, CRP binding and transcription activation at CRP-S sites: *J Mol Biol*, v. 383, p. 313-23.
- Cameron, A.D., Stoebel, D.M., and Dorman, C.J., 2011, DNA supercoiling is differentially regulated by environmental factors and FIS in *Escherichia coli* and *Salmonella enterica*: *Mol Microbiol*, v. 80, p. 85-101.
- Cameron, D.M., Gregory, S.T., Thompson, J., Suh, M.J., Limbach, P.A., and Dahlberg, A.E., 2004, *Thermus thermophilus* L11 methyltransferase, PrmA, is dispensable for growth and preferentially modifies free ribosomal protein L11 prior to ribosome assembly: *J Bacteriol*, v. 186, p. 5819-25.
- Ceci, P., Cellai, S., Falvo, E., Rivetti, C., Rossi, G.L., and Chiancone, E., 2004, DNA condensation and self-aggregation of *Escherichia coli* Dps are coupled phenomena related to the properties of the N-terminus: *Nucleic Acids Res*, v. 32, p. 5935-44.
- Ceci, P., Ilari, A., Falvo, E., and Chiancone, E., 2003, The Dps protein of *Agrobacterium tumefaciens* does not bind to DNA but protects it toward oxidative cleavage: x-ray crystal structure, iron binding, and hydroxyl-radical scavenging properties: *J Biol Chem*, v. 278, p. 20319-26.
- Chakraborty, R., and Bibb, M., 1997, The ppGpp synthetase gene (*relA*) of *Streptomyces coelicolor* A3(2) plays a conditional role in antibiotic production and morphological differentiation: *J Bacteriol*, v. 179, p. 5854-61.
- Champoux, J.J., 2001, DNA topoisomerases: structure, function, and mechanism: *Annu Rev Biochem*, v. 70, p. 369-413.
- Chang, C.N., and Chang, N., 1975, Methylation of the ribosomal proteins in *Escherichia coli*. Nature and stoichiometry of the methylated amino acids in 50S ribosomal proteins: *Biochemistry*, v. 14, p. 468-77.
- Chang, F.N., Chang, C.N., and Paik, W.K., 1974, Methylation of ribosomal proteins in *Escherichia coli*: *J Bacteriol*, v. 120, p. 651-6.

- Chao, C.C., Wu, S.L., and Ching, W.M., 2004, Using LC-MS with de novo software to fully characterize the multiple methylations of lysine residues in a recombinant fragment of an outer membrane protein from a virulent strain of *Rickettsia prowazekii*: *Biochim Biophys Acta*, v. 1702, p. 145-52.
- Chen, C.S., Smits, C., Dodson, G.G., Shevtsov, M.B., Merlino, N., Gollnick, P., and Antson, A.A., 2011, How to change the oligomeric state of a circular protein assembly: switch from 11-subunit to 12-subunit TRAP suggests a general mechanism: *PLoS One*, v. 6, p. e25296.
- Chen, J.M., German, G.J., Alexander, D.C., Ren, H., Tan, T., and Liu, J., 2006, Roles of Lsr2 in colony morphology and biofilm formation of *Mycobacterium smegmatis*: *J Bacteriol*, v. 188, p. 633-41.
- Chen, J.M., Ren, H., Shaw, J.E., Wang, Y.J., Li, M., Leung, A.S., Tran, V., Berbenetz, N.M., Kocincova, D., Yip, C.M., Reyrat, J.M., and Liu, J., 2008, Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein: *Nucleic Acids Res*, v. 36, p. 2123-35.
- Chen, S., and Calvo, J.M., 2002, Leucine-induced dissociation of *Escherichia coli* Lrp hexadecamers to octamers: *J Mol Biol*, v. 318, p. 1031-42.
- Chen, T., Damaj, B.B., Herrera, C., Lasko, P., and Richard, S., 1997, Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain: *Mol Cell Biol*, v. 17, p. 5707-18.
- Chiang, S.K., Lou, Y.C., and Chen, C., 2008, NMR solution structure of KP-TerB, a tellurite-resistance protein from *Klebsiella pneumoniae*: *Protein Sci*, v. 17, p. 785-9.
- Cho, B.K., Knight, E.M., Barrett, C.L., and Palsson, B.O., 2008, Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts: *Genome Res*, v. 18, p. 900-10.
- Christiansen, J.K., Larsen, M.H., Ingmer, H., Sogaard-Andersen, L., and Kallipolitis, B.H., 2004, The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence: *J Bacteriol*, v. 186, p. 3355-62.
- Claessen, D., de Jong, W., Dijkhuizen, L., and Wosten, H.A., 2006, Regulation of *Streptomyces* development: reach for the sky!: *Trends Microbiol*, v. 14, p. 313-9.
- Claessen, D., Rink, R., de Jong, W., Siebring, J., de Vreugd, P., Boersma, F.G., Dijkhuizen, L., and Wosten, H.A., 2003, A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in *Streptomyces coelicolor* by forming amyloid-like fibrils: *Genes Dev*, v. 17, p. 1714-26.
- Clarkson, S., and Bates, A.D., 1996, Action of DNA gyrase at RIP elements in *E. coli*: *Biochem Soc Trans*, v. 24, p. 420S.
- Colangeli, R., Haq, A., Arcus, V.L., Summers, E., Magliozzo, R.S., McBride, A., Mitra, A.K., Radjainia, M., Khajo, A., Jacobs, W.R., Jr., Salgame, P., and Alland, D., 2009, The multifunctional histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates: *Proc Natl Acad Sci U S A*, v. 106, p. 4414-8.
- Colangeli, R., Helb, D., Vilcheze, C., Hazbon, M.H., Lee, C.G., Safi, H., Sayers, B., Sardone, I., Jones, M.B., Fleischmann, R.D., Peterson, S.N., Jacobs, W.R., Jr., and Alland, D., 2007, Transcriptional regulation of multi-drug tolerance and antibiotic-induced responses by the histone-like protein Lsr2 in *M. tuberculosis*: *PLoS Pathog*, v. 3, p. e87.
- Combet, C., Blanchet, C., Geourjon, C., and Deleage, G., 2000, NPS@: network protein sequence analysis: *Trends Biochem Sci*, v. 25, p. 147-50.
- Couturier, E., and Rocha, E.P., 2006, Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes: *Mol Microbiol*, v. 59, p. 1506-18.
- Cox, J., and Mann, M., 2008, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification: *Nat Biotechnol*, v. 26, p. 1367-72.
- Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D., Zhou, D., Luo, S., Vasicek, T.J., Daly, M.J., Wolfsberg, T.G., and

- Collins, F.S., 2006, Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS): *Genome Res*, v. 16, p. 123-31.
- Cristofari, G., and Darlix, J.L., 2002, The ubiquitous nature of RNA chaperone proteins: *Prog Nucleic Acid Res Mol Biol*, v. 72, p. 223-68.
- Cunha, S., Odijk, T., Suleymanoglu, E., and Woldringh, C.L., 2001, Isolation of the *Escherichia coli* nucleoid: *Biochimie*, v. 83, p. 149-54.
- Dame, R.T., and Dorman, C.J., 2010, *Bacterial chromatin*: Dordrecht ; London, Springer.
- Dame, R.T., Kalmykova, O.J., and Grainger, D.C., 2011, Chromosomal macrodomains and associated proteins: implications for DNA organization and replication in gram negative bacteria: *PLoS Genet*, v. 7, p. e1002123.
- Dame, R.T., Luijsterburg, M.S., Krin, E., Bertin, P.N., Wagner, R., and Wuite, G.J., 2005, DNA bridging: a property shared among H-NS-like proteins: *J Bacteriol*, v. 187, p. 1845-8.
- Dame, R.T., Noom, M.C., and Wuite, G.J., 2006, Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation: *Nature*, v. 444, p. 387-90.
- Datsenko, K.A., and Wanner, B.L., 2000, One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products: *Proc Natl Acad Sci U S A*, v. 97, p. 6640-5.
- Davie, J.R., 2003, Inhibition of histone deacetylase activity by butyrate: *J Nutr*, v. 133, p. 2485S-2493S.
- de Jong, W., Wosten, H.A., Dijkhuizen, L., and Claessen, D., 2009, Attachment of *Streptomyces coelicolor* is mediated by amyloid fimbriae that are anchored to the cell surface via cellulose: *Mol Microbiol*, v. 73, p. 1128-40.
- de la Cruz, X., Lois, S., Sanchez-Molina, S., and Martinez-Balbas, M.A., 2005, Do protein motifs read the histone code?: *Bioessays*, v. 27, p. 164-75.
- de Vries, R., 2010, DNA condensation in bacteria: Interplay between macromolecular crowding and nucleoid proteins: *Biochimie*, v. 92, p. 1715-21.
- Deng, S., Stein, R.A., and Higgins, N.P., 2004, Transcription-induced barriers to supercoil diffusion in the *Salmonella typhimurium* chromosome: *Proc Natl Acad Sci U S A*, v. 101, p. 3398-403.
- Derouaux, A., Dehareng, D., Lecocq, E., Halici, S., Nothaft, H., Giannotta, F., Moutzourelis, G., Dusart, J., Devreese, B., Titgemeyer, F., Van Beeumen, J., and Rigali, S., 2004, Crp of *Streptomyces coelicolor* is the third transcription factor of the large CRP-FNR superfamily able to bind cAMP: *Biochem Biophys Res Commun*, v. 325, p. 983-90.
- Dersch, P., Kneip, S., and Bremer, E., 1994, The nucleoid-associated DNA-binding protein H-NS is required for the efficient adaptation of *Escherichia coli* K-12 to a cold environment: *Mol Gen Genet*, v. 245, p. 255-9.
- DiChiara, J.M., Contreras-Martinez, L.M., Livny, J., Smith, D., McDonough, K.A., and Belfort, M., 2010, Multiple small RNAs identified in *Mycobacterium bovis* BCG are also expressed in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*: *Nucleic Acids Res*, v. 38, p. 4067-78.
- Dillon, S.C., and Dorman, C.J., 2010, Bacterial nucleoid-associated proteins, nucleoid structure and gene expression: *Nat Rev Microbiol*, v. 8, p. 185-95.
- Dorman, C.J., 2004, H-NS: a universal regulator for a dynamic genome: *Nat Rev Microbiol*, v. 2, p. 391-400.
- , 2010, Horizontally acquired homologues of the nucleoid-associated protein H-NS: implications for gene regulation: *Mol Microbiol*, v. 75, p. 264-7.
- Dorman, C.J., and Deighan, P., 2003, Regulation of gene expression by histone-like proteins in bacteria: *Curr Opin Genet Dev*, v. 13, p. 179-84.
- Dorman, C.J., Hinton, J.C., and Free, A., 1999, Domain organization and oligomerization among H-NS-like nucleoid-associated proteins in bacteria: *Trends Microbiol*, v. 7, p. 124-8.
- Doyle, M., Fookes, M., Ivens, A., Mangan, M.W., Wain, J., and Dorman, C.J., 2007, An H-NS-like stealth protein aids horizontal DNA transmission in bacteria: *Science*, v. 315, p. 251-2.
- Drew, H.R., Weeks, J.R., and Travers, A.A., 1985, Negative supercoiling induces spontaneous unwinding of a bacterial promoter: *EMBO J*, v. 4, p. 1025-32.

- Drlica, K., Chen, C.R., and Kayman, S., 1999, Sedimentation analysis of bacterial nucleoid structure: *Methods Mol Biol*, v. 94, p. 87-98.
- Dworkin, J., and Shah, I.M., 2010, Exit from dormancy in microbial organisms: *Nat Rev Microbiol*, v. 8, p. 890-6.
- Eaton, D., and Ensign, J.C., 1980, *Streptomyces viridochromogenes* spore germination initiated by calcium ions: *J Bacteriol*, v. 143, p. 377-82.
- Elliot, M.A., Bibb, M.J., Buttner, M.J., and Leskiw, B.K., 2001, BldD is a direct regulator of key developmental genes in *Streptomyces coelicolor* A3(2): *Mol Microbiol*, v. 40, p. 257-69.
- Ernsting, B.R., Atkinson, M.R., Ninfa, A.J., and Matthews, R.G., 1992, Characterization of the regulon controlled by the leucine-responsive regulatory protein in *Escherichia coli*: *J Bacteriol*, v. 174, p. 1109-18.
- Eshghi, A., Pinne, M., Haake, D.A., Zuerner, R.L., Frank, A., and Cameron, C.E., 2012, Methylation and in vivo expression of the surface-exposed *Leptospira interrogans* outer-membrane protein OmpL32: *Microbiology*, v. 158, p. 622-35.
- Espeli, O., and Boccard, F., 1997, In vivo cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: genetic characterization of the target and identification of the cut site: *Mol Microbiol*, v. 26, p. 767-77.
- Espeli, O., Mercier, R., and Boccard, F., 2008, DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome: *Mol Microbiol*, v. 68, p. 1418-27.
- Esposito, D., Petrovic, A., Harris, R., Ono, S., Eccleston, J.F., Mbabaali, A., Haq, I., Higgins, C.F., Hinton, J.C., Driscoll, P.C., and Ladbury, J.E., 2002, H-NS oligomerization domain structure reveals the mechanism for high order self-association of the intact protein: *J Mol Biol*, v. 324, p. 841-50.
- Evinger, M., and Agabian, N., 1977, Envelope-associated nucleoid from *Caulobacter crescentus* stalked and swarmer cells: *J Bacteriol*, v. 132, p. 294-301.
- Facey, P.D., Hitchings, M.D., Saavedra-Garcia, P., Fernandez-Martinez, L., Dyson, P.J., and Del Sol, R., 2009, *Streptomyces coelicolor* Dps-like proteins: differential dual roles in response to stress during vegetative growth and in nucleoid condensation during reproductive cell division: *Mol Microbiol*, v. 73, p. 1186-202.
- Facey, P.D., Sevcikova, B., Novakova, R., Hitchings, M.D., Crack, J.C., Kormanec, J., Dyson, P.J., and Del Sol, R., 2011, The *dpsA* gene of *Streptomyces coelicolor*: induction of expression from a single promoter in response to environmental stress or during development: *PLoS One*, v. 6, p. e25593.
- Fairhead, H., Setlow, B., and Setlow, P., 1993, Prevention of DNA damage in spores and in vitro by small, acid-soluble proteins from *Bacillus* species: *J Bacteriol*, v. 175, p. 1367-74.
- Frenkiel-Krispin, D., Ben-Avraham, I., Englander, J., Shimon, E., Wolf, S.G., and Minsky, A., 2004, Nucleoid restructuring in stationary-state bacteria: *Mol Microbiol*, v. 51, p. 395-405.
- Frye, R.A., 2000, Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins: *Biochem Biophys Res Commun*, v. 273, p. 793-8.
- Gao, R., Mack, T.R., and Stock, A.M., 2007, Bacterial response regulators: versatile regulatory strategies from common domains: *Trends Biochem Sci*, v. 32, p. 225-34.
- Garcia, J., Cordeiro, T.N., Nieto, J.M., Pons, I., Juarez, A., and Pons, M., 2005, Interaction between the bacterial nucleoid associated proteins Hha and H-NS involves a conformational change of Hha: *Biochem J*, v. 388, p. 755-62.
- Garrido, N., Griparic, L., Jokitalo, E., Wartiovaara, J., van der Blik, A.M., and Spelbrink, J.N., 2003, Composition and dynamics of human mitochondrial nucleoids: *Mol Biol Cell*, v. 14, p. 1583-96.
- Gaston, K., Bell, A., Busby, S., and Fried, M., 1992, A comparison of the DNA bending activities of the DNA binding proteins CRP and TFIID: *Nucleic Acids Res*, v. 20, p. 3391-6.
- Geiduschek, E.P., Schneider, G.J., and Sayre, M.H., 1990, TF1, a bacteriophage-specific DNA-binding and DNA-bending protein: *J Struct Biol*, v. 104, p. 84-90.
- Ghangas, G.S., and Wilson, D.B., 1987, Expression of a *Thermomonospora fusca* Cellulase Gene in *Streptomyces lividans* and *Bacillus subtilis*: *Appl Environ Microbiol*, v. 53, p. 1470-5.

- Ghosh, S., and Grove, A., 2006, The *Deinococcus radiodurans*-encoded HU protein has two DNA-binding domains: *Biochemistry*, v. 45, p. 1723-33.
- Giangrossi, M., Giuliodori, A.M., Gualerzi, C.O., and Pon, C.L., 2002, Selective expression of the beta-subunit of nucleoid-associated protein HU during cold shock in *Escherichia coli*: *Mol Microbiol*, v. 44, p. 205-16.
- Goff, C.G., and Setzer, J., 1980, ADP ribosylation of *Escherichia coli* RNA polymerase is nonessential for bacteriophage T4 development: *J Virol*, v. 33, p. 547-9.
- Goldberg, M.D., Johnson, M., Hinton, J.C., and Williams, P.H., 2001, Role of the nucleoid-associated protein Fis in the regulation of virulence properties of enteropathogenic *Escherichia coli*: *Mol Microbiol*, v. 41, p. 549-59.
- Gonzalez-Ceron, G., Miranda-Olivares, O.J., and Servin-Gonzalez, L., 2009, Characterization of the methyl-specific restriction system of *Streptomyces coelicolor* A3(2) and of the role played by laterally acquired nucleases: *FEMS Microbiol Lett*, v. 301, p. 35-43.
- Goodrich, J.A., Schwartz, M.L., and McClure, W.R., 1990, Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF): *Nucleic Acids Res*, v. 18, p. 4993-5000.
- Gordon, B.R., Li, Y., Cote, A., Weirauch, M.T., Ding, P., Hughes, T.R., Navarre, W.W., Xia, B., and Liu, J., 2011, Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins: *Proc Natl Acad Sci U S A*, v. 108, p. 10690-5.
- Gottesman, S., McCullen, C.A., Guillier, M., Vanderpool, C.K., Majdalani, N., Benhammou, J., Thompson, K.M., Fitzgerald, P.C., Sowa, N.A., and Fitzgerald, D.J., 2006, Small RNA regulators and the bacterial response to stress: *Cold Spring Harb Symp Quant Biol*, v. 71, p. 1-11.
- Grainger, D.C., Hurd, D., Goldberg, M.D., and Busby, S.J., 2006, Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome: *Nucleic Acids Res*, v. 34, p. 4642-52.
- Gregoret, I.V., Lee, Y.M., and Goodson, H.V., 2004, Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis: *J Mol Biol*, v. 338, p. 17-31.
- 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*, v. 185, p. 5320-3.
- Griffith, J., Makhov, A., Santiago-Lara, L., and Setlow, P., 1994, Electron microscopic studies of the interaction between a *Bacillus subtilis* alpha/beta-type small, acid-soluble spore protein with DNA: protein binding is cooperative, stiffens the DNA, and induces negative supercoiling: *Proc Natl Acad Sci U S A*, v. 91, p. 8224-8.
- Grishin, N.V., 2001, KH domain: one motif, two folds: *Nucleic Acids Res*, v. 29, p. 638-43.
- Grove, A., 2011, Functional evolution of bacterial histone-like HU proteins: *Curr Issues Mol Biol*, v. 13, p. 1-12.
- Grove, A., and Wilkinson, S.P., 2005, Differential DNA binding and protection by dimeric and dodecameric forms of the ferritin homolog Dps from *Deinococcus radiodurans*: *J Mol Biol*, v. 347, p. 495-508.
- Guan, K.L., Yu, W., Lin, Y., Xiong, Y., and Zhao, S., 2010, Generation of acetyllysine antibodies and affinity enrichment of acetylated peptides: *Nat Protoc*, v. 5, p. 1583-95.
- Guillen, N., Le Hegaret, F., Fleury, A.M., and Hirschbein, L., 1978, Folded chromosomes of vegetative *Bacillus subtilis*: composition and properties: *Nucleic Acids Res*, v. 5, p. 475-89.
- Guo, F., and Adhya, S., 2007, Spiral structure of *Escherichia coli* HUalpha-beta provides foundation for DNA supercoiling: *Proc Natl Acad Sci U S A*, v. 104, p. 4309-14.
- Gupta, S., and Chatterji, D., 2003, Bimodal protection of DNA by *Mycobacterium smegmatis* DNA-binding protein from stationary phase cells: *J Biol Chem*, v. 278, p. 5235-41.
- 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*, v. 100, p. 1541-6.

- Hajdu, J., Bartha, F., and Friedrich, P., 1976, Crosslinking with bifunctional reagents as a means for studying the symmetry of oligomeric proteins: *Eur J Biochem*, v. 68, p. 373-83.
- Haniford, D.B., 2006, Transpososome dynamics and regulation in Tn10 transposition: *Crit Rev Biochem Mol Biol*, v. 41, p. 407-24.
- Hansen, A.M., and Kaper, J.B., 2009, Hfq affects the expression of the LEE pathogenicity island in enterohaemorrhagic *Escherichia coli*: *Mol Microbiol*, v. 73, p. 446-65.
- Harrington, E.W., and Trun, N.J., 1997, Unfolding of the bacterial nucleoid both in vivo and in vitro as a result of exposure to camphor: *J Bacteriol*, v. 179, p. 2435-9.
- Hart, B.R., and Blumenthal, R.M., 2011, Unexpected coregulator range for the global regulator Lrp of *Escherichia coli* and *Proteus mirabilis*: *J Bacteriol*, v. 193, p. 1054-64.
- Hatfield, G.W., and Benham, C.J., 2002, DNA topology-mediated control of global gene expression in *Escherichia coli*: *Annu Rev Genet*, v. 36, p. 175-203.
- Hesketh, A.R., Chandra, G., Shaw, A.D., Rowland, J.J., Kell, D.B., Bibb, M.J., and Chater, K.F., 2002, Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor*: *Mol Microbiol*, v. 46, p. 917-32.
- Hirsch, C.F., and Ensign, J.C., 1976, Nutritionally defined conditions for germination of *Streptomyces viridochromogenes* spores: *J Bacteriol*, v. 126, p. 13-23.
- Hiszczynska-Sawicka, E., and Kur, J., 1997, Effect of *Escherichia coli* IHF mutations on plasmid p15A copy number: *Plasmid*, v. 38, p. 174-9.
- Hong, H.J., Hutchings, M.I., Hill, L.M., and Buttner, M.J., 2005, The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*: *J Biol Chem*, v. 280, p. 13055-61.
- Hopwood, D.A., and Glauert, A.M., 1960a, The fine structure of *Streptomyces coelicolor*. II. The nuclear material: *J Biophys Biochem Cytol*, v. 8, p. 267-78.
- , 1960b, Observations on the chromatinic bodies of *Streptomyces coelicolor*: *J Biophys Biochem Cytol*, v. 8, p. 257-65.
- Horinouchi, S., Malpartida, F., Hopwood, D.A., and Beppu, T., 1989, afsB stimulates transcription of the actinorhodin biosynthetic pathway in *Streptomyces coelicolor* A3(2) and *Streptomyces lividans*: *Mol Gen Genet*, v. 215, p. 355-7.
- Horz, W., Fittler, F., and Zachau, H.G., 1983, Sequence specific cleavage of African green monkey alpha-satellite DNA by micrococcal nuclease: *Nucleic Acids Res*, v. 11, p. 4275-85.
- Hoskisson, P.A., and Hutchings, M.I., 2006, MtrAB-LpqB: a conserved three-component system in actinobacteria?: *Trends Microbiol*, v. 14, p. 444-9.
- Hsieh, L.S., Rouviere-Yaniv, J., and Drlica, K., 1991, Bacterial DNA supercoiling and [ATP]/[ADP] ratio: changes associated with salt shock: *J Bacteriol*, v. 173, p. 3914-7.
- Hsu, Y.H., Chung, M.W., and Li, T.K., 2006, Distribution of gyrase and topoisomerase IV on bacterial nucleoid: implications for nucleoid organization: *Nucleic Acids Res*, v. 34, p. 3128-38.
- Hu, K.H., Liu, E., Dean, K., Gingras, M., DeGraff, W., and Trun, N.J., 1996, Overproduction of three genes leads to camphor resistance and chromosome condensation in *Escherichia coli*: *Genetics*, v. 143, p. 1521-32.
- Huang, L., Tsui, P., and Freundlich, M., 1990, Integration host factor is a negative effector of in vivo and in vitro expression of ompC in *Escherichia coli*: *J Bacteriol*, v. 172, p. 5293-8.
- Huh, J.W., Shima, J., and Ochi, K., 1996, ADP-ribosylation of proteins in *Bacillus subtilis* and its possible importance in sporulation: *J Bacteriol*, v. 178, p. 4935-41.
- Huisman, O., Faalen, M., Girard, D., Jaffe, A., Toussaint, A., and Rouviere-Yaniv, J., 1989, Multiple defects in *Escherichia coli* mutants lacking HU protein: *J Bacteriol*, v. 171, p. 3704-12.
- Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M., 2005, Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein: *Mol Cell Proteomics*, v. 4, p. 1265-72.
- Jenuwein, T., and Allis, C.D., 2001, Translating the histone code: *Science*, v. 293, p. 1074-80.

- Jeong, K.S., Ahn, J., and Khodursky, A.B., 2004, Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*: *Genome Biol*, v. 5, p. R86.
- Johansson, J., and Uhlin, B.E., 1999, Differential protease-mediated turnover of H-NS and StpA revealed by a mutation altering protein stability and stationary-phase survival of *Escherichia coli*: *Proc Natl Acad Sci U S A*, v. 96, p. 10776-81.
- Jonsbu, E., McIntyre, M., and Nielsen, J., 2002, The influence of carbon sources and morphology on nystatin production by *Streptomyces noursei*: *J Biotechnol*, v. 95, p. 133-44.
- Jyothikumar, V., Tilley, E.J., Wali, R., and Herron, P.R., 2008, Time-lapse microscopy of *Streptomyces coelicolor* growth and sporulation: *Appl Environ Microbiol*, v. 74, p. 6774-81.
- Kaidow, A., Wachi, M., Nakamura, J., Magae, J., and Nagai, K., 1995, Anucleate cell production by *Escherichia coli* delta hns mutant lacking a histone-like protein, H-NS: *J Bacteriol*, v. 177, p. 3589-92.
- Kamashev, D., and Rouviere-Yaniv, J., 2000, The histone-like protein HU binds specifically to DNA recombination and repair intermediates: *EMBO J*, v. 19, p. 6527-35.
- Kar, S., Edgar, R., and Adhya, S., 2005, Nucleoid remodeling by an altered HU protein: reorganization of the transcription program: *Proc Natl Acad Sci U S A*, v. 102, p. 16397-402.
- Karambelkar, S., Swapna, G., and Nagaraja, V., 2012, Silencing of toxic gene expression by Fis: *Nucleic Acids Res*, v. 40, p. 4358-67.
- Karoonuthaisiri, N., Weaver, D., Huang, J., Cohen, S.N., and Kao, C.M., 2005, Regional organization of gene expression in *Streptomyces coelicolor*: *Gene*, v. 353, p. 53-66.
- Katsube, T., Matsumoto, S., Takatsuka, M., Okuyama, M., Ozeki, Y., Naito, M., Nishiuchi, Y., Fujiwara, N., Yoshimura, M., Tsuboi, T., Torii, M., Oshitani, N., Arakawa, T., and Kobayashi, K., 2007, Control of cell wall assembly by a histone-like protein in *Mycobacteria*: *J Bacteriol*, v. 189, p. 8241-9.
- Kavenoff, R., and Bowen, B.C., 1976, Electron microscopy of membrane-free folded chromosomes from *Escherichia coli*: *Chromosoma*, v. 59, p. 89-101.
- Kavenoff, R., and Ryder, O.A., 1976, Electron microscopy of membrane-associated folded chromosomes of *Escherichia coli*: *Chromosoma*, v. 55, p. 13-25.
- Khoury, G.A., Baliban, R.C., and Floudas, C.A., 2011, Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database: *Sci Rep*, v. 1.
- Kienhofer, J., Haussler, D.J., Ruckelshausen, F., Muessig, E., Weber, K., Pimentel, D., Ullrich, V., Burkle, A., and Bachschmid, M.M., 2009, Association of mitochondrial antioxidant enzymes with mitochondrial DNA as integral nucleoid constituents: *FASEB J*, v. 23, p. 2034-44.
- Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A., 2000, *Practical Streptomyces Genetics*, John Innes Foundation.
- Kim, D.W., Chater, K., Lee, K.J., and Hesketh, A., 2005a, Changes in the extracellular proteome caused by the absence of the bldA gene product, a developmentally significant tRNA, reveal a new target for the pleiotropic regulator AdpA in *Streptomyces coelicolor*: *J Bacteriol*, v. 187, p. 2957-66.
- Kim, D.W., Chater, K.F., Lee, K.J., and Hesketh, A., 2005b, Effects of growth phase and the developmentally significant bldA-specified tRNA on the membrane-associated proteome of *Streptomyces coelicolor*: *Microbiology*, v. 151, p. 2707-20.
- Kim, E.S., Hong, H.J., Choi, C.Y., and Cohen, S.N., 2001, Modulation of actinorhodin biosynthesis in *Streptomyces lividans* by glucose repression of afsR2 gene transcription: *J Bacteriol*, v. 183, p. 2198-203.
- Kim, J., Yoshimura, S.H., Hizume, K., Ohniwa, R.L., Ishihama, A., and Takeyasu, K., 2004, Fundamental structural units of the *Escherichia coli* nucleoid revealed by atomic force microscopy: *Nucleic Acids Res*, v. 32, p. 1982-92.

- Kinashi, H., Shimaji-Murayama, M., and Hanafusa, T., 1992, Integration of SCP1, a giant linear plasmid, into the *Streptomyces coelicolor* chromosome: *Gene*, v. 115, p. 35-41.
- Knight, S.W., and Samuels, D.S., 1999, Natural synthesis of a DNA-binding protein from the C-terminal domain of DNA gyrase A in *Borrelia burgdorferi*: *EMBO J*, v. 18, p. 4875-81.
- Kobayashi, T., Takahara, M., Miyagishima, S.Y., Kuroiwa, H., Sasaki, N., Ohta, N., Matsuzaki, M., and Kuroiwa, T., 2002, Detection and localization of a chloroplast-encoded HU-like protein that organizes chloroplast nucleoids: *Plant Cell*, v. 14, p. 1579-89.
- Kobryn, K., Lavoie, B.D., and Chaconas, G., 1999, Supercoiling-dependent site-specific binding of HU to naked Mu DNA: *J Mol Biol*, v. 289, p. 777-84.
- Kohler, P., and Marahiel, M.A., 1998, Mutational analysis of the nucleoid-associated protein HBSu of *Bacillus subtilis*: *Mol Gen Genet*, v. 260, p. 487-91.
- Kohno, K., Yasuzawa, K., Hirose, M., Kano, Y., Goshima, N., Tanaka, H., and Imamoto, F., 1994, Autoregulation of transcription of the *hupA* gene in *Escherichia coli*: evidence for steric hindrance of the functional promoter domains induced by HU: *J Biochem*, v. 115, p. 1113-8.
- Korner, H., Sofia, H.J., and Zumft, W.G., 2003, Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs: *FEMS Microbiol Rev*, v. 27, p. 559-92.
- Kumar, M., Gromiha, M.M., and Raghava, G.P., 2007, Identification of DNA-binding proteins using support vector machines and evolutionary profiles: *BMC Bioinformatics*, v. 8, p. 463.
- Kumar, S., Sardesai, A.A., Basu, D., Muniyappa, K., and Hasnain, S.E., 2010, DNA clasping by mycobacterial HU: the C-terminal region of HupB mediates increased specificity of DNA binding: *PLoS One*, v. 5.
- Kuroiwa, T., and Suzuki, T., 1981, Circular nucleoids isolated from chloroplasts in a brown alga *Ectocarpus siliculosus*: *Exp Cell Res*, v. 134, p. 457-61.
- Lane, A.A., and Chabner, B.A., 2009, Histone deacetylase inhibitors in cancer therapy: *J Clin Oncol*, v. 27, p. 5459-68.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L., 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome: *Genome Biol*, v. 10, p. R25.
- Lautier, T., and Nasser, W., 2007, The DNA nucleoid-associated protein Fis co-ordinates the expression of the main virulence genes in the phytopathogenic bacterium *Erwinia chrysanthemi*: *Mol Microbiol*, v. 66, p. 1474-90.
- Lavoie, B.D., and Chaconas, G., 1993, Site-specific HU binding in the Mu transpososome: conversion of a sequence-independent DNA-binding protein into a chemical nuclease: *Genes Dev*, v. 7, p. 2510-9.
- Lee, B.H., Murugasu-Oei, B., and Dick, T., 1998, Upregulation of a histone-like protein in dormant *Mycobacterium smegmatis*: *Mol Gen Genet*, v. 260, p. 475-9.
- Lee, E.C., Hales, L.M., Gumport, R.I., and Gardner, J.F., 1992, The isolation and characterization of mutants of the integration host factor (IHF) of *Escherichia coli* with altered, expanded DNA-binding specificities: *EMBO J*, v. 11, p. 305-13.
- Lenz, D.H., and Bassler, B.L., 2007, The small nucleoid protein Fis is involved in *Vibrio cholerae* quorum sensing: *Mol Microbiol*, v. 63, p. 859-71.
- Leonard, P.M., Smits, S.H., Sedelnikova, S.E., Brinkman, A.B., de Vos, W.M., van der Oost, J., Rice, D.W., and Rafferty, J.B., 2001, Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*: *EMBO J*, v. 20, p. 990-7.
- Letunic, I., and Bork, P., 2007, Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation: *Bioinformatics*, v. 23, p. 127-8.
- , 2011, Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy: *Nucleic Acids Res*, v. 39, p. W475-8.
- Lewin, A., Baus, D., Kamal, E., Bon, F., Kunisch, R., Maurischat, S., Adonopoulou, M., and Eich, K., 2008, The mycobacterial DNA-binding protein 1 (MDP1) from *Mycobacterium bovis* BCG influences various growth characteristics: *BMC Microbiol*, v. 8, p. 91.

- Lewis, D.E., Geanakopoulou, M., and Adhya, S., 1999, Role of HU and DNA supercoiling in transcription repression: specialized nucleoprotein repression complex at gal promoters in *Escherichia coli*: *Mol Microbiol*, v. 31, p. 451-61.
- Lewis, P.J., Doherty, G.P., and Clarke, J., 2008, Transcription factor dynamics: *Microbiology*, v. 154, p. 1837-44.
- Li, S., and Waters, R., 1998, *Escherichia coli* strains lacking protein HU are UV sensitive due to a role for HU in homologous recombination: *J Bacteriol*, v. 180, p. 3750-6.
- Li, X., 1997, *Streptomyces cellulolyticus* sp. nov., a new cellulolytic member of the genus *Streptomyces*: *Int J Syst Bacteriol*, v. 47, p. 443-5.
- Libby, E.A., Roggiani, M., and Goulian, M., 2012, Membrane protein expression triggers chromosomal locus repositioning in bacteria: *Proc Natl Acad Sci U S A*, v. 109, p. 7445-50.
- Lin, R., Ernsting, B., Hirshfield, I.N., Matthews, R.G., Neidhardt, F.C., Clark, R.L., and Newman, E.B., 1992, The Irp gene product regulates expression of *lysU* in *Escherichia coli* K-12: *J Bacteriol*, v. 174, p. 2779-84.
- Link, T.M., Valentin-Hansen, P., and Brennan, R.G., 2009, Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA: *Proc Natl Acad Sci U S A*, v. 106, p. 19292-7.
- Liu, H., Wang, Q., Liu, Q., Cao, X., Shi, C., and Zhang, Y., 2011, Roles of Hfq in the stress adaptation and virulence in fish pathogen *Vibrio alginolyticus* and its potential application as a target for live attenuated vaccine: *Appl Microbiol Biotechnol*, v. 91, p. 353-64.
- Liu, Y., Chen, H., Kenney, L.J., and Yan, J., 2010a, A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes: *Genes Dev*, v. 24, p. 339-44.
- Liu, Y., Wu, N., Dong, J., Gao, Y., Zhang, X., Mu, C., Shao, N., and Yang, G., 2010b, Hfq is a global regulator that controls the pathogenicity of *Staphylococcus aureus*: *PLoS One*, v. 5.
- Lucchini, S., Rowley, G., Goldberg, M.D., Hurd, D., Harrison, M., and Hinton, J.C., 2006, H-NS mediates the silencing of laterally acquired genes in bacteria: *PLoS Pathog*, v. 2, p. e81.
- Luijsterburg, M.S., Noom, M.C., Wuite, G.J., and Dame, R.T., 2006, The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective: *J Struct Biol*, v. 156, p. 262-72.
- Lukong, K.E., and Richard, S., 2003, Sam68, the KH domain-containing superSTAR: *Biochim Biophys Acta*, v. 1653, p. 73-86.
- Lynch, A.S., and Wang, J.C., 1993, Anchoring of DNA to the bacterial cytoplasmic membrane through cotranscriptional synthesis of polypeptides encoding membrane proteins or proteins for export: a mechanism of plasmid hypernegative supercoiling in mutants deficient in DNA topoisomerase I: *J Bacteriol*, v. 175, p. 1645-55.
- Macchi, R., Montesissa, L., Murakami, K., Ishihama, A., De Lorenzo, V., and Berton, G., 2003, Recruitment of sigma54-RNA polymerase to the Pu promoter of *Pseudomonas putida* through integration host factor-mediated positioning switch of alpha subunit carboxyl-terminal domain on an UP-like element: *J Biol Chem*, v. 278, p. 27695-702.
- MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and MacNeil, T., 1992, Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector: *Gene*, v. 111, p. 61-8.
- Madden, T., Ward, J.M., and Ison, A.P., 1996, Organic acid excretion by *Streptomyces lividans* TK24 during growth on defined carbon and nitrogen sources: *Microbiology*, v. 142 (Pt 11), p. 3181-5.
- Malik, M., Bensaid, A., Rouviere-Yaniv, J., and Drlica, K., 1996, Histone-like protein HU and bacterial DNA topology: suppression of an HU deficiency by gyrase mutations: *J Mol Biol*, v. 256, p. 66-76.
- Mangan, M.W., Lucchini, S., Danino, V., Croinin, T.O., Hinton, J.C., and Dorman, C.J., 2006, The integration host factor (IHF) integrates stationary-phase and virulence gene expression in *Salmonella enterica* serovar Typhimurium: *Mol Microbiol*, v. 59, p. 1831-47.
- Manteca, A., Jung, H.R., Schwammle, V., Jensen, O.N., and Sanchez, J., 2010, Quantitative proteome analysis of *Streptomyces coelicolor* Nonsporulating liquid cultures

- demonstrates a complex differentiation process comparable to that occurring in sporulating solid cultures: *J Proteome Res*, v. 9, p. 4801-11.
- Marsh, V.L., Peak-Chew, S.Y., and Bell, S.D., 2005, Sir2 and the acetyltransferase, Pat, regulate the archaeal chromatin protein, Alba: *J Biol Chem*, v. 280, p. 21122-8.
- Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., Gevaert, K., Vandekerckhove, J., and Apweiler, R., 2005, PRIDE: the proteomics identifications database: *Proteomics*, v. 5, p. 3537-45.
- Matsumoto, S., Furugen, M., Yukitake, H., and Yamada, T., 2000, The gene encoding mycobacterial DNA-binding protein I (MDPI) transformed rapidly growing bacteria to slowly growing bacteria: *FEMS Microbiol Lett*, v. 182, p. 297-301.
- Matsumoto, S., Matsumoto, M., Umemori, K., Ozeki, Y., Furugen, M., Tatsuo, T., Hirayama, Y., Yamamoto, S., Yamada, T., and Kobayashi, K., 2005, DNA augments antigenicity of mycobacterial DNA-binding protein 1 and confers protection against *Mycobacterium tuberculosis* infection in mice: *J Immunol*, v. 175, p. 441-9.
- Maurer, S., Fritz, J., and Muskhelishvili, G., 2009, A systematic in vitro study of nucleoprotein complexes formed by bacterial nucleoid-associated proteins revealing novel types of DNA organization: *J Mol Biol*, v. 387, p. 1261-76.
- Mayer, O., Rajkowitsch, L., Lorenz, C., Konrat, R., and Schroeder, R., 2007, RNA chaperone activity and RNA-binding properties of the *E. coli* protein StpA: *Nucleic Acids Res*, v. 35, p. 1257-69.
- McArthur, M., and Bibb, M., 2006, In vivo DNase I sensitivity of the *Streptomyces coelicolor* chromosome correlates with gene expression: implications for bacterial chromosome structure: *Nucleic Acids Res*, v. 34, p. 5395-401.
- McKenzie, N.L., and Nodwell, J.R., 2007, Phosphorylated AbsA2 negatively regulates antibiotic production in *Streptomyces coelicolor* through interactions with pathway-specific regulatory gene promoters: *J Bacteriol*, v. 189, p. 5284-92.
- McLeod, S.M., and Johnson, R.C., 2001, Control of transcription by nucleoid proteins: *Curr Opin Microbiol*, v. 4, p. 152-9.
- McLeod, S.M., Xu, J., Cramton, S.E., Gaal, T., Gourse, R.L., and Johnson, R.C., 1999, Localization of amino acids required for Fis to function as a class II transcriptional activator at the RpoS-dependent proP P2 promoter: *J Mol Biol*, v. 294, p. 333-46.
- McLeod, S.M., Xu, J., and Johnson, R.C., 2000, Coactivation of the RpoS-dependent proP P2 promoter by fis and cyclic AMP receptor protein: *J Bacteriol*, v. 182, p. 4180-7.
- Melzoch, K., de Mattos, M.J., and Neijssel, O.M., 1997, Production of actinorhodin by *Streptomyces coelicolor* A3(2) grown in chemostat culture: *Biotechnol Bioeng*, v. 54, p. 577-82.
- Mercante, J., Edwards, A.N., Dubey, A.K., Babitzke, P., and Romeo, T., 2009, Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression: *J Mol Biol*, v. 392, p. 511-28.
- Michels, P.C., Khmel'nitsky, Y.L., Dordick, J.S., and Clark, D.S., 1998, Combinatorial biocatalysis: a natural approach to drug discovery: *Trends Biotechnol*, v. 16, p. 210-5.
- Micka, B., and Marahiel, M.A., 1992, The DNA-binding protein HBSu is essential for normal growth and development in *Bacillus subtilis*: *Biochimie*, v. 74, p. 641-50.
- Mikulik, K., Felsberg, J., Kudrnacova, E., Bezouskova, S., Setinova, D., Stodulkova, E., Zidkova, J., and Zidek, V., 2012, CobB1 deacetylase activity in *Streptomyces coelicolor*: *Biochem Cell Biol*, v. 90, p. 179-87.
- Miyabe, I., Zhang, Q.M., Kano, Y., and Yonei, S., 2000, Histone-like protein HU is required for recA gene-dependent DNA repair and SOS induction pathways in UV-irradiated *Escherichia coli*: *Int J Radiat Biol*, v. 76, p. 43-9.
- Moffatt, B.A., and Studier, F.W., 1987, T7 lysozyme inhibits transcription by T7 RNA polymerase: *Cell*, v. 49, p. 221-7.

- Mohanty, B.K., Maples, V.F., and Kushner, S.R., 2004, The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in *Escherichia coli*: *Mol Microbiol*, v. 54, p. 905-20.
- Mohr, S.C., Sokolov, N.V., He, C.M., and Setlow, P., 1991, Binding of small acid-soluble spore proteins from *Bacillus subtilis* changes the conformation of DNA from B to A: *Proc Natl Acad Sci U S A*, v. 88, p. 77-81.
- Mojica, F.J., and Higgins, C.F., 1997, In vivo supercoiling of plasmid and chromosomal DNA in an *Escherichia coli* hns mutant: *J Bacteriol*, v. 179, p. 3528-33.
- Molle, V., and Buttner, M.J., 2000, Different alleles of the response regulator gene *bldM* arrest *Streptomyces coelicolor* development at distinct stages: *Mol Microbiol*, v. 36, p. 1265-78.
- Moore, J.M., Bradshaw, E., Seipke, R.F., Hutchings, M.I., and McArthur, M., 2013, Use and Discovery of Chemical Elicitors That Stimulate Biosynthetic Gene Clusters in *Streptomyces* Bacteria, *Methods in Enzymology*, Volume 517.
- Morita, T., Maki, K., and Aiba, H., 2005, RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs: *Genes Dev*, v. 19, p. 2176-86.
- Mukherjee, A., Bhattacharyya, G., and Grove, A., 2008, The C-terminal domain of HU-related histone-like protein Hlp from *Mycobacterium smegmatis* mediates DNA end-joining: *Biochemistry*, v. 47, p. 8744-53.
- Mukherjee, S., Hao, Y.H., and Orth, K., 2007, A newly discovered post-translational modification--the acetylation of serine and threonine residues: *Trends Biochem Sci*, v. 32, p. 210-6.
- Muller, C.M., Dobrindt, U., Nagy, G., Emdy, L., Uhlin, B.E., and Hacker, J., 2006, Role of histone-like proteins H-NS and StpA in expression of virulence determinants of uropathogenic *Escherichia coli*: *J Bacteriol*, v. 188, p. 5428-38.
- Murphy, L.D., and Zimmerman, S.B., 1997, Isolation and characterization of spermidine nucleoids from *Escherichia coli*: *J Struct Biol*, v. 119, p. 321-35.
- Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kuromori, T., Motohashi, R., Shono, Y., Nagata, N., Ikeuchi, M., and Shinozaki, K., 2008, A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in *Arabidopsis*: *Plant Cell*, v. 20, p. 3148-62.
- Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S.J., and Fang, F.C., 2006, Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*: *Science*, v. 313, p. 236-8.
- Newman, E.B., and Lin, R., 1995, Leucine-responsive regulatory protein: a global regulator of gene expression in *E. coli*: *Annu Rev Microbiol*, v. 49, p. 747-75.
- Nguyen, K.T., Piastro, K., Gray, T.A., and Derbyshire, K.M., 2010, *Mycobacterial* biofilms facilitate horizontal DNA transfer between strains of *Mycobacterium smegmatis*: *J Bacteriol*, v. 192, p. 5134-42.
- Nguyen, K.T., Willey, J.M., Nguyen, L.D., Nguyen, L.T., Viollier, P.H., and Thompson, C.J., 2002, A central regulator of morphological differentiation in the multicellular bacterium *Streptomyces coelicolor*: *Mol Microbiol*, v. 46, p. 1223-38.
- Nielsen, J.S., Boggild, A., Andersen, C.B., Nielsen, G., Boysen, A., Brodersen, D.E., and Valentin-Hansen, P., 2007, An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from *Methanococcus jannaschii*: *RNA*, v. 13, p. 2213-23.
- Niki, H., and Hiraga, S., 1998, Polar localization of the replication origin and terminus in *Escherichia coli* nucleoids during chromosome partitioning: *Genes Dev*, v. 12, p. 1036-45.
- Niki, H., Yamaichi, Y., and Hiraga, S., 2000, Dynamic organization of chromosomal DNA in *Escherichia coli*: *Genes Dev*, v. 14, p. 212-23.
- Nishida, S., Mizushima, T., Miki, T., and Sekimizu, K., 1997, Immotile phenotype of an *Escherichia coli* mutant lacking the histone-like protein HU: *FEMS Microbiol Lett*, v. 150, p. 297-301.
- Nomoto, R., Tezuka, T., Miyazono, K.I., Tanokura, M., Horinouchi, S., and Ohnishi, Y., 2012, Purification, crystallization and preliminary X-ray analysis of SGR6054, a *Streptomyces*

- homologue of the mycobacterial integration host factor mIHF: *Acta Crystallogr Sect F Struct Biol Cryst Commun*, v. 68, p. 1085-8.
- Noom, M.C., Navarre, W.W., Oshima, T., Wuite, G.J., and Dame, R.T., 2007, H-NS promotes looped domain formation in the bacterial chromosome: *Curr Biol*, v. 17, p. R913-4.
- Nou, X., Braaten, B., Kaltenbach, L., and Low, D.A., 1995, Differential binding of Lrp to two sets of pap DNA binding sites mediated by Pap I regulates Pap phase variation in *Escherichia coli*: *EMBO J*, v. 14, p. 5785-97.
- Novotna, J., Vohradsky, J., Berndt, P., Gramajo, H., Langen, H., Li, X.M., Minas, W., Orsaria, L., Roeder, D., and Thompson, C.J., 2003, Proteomic studies of diauxic lag in the differentiating prokaryote *Streptomyces coelicolor* reveal a regulatory network of stress-induced proteins and central metabolic enzymes: *Mol Microbiol*, v. 48, p. 1289-303.
- Oberto, J., Drlica, K., and Rouviere-Yaniv, J., 1994, Histones, HMG, HU, IHF: Meme combat: *Biochimie*, v. 76, p. 901-8.
- Oberto, J., Nabti, S., Jooste, V., Mignot, H., and Rouviere-Yaniv, J., 2009, The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction: *PLoS One*, v. 4, p. e4367.
- Ohniwa, R.L., Ushijima, Y., Saito, S., and Morikawa, K., 2011, Proteomic analyses of nucleoid-associated proteins in *Escherichia coli*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, and *Staphylococcus aureus*: *PLoS One*, v. 6, p. e19172.
- Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M., 2001, Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites: *Proc Natl Acad Sci U S A*, v. 98, p. 12215-20.
- Osterberg, S., del Peso-Santos, T., and Shingler, V., 2011, Regulation of alternative sigma factor use: *Annu Rev Microbiol*, v. 65, p. 37-55.
- Osuna, R., Lienau, D., Hughes, K.T., and Johnson, R.C., 1995, Sequence, regulation, and functions of fis in *Salmonella typhimurium*: *J Bacteriol*, v. 177, p. 2021-32.
- Otaka, H., Ishikawa, H., Morita, T., and Aiba, H., 2011, PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action: *Proc Natl Acad Sci U S A*, v. 108, p. 13059-64.
- Paget, M.S., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J., 1999, Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2): *J Bacteriol*, v. 181, p. 204-11.
- Painbeni, E., Caroff, M., and Rouviere-Yaniv, J., 1997, Alterations of the outer membrane composition in *Escherichia coli* lacking the histone-like protein HU: *Proc Natl Acad Sci U S A*, v. 94, p. 6712-7.
- Paleckova, P., Kontrova, F., Kofronova, O., Bobek, J., Benada, O., and Mikulik, K., 2007, Effect of protein kinase inhibitors on protein phosphorylation and germination of aerial spores from *Streptomyces coelicolor*: *Folia Microbiol (Praha)*, v. 52, p. 215-22.
- Panek, J., Bobek, J., Mikulik, K., Basler, M., and Vohradsky, J., 2008, Biocomputational prediction of small non-coding RNAs in *Streptomyces*: *BMC Genomics*, v. 9, p. 217.
- Park, S.S., Yang, Y.H., Song, E., Kim, E.J., Kim, W.S., Sohng, J.K., Lee, H.C., Liou, K.K., and Kim, B.G., 2009, Mass spectrometric screening of transcriptional regulators involved in antibiotic biosynthesis in *Streptomyces coelicolor* A3(2): *J Ind Microbiol Biotechnol*, v. 36, p. 1073-83.
- Park, S.T., Kang, C.M., and Husson, R.N., 2008, Regulation of the SigH stress response regulon by an essential protein kinase in *Mycobacterium tuberculosis*: *Proc Natl Acad Sci U S A*, v. 105, p. 13105-10.
- Parker, J.L., Jones, A.M., Serazetdinova, L., Saalbach, G., Bibb, M.J., and Naldrett, M.J., 2010, Analysis of the phosphoproteome of the multicellular bacterium *Streptomyces coelicolor* A3(2) by protein/peptide fractionation, phosphopeptide enrichment and high-accuracy mass spectrometry: *Proteomics*, v. 10, p. 2486-97.

- Pedulla, M.L., and Hatfull, G.F., 1998, Characterization of the mIHf gene of *Mycobacterium smegmatis*: J Bacteriol, v. 180, p. 5473-7.
- Pedulla, M.L., Lee, M.H., Lever, D.C., and Hatfull, G.F., 1996, A novel host factor for integration of mycobacteriophage L5: Proc Natl Acad Sci U S A, v. 93, p. 15411-6.
- Peekhaus, N., Tolner, B., Poolman, B., and Kramer, R., 1995, The glutamate uptake regulatory protein (Grp) of *Zymomonas mobilis* and its relation to the global regulator Lrp of *Escherichia coli*: J Bacteriol, v. 177, p. 5140-7.
- Penn, K., Jenkins, C., Nett, M., Udway, D.W., Gontang, E.A., McGlinchey, R.P., Foster, B., Lapidus, A., Podell, S., Allen, E.E., Moore, B.S., and Jensen, P.R., 2009, Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria: ISME J, v. 3, p. 1193-203.
- Perugini, G., Valenti, A., D'Amaro, A., Rossi, M., and Ciaramella, M., 2009, Reverse gyrase and genome stability in hyperthermophilic organisms: Biochem Soc Trans, v. 37, p. 69-73.
- Peter, B.J., Arsuaga, J., Breier, A.M., Khodursky, A.B., Brown, P.O., and Cozzarelli, N.R., 2004, Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*: Genome Biol, v. 5, p. R87.
- Peterson, S.N., Dahlquist, F.W., and Reich, N.O., 2007, The role of high affinity non-specific DNA binding by Lrp in transcriptional regulation and DNA organization: J Mol Biol, v. 369, p. 1307-17.
- Pethe, K., Bifani, P., Drobecq, H., Sergheraert, C., Debie, A.S., Loch, C., and Menozzi, F.D., 2002, Mycobacterial heparin-binding hemagglutinin and laminin-binding protein share antigenic methyllysines that confer resistance to proteolysis: Proc Natl Acad Sci U S A, v. 99, p. 10759-64.
- Piette, A., Derouaux, A., Gerkens, P., Noens, E.E., Mazzucchelli, G., Vion, S., Koerten, H.K., Titgemeyer, F., De Pauw, E., Leprince, P., van Wezel, G.P., Galleni, M., and Rigali, S., 2005, From dormant to germinating spores of *Streptomyces coelicolor* A3(2): new perspectives from the crp null mutant: J Proteome Res, v. 4, p. 1699-708.
- Pope, M.K., Green, B.D., and Westpheling, J., 1996, The bld mutants of *Streptomyces coelicolor* are defective in the regulation of carbon utilization, morphogenesis and cell-cell signalling: Mol Microbiol, v. 19, p. 747-56.
- Portugal, M.I., Todeschini, A.R., de Lima, C.S., Silva, C.A., Mohana-Borges, R., Ottenhoff, T.H., Mendonca-Previato, L., Previato, J.O., and Pessolani, M.C., 2008, Characterization of two heparan sulphate-binding sites in the mycobacterial adhesin Hlp: BMC Microbiol, v. 8, p. 75.
- Postow, L., Hardy, C.D., Arsuaga, J., and Cozzarelli, N.R., 2004, Topological domain structure of the *Escherichia coli* chromosome: Genes Dev, v. 18, p. 1766-79.
- Prabhakar, S., Mishra, A., Singhal, A., Katoch, V.M., Thakral, S.S., Tyagi, J.S., and Prasad, H.K., 2004, Use of the hupB gene encoding a histone-like protein of *Mycobacterium tuberculosis* as a target for detection and differentiation of *M. tuberculosis* and *M. bovis*: J Clin Microbiol, v. 42, p. 2724-32.
- Prieto, A.I., Kahramanoglou, C., Ali, R.M., Fraser, G.M., Seshasayee, A.S., and Luscombe, N.M., 2012, Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12: Nucleic Acids Res, v. 40, p. 3524-37.
- Prujijn, G.J., 2005, Doughnuts dealing with RNA: Nat Struct Mol Biol, v. 12, p. 562-4.
- Qiao, X., Sun, Y., Qiao, J., and Mindich, L., 2008, The role of host protein YajQ in the temporal control of transcription in bacteriophage Phi6: Proc Natl Acad Sci U S A, v. 105, p. 15956-60.
- Rabhi, M., Espeli, O., Schwartz, A., Cayrol, B., Rahmouni, A.R., Arluison, V., and Boudvillain, M., 2011, The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators: EMBO J, v. 30, p. 2805-16.
- Raghavan, R., Groisman, E.A., and Ochman, H., 2011, Genome-wide detection of novel regulatory RNAs in *E. coli*: Genome Res, v. 21, p. 1487-97.

- Rajkowitsch, L., and Schroeder, R., 2007, Dissecting RNA chaperone activity: RNA, v. 13, p. 2053-60.
- Ramagopal, S., and Subramanian, A.R., 1974, Alteration in the acetylation level of ribosomal protein L12 during growth cycle of *Escherichia coli*: Proc Natl Acad Sci U S A, v. 71, p. 2136-40.
- Ramos, C.G., Sousa, S.A., Grilo, A.M., Feliciano, J.R., and Leitao, J.H., 2011, The second RNA chaperone, Hfq2, is also required for survival under stress and full virulence of *Burkholderia cenocepacia* J2315: J Bacteriol, v. 193, p. 1515-26.
- Reiff, S.B., Vaishnav, S., and Striepen, B., 2012, The HU protein is important for apicoplast genome maintenance and inheritance in *Toxoplasma gondii*: Eukaryot Cell, v. 11, p. 905-15.
- Reynolds, R.C., Ananthan, S., Faaleolea, E., Hobrath, J.V., Kwong, C.D., Maddox, C., Rasmussen, L., Sosa, M.I., Thammasuvimol, E., White, E.L., Zhang, W., and Secrist, J.A., 3rd, 2012, High throughput screening of a library based on kinase inhibitor scaffolds against *Mycobacterium tuberculosis* H37Rv: Tuberculosis (Edinb), v. 92, p. 72-83.
- Rice, P.A., Yang, S., Mizuuchi, K., and Nash, H.A., 1996, Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn: Cell, v. 87, p. 1295-306.
- Rocha, E.P., 2004, The replication-related organization of bacterial genomes: Microbiology, v. 150, p. 1609-27.
- Rochman, M., Aviv, M., Glaser, G., and Muskhelishvili, G., 2002, Promoter protection by a transcription factor acting as a local topological homeostat: EMBO Rep, v. 3, p. 355-60.
- Romeo, T., 1998, Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB: Mol Microbiol, v. 29, p. 1321-30.
- Sabo, P.J., Kuehn, M.S., Thurman, R., Johnson, B.E., Johnson, E.M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., Weaver, M., Shafer, A., Lee, K., Neri, F., Humbert, R., Singer, M.A., Richmond, T.A., Dorschner, M.O., McArthur, M., Hawrylycz, M., Green, R.D., Navas, P.A., Noble, W.S., and Stamatoyannopoulos, J.A., 2006, Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays: Nat Methods, v. 3, p. 511-8.
- Salerno, P., Larsson, J., Bucca, G., Laing, E., Smith, C.P., and Flardh, K., 2009, One of the two genes encoding nucleoid-associated HU proteins in *Streptomyces coelicolor* is developmentally regulated and specifically involved in spore maturation: J Bacteriol, v. 191, p. 6489-500.
- Sambrook, J., MacCallum, P., and Russell, D., 2001, Molecular Cloning: A Laboratory Manual,, Cold Spring Harbor Laboratory Press.
- Santero, E., Hoover, T.R., North, A.K., Berger, D.K., Porter, S.C., and Kustu, S., 1992, Role of integration host factor in stimulating transcription from the sigma 54-dependent nifH promoter: J Mol Biol, v. 227, p. 602-20.
- Sarfert, E., Zimmer, C., Gumpert, J., and Hartmut, L., 1983, Folded chromosome structure and DNA-binding protein of *Streptomyces hygroscopicus*: Biochim Biophys Acta, v. 740, p. 118-124.
- Sasaki, N., Kuroiwa, H., Nishitani, C., Takano, H., Higashiyama, T., Kobayashi, T., Shirai, Y., Sakai, A., Kawano, S., Murakami-Murofushi, K., and Kuroiwa, T., 2003, Glom is a novel mitochondrial DNA packaging protein in *Physarum polycephalum* and causes intense chromatin condensation without suppressing DNA functions: Mol Biol Cell, v. 14, p. 4758-69.
- Sauer, E., and Weichenrieder, O., 2011, Structural basis for RNA 3'-end recognition by Hfq: Proc Natl Acad Sci U S A, v. 108, p. 13065-70.
- Sawai, R., Suzuki, A., Takano, Y., Lee, P.C., and Horinouchi, S., 2004, Phosphorylation of AfsR by multiple serine/threonine kinases in *Streptomyces coelicolor* A3(2): Gene, v. 334, p. 53-61.
- Schneider, R., Travers, A., Kutateladze, T., and Muskhelishvili, G., 1999, A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*: Mol Microbiol, v. 34, p. 953-64.

- Schneider, R., Travers, A., and Muskhelishvili, G., 1997, FIS modulates growth phase-dependent topological transitions of DNA in *Escherichia coli*: *Mol Microbiol*, v. 26, p. 519-30.
- Schroder, O., and Wagner, R., 2000, The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex: *J Mol Biol*, v. 298, p. 737-48.
- Schumacher, M.A., Pearson, R.F., Moller, T., Valentin-Hansen, P., and Brennan, R.G., 2002, Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein: *EMBO J*, v. 21, p. 3546-56.
- Sharadamma, N., Harshavardhana, Y., Singh, P., and Muniyappa, K., 2010, Mycobacterium tuberculosis nucleoid-associated DNA-binding protein H-NS binds with high-affinity to the Holliday junction and inhibits strand exchange promoted by RecA protein: *Nucleic Acids Res*, v. 38, p. 3555-69.
- Sheridan, S.D., Benham, C.J., and Hatfield, G.W., 1998, Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activating sequence: *J Biol Chem*, v. 273, p. 21298-308.
- Shima, J., Penyige, A., and Ochi, K., 1996, Changes in patterns of ADP-ribosylated proteins during differentiation of *Streptomyces coelicolor* A3(2) and its development mutants: *J Bacteriol*, v. 178, p. 3785-90.
- Shimoji, Y., Ng, V., Matsumura, K., Fischetti, V.A., and Rambukkana, A., 1999, A 21-kDa surface protein of *Mycobacterium leprae* binds peripheral nerve laminin-2 and mediates Schwann cell invasion: *Proc Natl Acad Sci U S A*, v. 96, p. 9857-62.
- Sinden, R.R., and Pettijohn, D.E., 1981, Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling: *Proc Natl Acad Sci U S A*, v. 78, p. 224-8.
- Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J., 2007, The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*: *Mol Microbiol*, v. 63, p. 193-217.
- Skenneron, C.T., Angly, F.E., Breitbart, M., Bragg, L., He, S., McMahon, K.D., Hugenholtz, P., and Tyson, G.W., 2011, Phage encoded H-NS: a potential achilles heel in the bacterial defence system: *PLoS One*, v. 6, p. e20095.
- Smith, D.D., Wood, N.J., and Hodgson, D.A., 1995, Interaction between primary and secondary metabolism in *Streptomyces coelicolor* A3(2): role of pyrroline-5-carboxylate dehydrogenase: *Microbiology*, v. 141 (Pt 7), p. 1739-44.
- Smits, W.K., and Grossman, A.D., 2010, The transcriptional regulator Rok binds A+T-rich DNA and is involved in repression of a mobile genetic element in *Bacillus subtilis*: *PLoS Genet*, v. 6, p. e1001207.
- Soares de Lima, C., Zulianello, L., Marques, M.A., Kim, H., Portugal, M.I., Antunes, S.L., Menozzi, F.D., Ottenhoff, T.H., Brennan, P.J., and Pessolani, M.C., 2005, Mapping the laminin-binding and adhesive domain of the cell surface-associated Hlp/LBP protein from *Mycobacterium leprae*: *Microbes Infect*, v. 7, p. 1097-109.
- Sobetzko, P., Travers, A., and Muskhelishvili, G., 2012, Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle: *Proc Natl Acad Sci U S A*, v. 109, p. E42-50.
- Sonden, B., and Uhlin, B.E., 1996, Coordinated and differential expression of histone-like proteins in *Escherichia coli*: regulation and function of the H-NS analog StpA: *EMBO J*, v. 15, p. 4970-80.
- Sonnleitner, E., and Haas, D., 2011, Small RNAs as regulators of primary and secondary metabolism in *Pseudomonas* species: *Appl Microbiol Biotechnol*, v. 91, p. 63-79.
- Sousa, S.A., Ramos, C.G., Moreira, L.M., and Leitao, J.H., 2010, The hfq gene is required for stress resistance and full virulence of *Burkholderia cepacia* to the nematode *Caenorhabditis elegans*: *Microbiology*, v. 156, p. 896-908.
- Starai, V.J., Celic, I., Cole, R.N., Boeke, J.D., and Escalante-Semerena, J.C., 2002, Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine: *Science*, v. 298, p. 2390-2.

- Steinman, H.M., Weinstein, L., and Brenowitz, M., 1994, The manganese superoxide dismutase of *Escherichia coli* K-12 associates with DNA: *J Biol Chem*, v. 269, p. 28629-34.
- Stock, A., Schaeffer, E., Koshland, D.E., Jr., and Stock, J., 1987, A second type of protein methylation reaction in bacterial chemotaxis: *J Biol Chem*, v. 262, p. 8011-4.
- Stock, J.B., Ninfa, A.J., and Stock, A.M., 1989, Protein phosphorylation and regulation of adaptive responses in bacteria: *Microbiol Rev*, v. 53, p. 450-90.
- Storz, G., Altuvia, S., and Wassarman, K.M., 2005, An abundance of RNA regulators: *Annu Rev Biochem*, v. 74, p. 199-217.
- Struhl, K., 1999, Fundamentally different logic of gene regulation in eukaryotes and prokaryotes: *Cell*, v. 98, p. 1-4.
- Studier, F.W., and Moffatt, B.A., 1986, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes: *J Mol Biol*, v. 189, p. 113-30.
- Summers, E.L., Meindl, K., Uson, I., Mitra, A.K., Radjainia, M., Colangeli, R., Alland, D., and Arcus, V.L., 2012, The Structure of the Oligomerization Domain of Lsr2 from *Mycobacterium tuberculosis* Reveals a Mechanism for Chromosome Organization and Protection: *PLoS One*, v. 7, p. e38542.
- Sun, X., Zhulin, I., and Wartell, R.M., 2002, Predicted structure and phyletic distribution of the RNA-binding protein Hfq: *Nucleic Acids Res*, v. 30, p. 3662-71.
- Sun, Y., Jiang, X., Chen, S., and Price, B.D., 2006, Inhibition of histone acetyltransferase activity by anacardic acid sensitizes tumor cells to ionizing radiation: *FEBS Lett*, v. 580, p. 4353-6.
- Swiercz, J.P., Hindra, Bobek, J., Haiser, H.J., Di Berardo, C., Tjaden, B., and Elliot, M.A., 2008, Small non-coding RNAs in *Streptomyces coelicolor*: *Nucleic Acids Res*, v. 36, p. 7240-51.
- Swinger, K.K., and Rice, P.A., 2004, IHF and HU: flexible architects of bent DNA: *Curr Opin Struct Biol*, v. 14, p. 28-35.
- Takatsuka, M., Osada-Oka, M., Satoh, E.F., Kitadokoro, K., Nishiuchi, Y., Niki, M., Inoue, M., Iwai, K., Arakawa, T., Shimoji, Y., Ogura, H., Kobayashi, K., Rambukkana, A., and Matsumoto, S., 2011, A histone-like protein of mycobacteria possesses ferritin superfamily protein-like activity and protects against DNA damage by Fenton reaction: *PLoS One*, v. 6, p. e20985.
- Tanaka, H., Yasuzawa, K., Kohno, K., Goshima, N., Kano, Y., Saiki, T., and Imamoto, F., 1995, Role of HU proteins in forming and constraining supercoils of chromosomal DNA in *Escherichia coli*: *Mol Gen Genet*, v. 248, p. 518-26.
- Tani, T.H., Khodursky, A., Blumenthal, R.M., Brown, P.O., and Matthews, R.G., 2002, Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis: *Proc Natl Acad Sci U S A*, v. 99, p. 13471-6.
- Taylor, D.E., and Summers, A.O., 1979, Association of tellurium resistance and bacteriophage inhibition conferred by R plasmids: *J Bacteriol*, v. 137, p. 1430-3.
- Tendeng, C., Soutourina, O.A., Danchin, A., and Bertin, P.N., 2003, MvaT proteins in *Pseudomonas* spp.: a novel class of H-NS-like proteins: *Microbiology*, v. 149, p. 3047-50.
- Teplyakov, A., Obmolova, G., Bir, N., Reddy, P., Howard, A.J., and Gilliland, G.L., 2003, Crystal structure of the YajQ protein from *Haemophilus influenzae* reveals a tandem of RNP-like domains: *J Struct Funct Genomics*, v. 4, p. 1-9.
- Teramoto, J., Yoshimura, S.H., Takeyasu, K., and Ishihama, A., 2010, A novel nucleoid protein of *Escherichia coli* induced under anaerobic growth conditions: *Nucleic Acids Res*, v. 38, p. 3605-18.
- Tezuka, T., Hara, H., Ohnishi, Y., and Horinouchi, S., 2009, Identification and gene disruption of small noncoding RNAs in *Streptomyces griseus*: *J Bacteriol*, v. 191, p. 4896-904.
- Thao, S., Chen, C.S., Zhu, H., and Escalante-Semerena, J.C., 2010, Nepsilon-lysine acetylation of a bacterial transcription factor inhibits its DNA-binding activity: *PLoS One*, v. 5, p. e15123.
- Thingholm, T.E., Jorgensen, T.J., Jensen, O.N., and Larsen, M.R., 2006, Highly selective enrichment of phosphorylated peptides using titanium dioxide: *Nat Protoc*, v. 1, p. 1929-35.
- Tillotson, R.D., Wosten, H.A., Richter, M., and Willey, J.M., 1998, A surface active protein involved in aerial hyphae formation in the filamentous fungus *Schizophyllum commune* restores

- the capacity of a bald mutant of the filamentous bacterium *Streptomyces coelicolor* to erect aerial structures: *Mol Microbiol*, v. 30, p. 595-602.
- Tonthat, N.K., Arold, S.T., Pickering, B.F., Van Dyke, M.W., Liang, S., Lu, Y., Beuria, T.K., Margolin, W., and Schumacher, M.A., 2011, Molecular mechanism by which the nucleoid occlusion factor, SImA, keeps cytokinesis in check: *EMBO J*, v. 30, p. 154-64.
- Travers, A., 1997, DNA-protein interactions: IHF--the master bender: *Curr Biol*, v. 7, p. R252-4.
- Travers, A., and Muskhelishvili, G., 1998, DNA microloops and microdomains: a general mechanism for transcription activation by torsional transmission: *J Mol Biol*, v. 279, p. 1027-43.
- , 2005, DNA supercoiling - a global transcriptional regulator for enterobacterial growth?: *Nat Rev Microbiol*, v. 3, p. 157-69.
- Travers, A., Schneider, R., and Muskhelishvili, G., 2001, DNA supercoiling and transcription in *Escherichia coli*: The FIS connection: *Biochimie*, v. 83, p. 213-7.
- Tsai, H.H., Huang, C.H., Tessmer, I., Erie, D.A., and Chen, C.W., 2011, Linear *Streptomyces* plasmids form superhelical circles through interactions between their terminal proteins: *Nucleic Acids Res*, v. 39, p. 2165-74.
- Tsui, P., Helu, V., and Freundlich, M., 1988, Altered osmoregulation of ompF in integration host factor mutants of *Escherichia coli*: *J Bacteriol*, v. 170, p. 4950-3.
- Udwary, D.W., Zeigler, L., Asolkar, R.N., Singan, V., Lapidus, A., Fenical, W., Jensen, P.R., and Moore, B.S., 2007, Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*: *Proc Natl Acad Sci U S A*, v. 104, p. 10376-81.
- Umbarger, M.A., Toro, E., Wright, M.A., Porreca, G.J., Bau, D., Hong, S.H., Fero, M.J., Zhu, L.J., Marti-Renom, M.A., McAdams, H.H., Shapiro, L., Dekker, J., and Church, G.M., 2011, The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation: *Mol Cell*, v. 44, p. 252-64.
- Updegrove, T.B., Correia, J.J., Galletto, R., Bujalowski, W., and Wartell, R.M., 2010, *E. coli* DNA associated with isolated Hfq interacts with Hfq's distal surface and C-terminal domain: *Biochim Biophys Acta*, v. 1799, p. 588-96.
- Uyar, E., Kurokawa, K., Yoshimura, M., Ishikawa, S., Ogasawara, N., and Oshima, T., 2009, Differential binding profiles of StpA in wild-type and hns mutant cells: a comparative analysis of cooperative partners by chromatin immunoprecipitation-microarray analysis: *J Bacteriol*, v. 191, p. 2388-91.
- Valens, M., Penaud, S., Rossignol, M., Cornet, F., and Boccard, F., 2004, Macrodome organization of the *Escherichia coli* chromosome: *EMBO J*, v. 23, p. 4330-41.
- van Noort, J., Verbrugge, S., Goosen, N., Dekker, C., and Dame, R.T., 2004, Dual architectural roles of HU: formation of flexible hinges and rigid filaments: *Proc Natl Acad Sci U S A*, v. 101, p. 6969-74.
- Vanet, A., Plumbridge, J.A., Guerin, M.F., and Alix, J.H., 1994, Ribosomal protein methylation in *Escherichia coli*: the gene *prmA*, encoding the ribosomal protein L11 methyltransferase, is dispensable: *Mol Microbiol*, v. 14, p. 947-58.
- Viollier, P.H., Minas, W., Dale, G.E., Folcher, M., and Thompson, C.J., 2001, Role of acid metabolism in *Streptomyces coelicolor* morphological differentiation and antibiotic biosynthesis: *J Bacteriol*, v. 183, p. 3184-92.
- Vockenhuber, M.P., Sharma, C.M., Statt, M.G., Schmidt, D., Xu, Z., Dietrich, S., Liesegang, H., Mathews, D.H., and Suess, B., 2011, Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*: *RNA Biol*, v. 8, p. 468-77.
- Vockenhuber, M.P., and Suess, B., 2012, *Streptomyces coelicolor* sRNA scr5239 inhibits agarase expression by direct base pairing to the *dagA* coding region: *Microbiology*, v. 158, p. 424-35.
- Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jager, J.G., Huttenhofer, A., and Wagner, E.G., 2003, RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria: *Nucleic Acids Res*, v. 31, p. 6435-43.
- Vogel, J., and Luisi, B.F., 2011, Hfq and its constellation of RNA: *Nat Rev Microbiol*, v. 9, p. 578-89.

- von Freiesleben, U., Rasmussen, K.V., and Schaechter, M., 1994, SeqA limits DnaA activity in replication from oriC in *Escherichia coli*: *Mol Microbiol*, v. 14, p. 763-72.
- Wang, G., Huang, X., Li, S., Huang, J., Wei, X., Li, Y., and Xu, Y., 2012, The RNA chaperone Hfq regulates antibiotic biosynthesis in the rhizobacterium *Pseudomonas aeruginosa* M18: *J Bacteriol*, v. 194, p. 2443-57.
- Wang, R.F., and Kushner, S.R., 1991, Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*: *Gene*, v. 100, p. 195-9.
- Wang, W., Li, G.W., Chen, C., Xie, X.S., and Zhuang, X., 2011, Chromosome organization by a nucleoid-associated protein in live bacteria: *Science*, v. 333, p. 1445-9.
- Wang, W., Shu, D., Chen, L., Jiang, W., and Lu, Y., 2009, Cross-talk between an orphan response regulator and a noncognate histidine kinase in *Streptomyces coelicolor*: *FEMS Microbiol Lett*, v. 294, p. 150-6.
- Watve, M.G., Tickoo, R., Jog, M.M., and Bhole, B.D., 2001, How many antibiotics are produced by the genus *Streptomyces*?: *Arch Microbiol*, v. 176, p. 386-90.
- White, J., and Bibb, M., 1997, bldA dependence of undecylprodigiosin production in *Streptomyces coelicolor* A3(2) involves a pathway-specific regulatory cascade: *J Bacteriol*, v. 179, p. 627-33.
- Williams, R.B., Henrikson, J.C., Hoover, A.R., Lee, A.E., and Cichewicz, R.H., 2008, Epigenetic remodeling of the fungal secondary metabolome: *Org Biomol Chem*, v. 6, p. 1895-7.
- Wilson, R.L., Libby, S.J., Freet, A.M., Boddicker, J.D., Fahlen, T.F., and Jones, B.D., 2001, Fis, a DNA nucleoid-associated protein, is involved in *Salmonella typhimurium* SPI-1 invasion gene expression: *Mol Microbiol*, v. 39, p. 79-88.
- Wright, P.E., and Dyson, H.J., 2009, Linking folding and binding: *Curr Opin Struct Biol*, v. 19, p. 31-8.
- Yang, Y.H., Song, E., Willemse, J., Park, S.H., Kim, W.S., Kim, E.J., Lee, B.R., Kim, J.N., van Wezel, G.P., and Kim, B.G., 2012, A novel function of *Streptomyces* integration host factor (SIHF) in the control of antibiotic production and sporulation in *Streptomyces coelicolor*: *Antonie Van Leeuwenhoek*, v. 101, p. 479-92.
- Yim, G., Wang, H.H., and Davies, J., 2007, Antibiotics as signalling molecules: *Philos Trans R Soc Lond B Biol Sci*, v. 362, p. 1195-200.
- Yokoyama, E., Doi, K., Kimura, M., and Ogata, S., 2001, Disruption of the hup gene encoding a histone-like protein HS1 and detection of HS12 of *Streptomyces lividans*: *Res Microbiol*, v. 152, p. 717-23.
- Yokoyama, K., Ishijima, S.A., Clowney, L., Koike, H., Aramaki, H., Tanaka, C., Makino, K., and Suzuki, M., 2006, Feast/famine regulatory proteins (FFRPs): *Escherichia coli* Lrp, AsnC and related archaeal transcription factors: *FEMS Microbiol Rev*, v. 30, p. 89-108.
- Yoshikawa, A., Isono, S., Sheback, A., and Isono, K., 1987, Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of *Escherichia coli* K12: *Mol Gen Genet*, v. 209, p. 481-8.
- Yu, B.J., Kim, J.A., Moon, J.H., Ryu, S.E., and Pan, J.G., 2008, The diversity of lysine-acetylated proteins in *Escherichia coli*: *J Microbiol Biotechnol*, v. 18, p. 1529-36.
- Zahrt, T.C., and Deretic, V., 2000, An essential two-component signal transduction system in *Mycobacterium tuberculosis*: *J Bacteriol*, v. 182, p. 3832-8.
- Zemanova, M., Kaderabkova, P., Patek, M., Knoppova, M., Silar, R., and Nesvera, J., 2008, Chromosomally encoded small antisense RNA in *Corynebacterium glutamicum*: *FEMS Microbiol Lett*, v. 279, p. 195-201.
- Zhang, A., Altuvia, S., Tiwari, A., Argaman, L., Hengge-Aronis, R., and Storz, G., 1998, The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein: *EMBO J*, v. 17, p. 6061-8.
- Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G., and Gottesman, S., 2003, Global analysis of small RNA and mRNA targets of Hfq: *Mol Microbiol*, v. 50, p. 1111-24.
- Zhang, Q., and Wang, Y., 2010, HMG modifications and nuclear function: *Biochim Biophys Acta*, v. 1799, p. 28-36.

- Zheng, D., Constantinidou, C., Hobman, J.L., and Minchin, S.D., 2004, Identification of the CRP regulon using in vitro and in vivo transcriptional profiling: *Nucleic Acids Res*, v. 32, p. 5874-93.
- Zimmerman, S.B., 2006, Shape and compaction of *Escherichia coli* nucleoids: *J Struct Biol*, v. 156, p. 255-61.