



**A University of Sussex DPhil thesis**

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

**Structural and Functional Studies on RbpA, a RNA Polymerase  
Binding Protein in *Streptomyces coelicolor* A3 (2)**

**By**

**ALINE TABIB-SALAZAR**

Submitted in total fulfilment of the requirements of the Degree of  
**Doctor of Philosophy**

**Department of Biochemistry  
School of Life Sciences  
University of Sussex  
September 2012**

UNIVERSITY OF SUSSEX

ALINE TABIB-SALAZAR

For the degree of Doctor of Philosophy

Structural and Functional Studies on RbpA, a RNA Polymerase Binding Proteinin *Streptomyces coelicolor* A3 (2)SUMMARY

RbpA is a RNA polymerase-binding protein that was identified in *Streptomyces coelicolor*. It is found in all Actinobacteria, including the pathogenic agent *Mycobacterium tuberculosis*. *Streptomyces* strains that have an *rbpA* mutation grow at a slower rate than the wild-type and are more sensitive to the RNAP-targeting antibiotic, rifampicin. RbpA binds to and activates  $\sigma^{\text{HrdB}}$ , the principal sigma factor that directs transcription of most housekeeping genes in *S. coelicolor*. Using bacterial two-hybrid analysis and *in vitro* pull down assays, RbpA was shown to interact with region 1.2-2.4 of  $\sigma^{\text{HrdB}}$ . This region forms part of a major interface with core RNA polymerase and is involved in the recognition of, and binding to, the -10 promoter element. Rv2050, the homologue of RbpA in *M. tuberculosis*, was also shown to interact with the principal sigma factor of this organism,  $\sigma^A$ . Structural studies on RbpA and Rv2050 revealed that it is composed of two regions, a structured N-terminal  $\beta$ -fold region and an flexible or unstable C-terminal region, which interacts with sigma. Alanine-scanning site-directed mutagenesis on the C-terminal region of RbpA identified important residues involved in  $\sigma^{\text{HrdB}}$  interaction as well as residues that might be involved in transcriptional activation.

## **Acknowledgements**

I would like to acknowledge the University of Sussex and the Overseas Research Studentship for awarding me a scholarship. I would like to thank my supervisor, Dr. Mark Paget, for all his support, guidance and patience throughout my PhD degree. I would like to express thanks to everyone in the Paget lab, past and present, especially: Richard, Philip, Claire and Laurence for all their support. I would like to acknowledge Professor Stephen Matthews together with his lab group at Imperial College London for their collaboration with all the structural work. Thanks to Kyle for his help with the structural figures.

Special thanks to my parents for all their emotional and financial support without which none of this would have been possible especially my mum for all her kind words, and her faith in me. I would also like to extend special thanks to my husband, Juan, for his patience, providing me with constant motivation and assurance throughout my degree. Also, special thanks to my amazing friend, Eva, for keeping me sane and being there for me at all times and to my brother, Elias, who was patient with me even when I became unbearable during the final stages of my writing. Finally, I would like to express appreciation to all my family especially Shadia, John, Charlene, William, Jimena, and my grandmas Najet and Rene. Also thanks to my friends; Jatu, Terri, Jasmine, Pooja, Prash, Maria, Yari, Ela, Raymond, Nick and Caroline who helped in all their little ways.

## **Table of Contents:**

<b>1- Introduction.....</b>	<b>17</b>
1.0 Overview .....	18
1.1 The Actinobacteria phylum.....	18
1.1.1 <i>Streptomyces coelicolor</i> A3 (2).....	19
1.1.1.1 The genome .....	19
1.1.1.2 Life cycle and regulation.....	20
1.1.1.3 Regulation of antibiotic production .....	23
1.1.2 <i>Mycobacterium tuberculosis</i> .....	25
1.1.2.1 The genome .....	25
1.1.2.2 Pathogenicity.....	26
1.1.2.3 Therapeutics and importance of Rifampicin .....	27
1.2 Bacterial gene regulation.....	29
1.2.1 Bacterial DNA-dependent RNA polymerase .....	30
1.2.2 Sigma factors.....	32
1.2.2.1 The different groups of the $\sigma^{70}$ -related family .....	33
1.2.2.2 The domains of the principal sigma factors .....	35
1.2.3 The sigma factors of Actinobacteria.....	39
1.2.3.1 <i>S. coelicolor</i> .....	39
1.2.3.2 <i>M. tuberculosis</i> .....	40
1.2.4 Transcription in bacteria.....	41
1.2.4.1 Initiation .....	41
1.2.4.2 Elongation .....	46
1.2.4.3 Termination .....	47
1.2.5 Factors that affect RNAP activity .....	48
1.2.5.1 Guanosine 5'-diphosphate 3'-diphosphate (ppGpp).....	48
1.2.5.2 6S RNA .....	49
1.2.6 RNAP binding proteins.....	51
1.2.6.1 DksA.....	51
1.2.6.2 CarD.....	53
1.2.7 Anti-sigma factors .....	53
1.2.7.1 Rsd.....	54

1.2.7.2 RsrA .....	55
1.2.7.3 Gin.....	57
1.2.8 Proteins that bind to anti-sigma factors.....	58
1.2.9 Positive regulators of sigma factors .....	59
1.2.9.1 Crl.....	60
1.2.9.2 RbpA .....	61
1.3 Project aims.....	62
<b>2- Materials and Methods .....</b>	<b>64</b>
2.1 Chemicals and reagents.....	65
2.1.1 Enzymes .....	66
2.1.1.1 Polymerases .....	66
2.1.1.2 DNA modifying enzymes .....	66
2.1.1.3 DNA/RNA restriction enzymes .....	66
2.1.2 Plasmids and expression vectors .....	67
2.1.3 <i>S. coelicolor</i> and <i>E. coli</i> strains .....	68
2.1.3.1 <i>E.coli</i> .....	68
2.1.3.2 <i>S. coelicolor</i> .....	69
2.1.4 Growth media and selection .....	70
2.1.4.1 <i>E.coli</i> growth media .....	70
2.1.4.2 <i>S. coelicolor</i> growth media .....	71
2.1.4.3 Antibiotics and additives .....	72
2.1.4.4 Growth and storage of <i>E. coli</i> and <i>S. coelicolor</i> strains .....	73
2.1.5 Primers .....	73
2.1.5.1 Bacterial two-hybrid.....	73
2.1.5.2 Site-directed mutagenesis.....	75
2.1.5.3 Protein expression.....	76
2.1.5.4 <i>In vitro</i> transcription .....	76
2.1.6 Solutions/buffers .....	77
2.1.6.1 DNA manipulation .....	77
2.1.6.2 Protein purification.....	78
2.1.6.3 Western blotting .....	79
2.1.6.4 <i>In vitro</i> transcription assays.....	80

2.1.6.5 $\beta$ -galactosidase assay .....	80
2.1.6.6 Others .....	80
2.2 DNA manipulation and cloning .....	81
2.2.1 DNA manipulation .....	81
2.2.1.1 General Polymerase chain reaction (PCR) .....	81
2.2.1.2 Inverse PCR .....	81
2.2.1.3 Restriction digestion .....	82
2.2.1.4 DNA gel purification.....	82
2.2.1.5 Dephosphorylation of DNA.....	82
2.2.1.6 DNA ligation .....	83
2.2.2 <i>E. coli</i> DNA transformation.....	83
2.2.2.1 Heat shock .....	83
2.2.2.2 Electroporation .....	84
2.2.3 DNA extraction and purification .....	85
2.2.3.1 Small scale Wizard® Miniprep plasmid purification.....	85
2.2.3.2 Small scale plasmid purification by alkaline lysis .....	85
2.2.3.3 Large scale Midiprep plasmid purification (Qiagen).....	85
2.2.4 RNA extraction and purification .....	85
2.2.4.1 Small scale RNA miniprep.....	85
2.2.5 Microbiological methods .....	86
2.2.5.1 Conjugation of DNA from <i>E. coli</i> to <i>S. coelicolor</i> .....	86
2.2.5.2 Harvesting <i>S. coelicolor</i> spores.....	87
2.2.5.3 Determining the growth curve of <i>S. coelicolor</i> in liquid media ....	87
2.3 Protein expression and purification .....	87
2.3.1 Protein expression .....	87
2.3.1.1 Unlabelled proteins.....	87
2.3.1.2 $^{13}\text{C}$ $^{15}\text{N}$ labelling of protein.....	88
2.3.2 Protein purification .....	88
2.3.2.1 Preparing cell lysate .....	88
2.3.2.2 Ni-NTA sepharose hand-made column .....	89
2.3.2.3 Ni-NTA spin column .....	89
2.3.2.4 Dynabeads® His-tag isolation and pull-down.....	89
2.3.2.5 Anion-exchange chromatography.....	90

2.3.2.6 Gel filtration chromatography .....	90
2.3.3 Protein buffer exchange .....	91
2.3.4 Concentrating the protein.....	91
2.3.5 Determining the protein concentration .....	91
2.3.5.1 Bradford assay .....	91
2.3.5.2 Bicinchoninic acid assay (BCA).....	92
2.3.6 Protein sample analysis by SDS polyacrylamide gel .....	92
2.3.6.1 NuPAGE® 4-12% Bis-Tris gels .....	92
2.3.6.2 SDS polyacrylamide gel preparation .....	92
2.3.7 Hanging drop vapour diffusion .....	93
2.3.8 Protein preparation for ICP-MS.....	94
2.4 Western blotting .....	94
2.4.1 Preparing samples .....	94
2.4.2 Determining protein concentration .....	94
2.4.3 Immunoblotting protocol.....	95
2.4.4 Washes and antibodies.....	95
2.4.5 ECL detection and developing .....	95
2.5 <i>In vitro</i> transcription assays .....	96
2.5.1 Preparing the gel .....	96
2.5.2 <i>In vitro</i> transcription assays .....	96
2.6 $\beta$ -galactosidase assay .....	97
<b>3- RbpA Bioinformatics and its Biological Role.....</b>	<b>98</b>
3.0 Overview .....	99
3.1 Bioinformatic analysis of RbpA.....	99
3.1.1 RbpA in the Actinobacteria .....	99
3.1.2 The RbpA paralogue, RbpB.....	103
3.1.3 The secondary structural predictions of RbpA .....	106
3.1.3.1 The N-terminal region.....	106
3.1.3.2 The C-terminal region.....	107
3.2 Biological role of RbpA in <i>S. coelicolor</i> .....	108
3.2.1 Essential for normal growth .....	108
3.2.2 Gene expression in <i>S. coelicolor</i> $\Delta rbpA$ strain .....	112



3.2.2.1 Down-regulated genes .....	113
3.2.2.2 Up-regulated genes.....	115
3.2.3 The level of $\sigma^{\text{HrdB}}$ during growth of <i>S. coelicolor</i> .....	117
3.2.3.1 J1981 .....	117
3.2.3.2 S129.....	120
3.2.4 RbpA in holoenzyme formation.....	122
3.3 Discussion .....	125
3.3.1 Actinobacteria lacking RbpA homologues.....	125
3.3.2 RbpB in <i>S. coelicolor</i> .....	126
3.3.3 Gene expression in <i>S. coelicolor</i> $\Delta\text{rbpA}$ mutant.....	126
3.3.4 $\sigma^{\text{HrdB}}$ level during growth of J1981 and S129 .....	127
3.3.5 RbpA in holoenzyme formation.....	129
<b>4- Mapping RbpA interactions with <math>\sigma^{\text{HrdB}}</math> .....</b>	<b>130</b>
4.0 Overview .....	131
4.1 BACTH analysis of sigma factors with RbpA.....	131
4.1.1 RbpA interaction with $\sigma^{\text{HrdB}}$ .....	132
4.1.2 RbpA interaction with Group II .....	135
4.1.3 RbpA interaction with Group III and IV.....	137
4.2 Mapping the binding site of RbpA on $\sigma^{\text{HrdB}}$ .....	140
4.2.1 <i>In vivo</i> analysis .....	142
4.2.2 <i>In vitro</i> analysis .....	145
4.3 Rv2050 interaction with <i>M. tuberculosis</i> sigma factors.....	147
4.3.1 $\sigma^{\text{A}}$ and $\sigma^{\text{B}}$ .....	147
4.3.2 Rv2050 interaction with sigma domain 2 .....	150
4.3.2.1 <i>In vivo</i> analysis.....	150
4.3.2.2 <i>In vitro</i> analysis.....	151
4.4 Discussion .....	152
4.4.1 RbpA binds to $\sigma^{\text{HrdA}}$ and $\sigma^{\text{HrdB}}$ .....	152
4.4.2 The principal sigma factor, $\sigma^{\text{HrdB}}$ .....	155
4.4.3 RbpB interaction with sigma factors.....	156
4.4.4 RbpA interacts at domain 2 of $\sigma^{\text{HrdB}}$ .....	156
4.4.5 $\sigma^{\text{HrdB}}$ lacks a non-conserved region between region 1.2 to 2.1 .....	157

4.4.6 Crl, a possible analogue of RbpA .....	159
4.4.7 The importance of Rv2050 in <i>M. tuberculosis</i> .....	160
<b>5- Mutagenesis of <i>rbpA</i> .....</b>	<b>163</b>
5.0 Overview .....	164
5.1 Identifying the $\sigma^{\text{HrdB}}$ binding region of RbpA .....	164
5.1.1 $\sigma^{\text{HrdB}}$ interaction with RbpA fragments.....	165
5.2 Site-directed mutagenesis of C-terminal region of RbpA.....	168
5.2.1 Constructing <i>rbpA</i> mutants .....	168
5.2.2 Helix 1 .....	170
5.2.2.1 Effect of mutations in RbpA helix 1 on the phenotype of <i>S. coelicolor</i> .....	171
5.2.2.2 Effect of mutations in RbpA helix 1 on the growth curve of <i>S. coelicolor</i> .....	173
5.2.3 The ERR motif .....	174
5.2.3.1 Effect of mutations in RbpA ERR motif on the phenotype of <i>S. coelicolor</i> .....	175
5.2.3.2 Effect of mutations in RbpA ERR motif on the growth curve of <i>S. coelicolor</i> .....	176
5.2.4 Helix 2.....	177
5.2.4.1 Effect of mutations in RbpA helix 2 on the phenotype of <i>S. coelicolor</i> .....	177
5.3 BACTH analysis of $\sigma^{\text{HrdB}}$ interaction with RbpA mutants .....	179
5.3.1 Helix 1 .....	179
5.3.2 The ERR motif .....	181
5.3.3 Helix 2.....	182
5.4 $\sigma^{\text{HrdB}}$ interaction with selected RbpA mutants <i>in vitro</i> .....	185
5.4.1 RbpA <sub>R80A</sub> and RbpA <sub>M85A</sub> .....	185
5.4.2 RbpA <sub>R89A</sub> and RbpA <sub>R90A</sub> .....	188
5.5 Testing RbpA <sub>R80A</sub> and RbpA <sub>M85A</sub> for dominant-negative effects.....	189
5.6 Transcription activation by RbpA <sub>M85A</sub> and RbpA <sub>2RA</sub> proteins.....	192
5.7 Discussion .....	194
5.7.1 The C-terminal region of RbpA .....	194
5.7.2 R80 and M85 are important in RbpA functionality.....	194

5.7.3 The ERR motif of RbpA is essential for $\sigma^{\text{HrdB}}$ interaction.....	195
5.7.4 Helix 2 of RbpA is involved in interaction with $\sigma^{\text{HrdB}}$ .....	196
<b>6- Structural analysis of Rv2050/RbpA .....</b>	<b>198</b>
6.0 Overview .....	199
6.1 X-ray crystallography of Rv2050 .....	199
6.2 NMR spectroscopy of Rv2050 .....	200
6.2.1 Native $^{13}\text{C}$ & $^{15}\text{N}$ Rv2050 over-expression .....	200
6.2.2 Purification of $^{13}\text{C}$ $^{15}\text{N}$ Rv2050 .....	201
6.2.2.1 Anion-exchange - QFF column .....	201
6.2.2.2 Anion-exchange - Mono Q column .....	201
6.2.2.3 Gel filtration .....	201
6.2.3 NMR structure of Rv2050 .....	203
6.2.4 Homology search of Rv2050.....	205
6.3 Rv2050 is not a Zn metalloprotein.....	206
6.3.1 ICP-MS analysis of Rv2050 .....	206
6.3.2 Key residues that maintain the structural integrity of Rv2050 .....	207
6.4 Attempts to solve the structure of RbpA using NMR .....	208
6.4.1 Over-expression of $^{13}\text{C}$ $^{15}\text{N}$ RbpA .....	209
6.4.2 Purification of $^{13}\text{C}$ $^{15}\text{N}$ RbpA.....	209
6.4.2.1 Anion-exchange - QFF column .....	209
6.4.2.2 Anion-exchange - Mono Q column .....	209
6.4.2.3 Gel filtration .....	209
6.4.3 The C-terminal region .....	212
6.4.4 NMR analysis.....	212
6.5 Attempts to solve the structure of an RbpA/Rv2050-sigma complex.....	213
6.5.1 Isolation of Rv2050- $\sigma^{\text{A}(2)}$ .....	213
6.5.2 Isolation of RbpA- $\sigma^{\text{HrdB}(2)}$ .....	215
6.6 Discussion .....	217
6.6.1 Protein homology of the N-terminal region of Rv2050 .....	217
6.6.2 Zn metal as a structural component of RbpA.....	218
6.6.3 Protein homologues that exist in Zn and non-Zn form .....	219
6.6.3.1 Ribosomal proteins.....	219

6.6.3.2 DksA.....	220
6.6.3.3 Superoxide dismutase (SodC) .....	220
<b>7- General Discussions and Summary.....</b>	<b>222</b>
7.0 Overview .....	223
7.1 The RbpA family.....	223
7.1.1 Why is RbpA only present within the Actinobacteria? .....	224
7.1.2 Is the RbpA family essential in the Actinobacteria? .....	225
7.2 RbpA as a distinct transcriptional activator.....	225
7.3 The role of RbpA in transcription initiation .....	227
7.4 RbpA structure and region function .....	229
7.5 Future directions.....	230
<b>Appendices.....</b>	<b>232</b>
<b>References.....</b>	<b>249</b>

## **A List of the Figures:**

<b>1- Introduction.....</b>	<b>17</b>
Figure 1.1: The life cycle of <i>Streptomyces coelicolor</i> .....	22
Figure 1.2: The overall structure of the Rifamycins.....	28
Figure 1.3: The crystal structure of <i>Thermus thermophilus</i> RNAP holoenzyme at 2.6 Å resolution.....	32
Figure 1.4: Schematic diagram of the $\sigma^{70}$ principal sigma factor.....	34
Figure 1.5: A ribbon diagram of the principal sigma factor, $\sigma^{70}$ .....	36
Figure 1.6: A cross-section view of RNAP during transcription initiation.....	43
Figure 1.7: Structure of $\sigma^{70}$ region 4 in complex with anti- $\sigma$ factor, Rsd.....	55
Figure 1.8: Structure of $\sigma^E$ /RseA complex. ....	59
<b>3- RbpA Bioinformatics and its Biological Role.....</b>	<b>98</b>
Figure 3.1: Multiple sequence alignment of 22 homologues of RbpA in the Actinobacteria phylum.....	102
Figure 3.2: Multiple sequence alignment of 22 homologues of RbpB in the Actinobacteria phylum.....	105
Figure 3.3: A helical wheel representation of Helix 2 showing the amino acid distribution in the helix. ....	108
Figure 3.4: RbpA is essential to maintain the normal growth rate of <i>S. coelicolor</i> . ....	111
Figure 3.5: The concentration of $\sigma^{\text{HrdB}}$ during growth of J1981. ....	119
Figure 3.6: The concentration of $\sigma^{\text{HrdB}}$ during growth of S129.....	121
Figure 3.7: The level of $\sigma^{\text{HrdB}}$ relative to RNAP in J1981 and S129. ....	124
<b>4- Mapping RbpA interactions with <math>\sigma^{\text{HrdB}}</math> .....</b>	<b>130</b>
Figure 4.1: BACTH assay of RbpA interaction with principal sigma factor, $\sigma^{\text{HrdB}(2-4)}$ , of <i>S. coelicolor</i> . ....	134
Figure 4.2: BACTH assay of RbpA interaction with domain 2-4 of Group I and II sigma factors, $\sigma^{\text{HrdA}}$ , $\sigma^{\text{HrdB}}$ , $\sigma^{\text{HrdC}}$ , $\sigma^{\text{HrdD}}$ , of <i>S. coelicolor</i> .....	136
Figure 4.3: BACTH assay of RbpA interaction with Group III and Group IV sigma factors: $\sigma^{\text{B}}$ , $\sigma^{\text{R}}$ , $\sigma^{\text{E}}$ and $\sigma^{\text{WhiG}}$ , of <i>S. coelicolor</i> . ....	139
Figure 4.4: Alignment of <i>T. aquaticus</i> $\sigma^{\text{A}}$ and $\sigma^{\text{HrdB}}$ . ....	141
Figure 4.5: Constructed structural domains of $\sigma^{\text{HrdB}}$ .....	142
Figure 4.6: BACTH assay of RbpA interaction with the constructed fragments of $\sigma^{\text{HrdB}}$ . ....	144

Figure 4.7: The interaction of RbpA with $\sigma^{\text{HrdB}(2)}$ and $\sigma^{\text{HrdB}(2-4)}$ .....	146
Figure 4.8: BACTH assay of Rv2050 interaction with principal sigma factor, $\sigma^{\text{A}}$ , and principal-like sigma factor, $\sigma^{\text{B}}$ , of <i>M. tuberculosis</i> .....	149
Figure 4.9: The $\beta$ -galactosidase activity of the interaction of Rv2050 with $\sigma^{\text{A}(2)}$ and $\sigma^{\text{HrdB}(2)}$ .....	151
Figure 4.10: The interaction of Rv2050 with $\sigma^{\text{HrdB}(2)}$ .....	152
Figure 4.11: The multiple sequence alignment of domains 2-4 of $\sigma^{\text{HrdA}}$ , $\sigma^{\text{HrdB}}$ , $\sigma^{\text{HrdC}}$ and $\sigma^{\text{HrdD}}$ .....	154
Figure 4.12: The pairwise sequence alignment of $\sigma^{\text{HrdB}(2)}$ and <i>T. thermophilus</i> principal sigma factor, $\sigma^{\text{A}}$ .....	158
Figure 4.13: Secondary structure prediction of Crl.....	160
Figure 4.14: The multiple sequence alignment of $\sigma^{\text{HrdB}(2)}$ against $\sigma^{\text{A}}$ and $\sigma^{\text{B}}$ of <i>M. tuberculosis</i> .....	162
<b>5- Mutagenesis of <i>rbpA</i> .....</b>	<b>163</b>
Figure 5.1: Truncated fragments of RbpA.....	164
Figure 5.2: BACTH assay of $\sigma^{\text{HrdB}}$ interaction with truncated RbpA derivatives. ....	167
Figure 5.3: The amino acid sequence and secondary structure prediction of RbpA.....	168
Figure 5.4: Complementation of <i>S. coelicolor</i> S129 strain by introducing the engineered <i>rbpA_NdeI</i> vector. ....	170
Figure 5.5: Complementation of <i>S. coelicolor</i> S129 strains by introducing mutated residues of Helix 1 of RbpA.....	172
Figure 5.6: A graphical representation of the growth curve of J1981 (w/t), S129 ( $\Delta\text{rbpA}$ ), S129 (pSET $\Omega$ :: <i>rbpA</i> <sub>R80A</sub> ) and S129 (pSET $\Omega$ :: <i>rbpA</i> <sub>M85A</sub> ).....	174
Figure 5.7: Complementation of <i>S. coelicolor</i> S129 strains by introducing mutated residues of the ERR motif of RbpA. ....	175
Figure 5.8: A graphical representation of the growth curve of J1981 (w/t), S129 ( $\Delta\text{rbpA}$ ), S129 (pSET $\Omega$ :: <i>rbpA</i> <sub>R89A</sub> ) and S129 (pSET $\Omega$ :: <i>rbpA</i> <sub>R90A</sub> ).....	176
Figure 5.9: Complementation of <i>S. coelicolor</i> S129 strains by introducing mutated residues of Helix 2 of RbpA.....	178
Figure 5.10: BACTH assay of the interaction of $\sigma^{\text{HrdB}}$ with the mutated residues in Helix 1 of RbpA. ....	180
Figure 5.11: BACTH assay of the interaction of $\sigma^{\text{HrdB}}$ with the mutated residues of the ERR motif of RbpA.....	182

Figure 5.12: BACTH assay of the interaction of $\sigma^{\text{HrdB}}$ with mutated RbpA residues in Helix 2.....	184
Figure 5.13: The interaction of RbpA <sub>R80A</sub> and RbpA <sub>M85A</sub> with $\sigma^{\text{HrdB}(2)}$ and $\sigma^{\text{HrdB}(2-4)}$ . .....	187
Figure 5.14: The interaction of RbpA <sub>R89A</sub> , RbpA <sub>R90A</sub> and RbpA <sub>2RA</sub> with $\sigma^{\text{HrdB}(2-4)}$ . .....	189
Figure 5.15: Introducing pIJ6902:: <i>rbpA</i> <sub>w/t</sub> , pIJ6902:: <i>rbpA</i> <sub>R80A</sub> , and pIJ6902:: <i>rbpA</i> <sub>M85A</sub> into <i>S. coelicolor</i> S129, J1981 and M145 strains.....	191
Figure 5.16: <i>In vitro</i> transcription assays showing transcription activation in the presence of RbpA <sub>w/t</sub> , RbpA <sub>M85A</sub> , and RbpA <sub>2RA</sub> . ....	193
<b>6- Structural analysis of Rv2050/RbpA .....</b>	<b>198</b>
Figure 6.1: The purification of $^{13}\text{C}$ $^{15}\text{N}$ Rv2050.....	202
Figure 6.2: The amino acid sequence of Rv2050 indicating its structured region. .....	203
Figure 6.3: Backbone structure of N-terminus of Rv2050. ....	204
Figure 6.4: The surface charge and the cartoon representation of the N-terminus of Rv2050. ....	205
Figure 6.5: Residues in Rv2050 that are thought to be interacting instead of coordinating Zn.....	208
Figure 6.6: The purification of $^{13}\text{C}$ $^{15}\text{N}$ RbpA.....	211
Figure 6.7: The amino acid sequence of RbpA indicating its structured region. .....	212
Figure 6.8: The expression, solubility and interaction of Rv2050 and $\sigma^{\text{A}(2)}$ . ....	215
Figure 6.9: Gel filtration purification of $\sigma^{\text{HrdB}(2)}$ . ....	216
Figure 6.10: The X-ray crystal structure of the C-terminal domain of the peptide chain release factor 1 (RF1) dimer from <i>Methanosarcina mazei</i> overlayed with Rv2050 N-terminal region. ....	218

**Abbreviations:**

ACT: Actinorhodin

*aac(3)/IV*: apramycin resistance gene

*aadA*: Spectinomycin resistance gene

Amp<sup>R</sup>: Ampicillin resistance

BACTH: Bacterial adenylate cyclase two-hybrid

BCA: Bicinchoninic acid assay

ECF: Extra-Cytoplasmic function

ETZ: ET12567 cells containing pUZ8002 plasmid

HtH: Helix-turn-helix

ICP-MS: Inductively coupled plasma mass spectrometry

IPTG: Isopropyl- $\beta$ -D-thiogalactopyranoside

LA: Lennox agar

LB: Lennox broth

MCS: Multiple cloning site

MS: Mannitol soya flour medium

Mg: Magnesium

Mn: Manganese

NMR: Nuclear magnetic resonance

NCR: Non-conserved region

PMSF: Phenylmethylsulfonyl Fluoride

ppGpp: Guanosine 5'-diphosphate 3'-diphosphate

Rif: Rifampicin

RNAP: RNA polymerase

RT: Room temperature

*tsr*: Thiostepeton resistance gene

X-gal: 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside

YEME: Yeast extract-malt extract medium

ZAS: Zinc-binding anti-sigma factor

Zn: Zinc



## **1- Introduction**

## **1.0 Overview**

*Streptomyces coelicolor*, a Gram-positive bacterium, is the most studied representative of the *Streptomyces* genus. Streptomycetes are soil-living microorganisms that belong to the large phylum of Actinobacteria and this includes pathogenic agents such as *Mycobacterium tuberculosis* and *Mycobacterium leprae*. There are many cases where fundamental biological mechanisms are shared between all Actinobacteria and so discoveries made in the non-pathogenic streptomycetes can often have important implications for the understanding of pathogenic strains. The streptomycetes are medically important in their own right because they produce the majority of antibiotics in use today and are a potential source of novel antimicrobials. This project focuses on RNA polymerase (RNAP), the transcription machinery. Basic RNAP function is conserved in all bacteria and is a target for antibiotics, including rifamycins. However, recent discoveries suggest that the Actinobacteria possess some unique RNAP regulatory features that might be novel targets for antimicrobials.

## **1.1 The Actinobacteria phylum**

The Actinobacteria formerly known as the “*Actinomycetes*” are a large group of microorganisms that were discovered in the late 19<sup>th</sup> century. However, the name *Actinomycete* was first used in the late 18<sup>th</sup> century to describe a disease that affected cattle, “lumpy jaw”. Carl Otto Harz, a German botanist, saw that the structure of this new organism was similar to fungus and grew in the absence of oxygen. Therefore, he concluded that it was a fungus and he called it “*Actinomyces bovis*”. *Actinomyces* means “ray fungus” (Hopwood, 2007).

Selman Abraham Waksman discovered that these microbes produce antimicrobial agents and put forward the name ‘antibiotic’ (Kresge *et al.*, 2004). An antibiotic is a natural compound produced by one bacterium that inhibits the growth of another bacterium. The first antibiotic discovered by Waksman was Actinomycin, from *Actinomyces antibioticus*. Actinomycin has two components:

A and B. Actinomycin A has an orange-red colour and has bacteriostatic and bactericidal characteristics, whilst Actinomycin B is colourless and has only bactericidal properties (Kresge *et al.*, 2004). However, this antibiotic was very toxic and hence, was not useful at that time (Hopwood, 2007). Nowadays, it is used as a therapeutic agent in chemotherapy.

However, Waksman then discovered Streptomycin in 1943 from *Streptomyces griseus*, which was the first effective antibiotic for the treatment of tuberculosis. He was awarded the Nobel Prize for Physiology or Medicine in 1952. Interest in these microorganisms thereafter greatly increased (Kresge *et al.*, 2004) and currently these bacteria make over two-thirds of the world's antibiotics that we use for treatment of bacterial infections and fungal diseases, as well as several key anticancer and anti-parasitic therapies (Bentley *et al.*, 2002).

#### **1.1.1 *Streptomyces coelicolor* A3 (2)**

*Streptomyces* are ubiquitous soil bacteria that produce antibiotics as a result of complex secondary metabolic pathways. Antibiotics are produced to aid them in competition with other soil bacteria for basic nutrients and also for space so they can multiply freely. Streptomycetes are also involved in carbon and nitrogen recycling as they degrade insoluble remains (e.g. chitin) of organisms in the soil (Bentley *et al.*, 2002). *Streptomyces coelicolor* A3 (2) is the model organism used in research to study and understand the streptomycetes. It is an obligate aerobe but has been shown to stay viable and undergo microaerobic growth during periods of anaerobiosis (van Keulen *et al.*, 2007).

##### 1.1.1.1 The genome

*S. coelicolor* A3 (2) has a single linear chromosome with a high guanine and cytosine (G+C) content of 72.1%. DNA replication proceeds bi-directionally from the origin of replication (*oriC*) that is positioned centrally. The essential genes including genes coding for DNA replication, transcription, translation are located close to the *oriC* between 1.5 to 6.4 Mb (known as the 'core' region). However,

non-essential genes such as those coding for secondary metabolites enzymes, hydrolytic enzymes, are located on the 'arms' of the chromosome (Bentley *et al.*, 2002).

The chromosome of *S. coelicolor* has 8,667,507 bp and 7,825 predicted genes (Bentley *et al.*, 2002). It is larger and encodes about two-fold more genes than that of *M. tuberculosis*. The 'core' region of *S. coelicolor* and the whole chromosome of *M. tuberculosis* have evolved from a common ancestor, and so they have a large number of similar gene sequences and gene clusters. The most highly conserved gene clusters code for respiratory chain NADH dehydrogenase subunits. Others include genes coding for origin of replication, pyrimidine synthesis, cell division, arginine biosynthesis and ribosomal proteins. However, *S. coelicolor* has gained extra DNA at the 'arm' regions (Bentley *et al.*, 2002).

The *S. coelicolor* genome has expanded protein families involved in regulatory functions, transport, and the utilisation of extracellular nutrients. With regard to gene regulation, *S. coelicolor* has 965 regulators, of which 65 are sigma factors, and has many genes that encode sensor kinases and response regulators involved in two-component regulatory systems. In transport, it encodes 614 proteins, most of which are ABC transporter type, including permeases and ATP-binding proteins involved in transporting sugars, amino acids, peptides, metals and ions. It also encodes 819 proteins involved in degradation of nutrients, an example of which includes the secreted hydrolases that are comprised of 60 proteases/peptidase, thirteen chitinases/chitosanases, eight cellulases/endogluconases, three amylases and two pectate lyases (Bentley *et al.*, 2002).

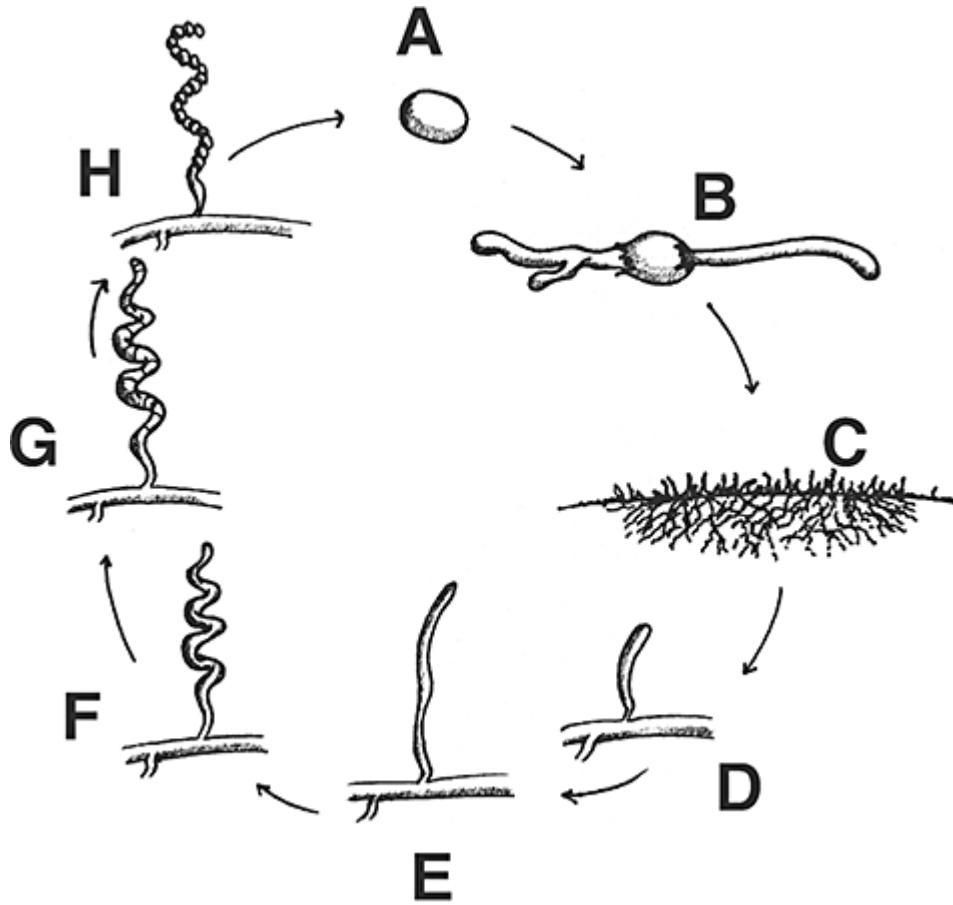
#### 1.1.1.2 Life cycle and regulation

*S. coelicolor* and other streptomycetes are distinguished from other bacteria by their complex multi-cellular development. They form a highly branched

vegetative mycelium growing horizontally on a surface. The mycelium then differentiates into aerial hyphae that project up from the mycelium giving rise to spores (Bentley et al., 2002, Hopwood, 2007).

When a spore of *S. coelicolor* lands on a suitable environment with nutrients, it starts germination by forming one or two “germ tubes” (Fig. 1.1A & B) (Flardh & Buttner, 2009). These germ tubes grow to produce a vegetative branching mass of mycelium (Fig. 1.1C). The growth of *S. coelicolor* is distinct from the other Actinobacteria in that they extend at the tip. The protein, DivIVA is crucial for tip extension and cell shape, recruiting cell wall biosynthesis machinery during branching (Flardh & Buttner, 2009). During nutrient limitation, aerial hyphae projects from the vegetative mycelium and grows into the air (Fig. 1.1D). This requires the aerial hyphae to be coated in a hydrophobic sheath and to produce SapB, a surfactant peptide. SapB, however, is only activated on rich media and reduces the surface tension allowing the aerial hyphae to protrude into the air. The hydrophobic sheath is made up of the chaplin and rodlin proteins. The chaplins assemble at the surface, lowering the surface tension, whilst the rodlin proteins organise the chaplins into well-ordered structures (Flardh & Buttner, 2009).

The aerial hyphae continue growing by tip extension, carrying DivIVA foci, and can form hyphal cross-walls. This forms a sporogenic cell, which has no septation and contains more than 50 copies of the genome (Fig. 1.1E). Growth is arrested before multiple cell divisions are initiated. Septation is directed by FtsZ, which is recruited by SsgB (van Wezel & McDowall, 2011). The segregation and arrangement of the chromosomes require ParAB and FtsK systems during sporulation. This forms the prespore compartments each of which contains a single chromosome (Fig. 1.1G). MreB then ensures accurate assembly of the spore wall (Fig. 1.1H) (Flardh & Buttner, 2009).



**Figure 1.1: The life cycle of *Streptomyces coelicolor*.**

**A:** Spore, **B:** spore germination and formation of germ-tube, **C:** vegetative growth, **D to F:** aerial growth, **G:** aerial hyphae with septation and uni-genomic pre-spore compartments, and **H:** spore maturation. Copyright license No.: 2937570798152 (Newton & Fahey, 2008).

Regulation of these different stages of the life cycle is controlled to some extent by sigma factors belonging to both groups III and IV (Section 1.2.2.1).  $\sigma^{\text{BldN}}$  and  $\sigma^{\text{N}}$  are involved in the formation of aerial hyphae (Bibb *et al.*, 2000, Dalton *et al.*, 2007) whereas WhiG, the sporulation sigma factor, commences the formation of spores (Chater *et al.*, 1989). The *bld* mutants lack the ability to form aerial mycelium whilst the *whi* mutants do not produce normal mature spores or the grey pigment (Chater, 2001). BldD is a repressor of *bldN* and *whiG*. When aerial

hyphae growth starts, BldD no longer represses *bldN* and hence, RNAP- $\sigma^{\text{BldN}}$  transcribes BldM, which is required for aerial growth (Chater, 2001).

The early sporulation genes include: *whi* A, B, G, H, I and J. WhiA, WhiB and possibly WhiI are involved in initiating septum formation (Flardh & Buttner, 2009), whilst WhiD is involved in later stages of sporulation. Both WhiB and WhiD contain iron-sulphur cluster, co-ordinated by four conserved cysteine residues, which possibly responds to redox stress causing the dissociation of the cluster (Chater, 2001, Crack *et al.*, 2009).  $\sigma^{\text{F}}$  is involved in later stages of sporulation and is only expressed when sporulation has initiated and therefore is reliant on WhiG and other early sporulation genes (Kelemen *et al.*, 1996). Another late sporulation gene is *whiE*, which produces the grey polyketide spore pigment (Flardh & Buttner, 2009).

#### 1.1.1.3 Regulation of antibiotic production

*S. coelicolor* has 22 gene clusters that encode enzymes involved in the synthesis of secondary products that can function as antibiotics, siderophores, pigments, lipids and other molecules. Of particular interest is the production of antibiotics. There are four gene clusters that encode enzymes that produce characterised antibiotics: Actinorhodin - an aromatic polyketide antibiotic (Wright & Hopwood, 1976a), Red - red complex of undecylprodigiosin antibiotic (Rudd & Hopwood, 1980), the non-ribosomal peptide CDA - calcium-dependent antibiotic (Lahey *et al.*, 1983), and methylenomycin (Wright & Hopwood, 1976b), which is located on the giant linear plasmid SCP1.

Antibiotic clusters are each regulated by a specific regulatory protein (ActII-ORF4, RedD and CdaR) that initiates a pathway specific signalling cascade, which leads to the production of the antibiotic. The transcription factors that regulate the synthesis of antibiotics also undergo regulation to either stimulate or inhibit their transcription. The antibiotic clusters are also regulated by global regulators of the cell in response to nutrient starvation and environmental stress,

which also control the development of *S. coelicolor* (van Wezel & McDowall, 2011).

Regulation of the actinorhodin (ACT) production has been widely studied. The major transcription factor that activates the *act* cluster of *S. coelicolor* is ActII-ORF4 and the regulation of its promoter involves several mechanisms (van Wezel & McDowall, 2011). One of the main regulators is AbsA2, which is a repressor of *actII*-ORF4. AbsA2 is phosphorylated by AbsA1 and this phosphorylated AbsA2 can now bind to the promoter and inhibit its transcription (Sheeler *et al.*, 2005). Another repressor of the *actII*-ORF4 promoter is DasR, which binds directly to the promoter. It has a dual function as it senses a nutrient, glucosamine-6-phosphate, which decreases its affinity for the *actII*-ORF4 promoter and hence, ACT production is enhanced (Rigali *et al.*, 2008). AtrA is the main activator of *actII*-ORF4 promoter by inducing its transcription (Uguru *et al.*, 2005).

The *red* and *cda* clusters are regulated by the RedD and CdaR transcription factors, respectively. AbsA2 is also the protein that binds to these promoters and represses their activity. Another system that was shown to affect ACT and RED production is the AfsKRS system. AfsK phosphorylates AfsR, which then binds to *afsS* promoter and stimulates its transcription. The precise mechanism of action of AfsS has not been resolved yet; however, it has been shown that increasing the induction of *afsS* enhances antibiotic production. The biosynthesis of methylenomycin is regulated by MmyB, which is a transcriptional activator. It has been shown that the level of phosphate, nitrogen, carbon, and metals such as Zn and iron all play key roles on antibiotic production (van Wezel & McDowall, 2011).

In conditions of low nutrients, *S. coelicolor* switches on the production of antibiotics. The highly phosphorylated guanine nucleotide ppGpp has been shown to have a major role in the stringent response in *E. coli* (Section 1.2.5.1).



In *S. coelicolor*, ppGpp production decreases rRNA synthesis, ribosomal production, carbon metabolism, fatty acid biosynthesis, and amino acid biosynthesis (Hesketh *et al.*, 2007), and is additionally thought to play a major role in development and secondary metabolism. *S. coelicolor* has two homologues of the *E. coli* RelA/SpoT that synthesise ppGpp (Section 1.2.5.1). The first one is called RelA and it synthesises ppGpp when there is a depletion of amino acids or glucose. The second homologue of RelA/SpoT is RshA (Sun *et al.*, 2001). RshA produces ppGpp when the level of phosphate is reduced (Ryu *et al.*, 2007). Both RelA and RshA are similar to *E. coli* SpoT in that they both contain (p)ppGpp hydrolase and synthetase activity. ppGpp synthesis leads to the induction of *redD* and *actII*-ORF4 transcription and therefore, the stimulation of ACT and RED production (van Wezel & McDowall, 2011).

### **1.1.2 *Mycobacterium tuberculosis***

*M. tuberculosis* was identified by Robert Koch in 1882. He was awarded the Nobel Prize in Physiology or Medicine for this discovery in 1905 (Hopwood, 2007). The tubercle is a slender, straight or slightly curved bacillus. It is non-motile and non-encapsulated and, unlike *S. coelicolor*, does not form spores. Mycobacteria are obligate aerobes and cannot grow in the absence of oxygen. However, *M. tuberculosis* can survive long periods, months to years, of oxygen limitation in their inactive state (Wayne & Sohaskey, 2001). Unlike *S. coelicolor*, *M. tuberculosis* cannot survive sudden anaerobiosis and it is thought that they reduce nitrate by nitrate reductase to provide enough energy to adjust to this anaerobic condition (Wayne & Hayes, 1996).

#### **1.1.2.1 The genome**

*M. tuberculosis* is a Gram-positive bacterium with a high G+C content of 65.6%. It has a circular genome of 4,411,529 bp and about 4,000 genes (Cole *et al.*, 1998). The genome has 50 genes that encode stable RNA of which 45 are transfer RNAs. It has pathways involved in amino acid biosynthesis, vitamins, and enzyme co-factors with extensive metabolic pathways for carbohydrates

and lipids. It has more than 100 regulatory proteins of which thirteen are sigma factors (Cole *et al.*, 1998). Unlike other bacteria, *M. tuberculosis* has a large part of its genome dedicated to enzymes involved in lipogenesis and lipolysis to produce a complex cell envelope that is involved in its pathogenicity. Also, about 10% of its genome codes for two newly discovered large families of glycine-rich proteins (PE and PPE protein families), which might be essential in immunology (Cole *et al.*, 1998). Additionally, it encodes proteins involved in responses to oxidative-stress, heat shock, carbon starvation and other environmental stresses (Rodrigue *et al.*, 2006).

#### 1.1.2.2 Pathogenicity

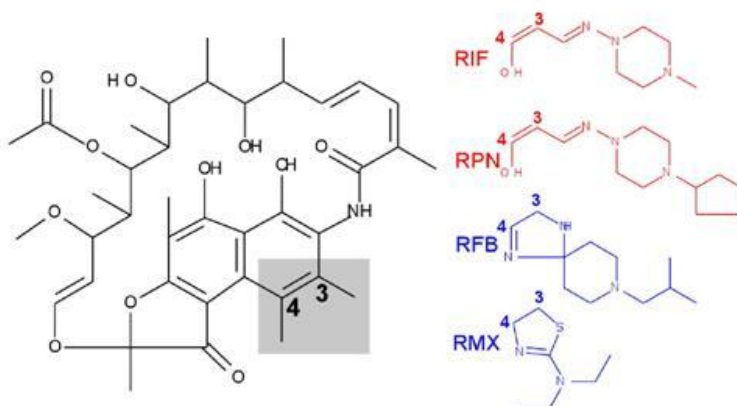
Tuberculosis is an airborne disease that continues to claim millions of lives around the world. There has been a rise in tuberculosis occurrence in people with immunodeficiency virus (HIV) or poor immune response, where it can be fatal (Cole *et al.*, 1998). Since *M. tuberculosis* can mutate into antibiotic-resistant strains, developing a cure for this disease is proving to be difficult (Cole *et al.*, 1998, Mukherjee & Chatterji, 2008). It is essential to try to understand these bacteria better so as to help develop new therapeutic interventions to treat and if possible to wipe out this disease.

*M. tuberculosis* is a slow growing bacterium that takes approximately 24 hours to reproduce. When it infects a human, the immune response controls the infection but does not destroy it. This bacterium then goes into a dormant state, rapidly decreasing rRNA and ribosome synthesis (Stallings *et al.*, 2009) and the immune system can no longer recognise it. Thus, later in life when the immune system is compromised, it can then attack the host and potentially cause death (Cole *et al.*, 1998). The tubercle has a complex cell envelope made up of peptidoglycan and a further layer rich in lipids (mycolic and mycocerosic acid), glycolipids (lipoarabinomannan) and polysaccharides, which might cause the inflammatory response by the host and play a role in pathogenesis (Cole *et al.*, 1998).

### 1.1.2.3 Therapeutics and importance of Rifampicin

*M. tuberculosis* has natural resistance to antibiotics, due in part, to the nature of the cell envelope, as well as resistance genes (Cole *et al.*, 1998). All these factors contribute to the persistency of this bacterium. Therefore, treatment of tuberculosis involves a combination of multi-drug therapy including Rifampicin (Rif), isoniazid, pyrazinamide, ethambutol and streptomycin (WHO) (Villain-Guillot *et al.*, 2007). As *M. tuberculosis* is an intracellular pathogen, the drug needs to diffuse into the tissues, the cells and bacteria to exert its effect. Rif has this property and is nowadays one of the most effective drugs against tuberculosis (Villain-Guillot *et al.*, 2007).

Rifamycins are naturally occurring antibiotics that were first discovered to belong to the ansamycin family. Ansamycin was isolated from the strain *Amycolatopsis mediteranei* by Sensi *et al.*, in 1959. It contains an aromatic naphthoquinone moiety and a “tail” on C3 and/or C4 position that can be altered (Fig. 1.2) (Artsimovitch *et al.*, 2005). Rif is a synthetic derivative of the rifamycins and is effective against Gram-positive and some Gram-negative bacteria (Tupin *et al.*, 2010). It binds to the  $\beta$ -subunit of RNAP in the DNA binding pocket deep within the DNA/RNA channel. However, it is separated from the  $Mg^{2+}$  in the RNAP active site by about 12.1 Å (Campbell *et al.*, 2001). There are 23 residues situated around the DNA binding pocket of RNAP that have been shown to interact with Rif. The bonds are mainly van der Waal’s forces interacting with the naphthol ring of Rif and some hydrogen bonding with polar groups of Rif. However, only twelve of these residues directly interact with Rif (Campbell *et al.*, 2001, Villain-Guillot *et al.*, 2007).



**Figure 1.2: The overall structure of the Rifamycins.**

On the left is the naphthoquinone moiety and on the right are the various tails that can be attached on C3 and/or C4 position. The red tails are attached to C3 (RIF: Rifampicin and RPN: rifapentin) and the blue tails are attached on C3/C4 (RFB: rifabutin and RMX: rifamexyl). Copyright license No.: 2937590289248 (Artsimovitch *et al.*, 2005).

The mechanism by which Rif inhibits RNAP is proposed to be by steric hindrance. Rif binds to the DNA channel within RNAP and directly blocks the path of a 2 to 3 nt long RNA transcript using its naphthol ring (Campbell *et al.*, 2001). However, as Rif activity is altered by different sigma factors, this proposed model was reviewed. It was proposed that Rif binds to sigma region 3 and causes an allosteric signal that dissociates the  $Mg^{2+}$  ion from the active centre of RNAP (Artsimovitch *et al.*, 2005), which diminishes the catalytic activity of the active site leading to the short unstable RNA/DNA hybrid to dissociate from the RNAP (Artsimovitch *et al.*, 2005, Villain-Guillot *et al.*, 2007). Recently, the allosteric model for Rif action proposed by Artsimovitch *et al.*, has been called into question (Feklistov *et al.*, 2008). Rif does not have any effect on the elongating RNAP (Campbell *et al.*, 2001).

The sequence homology is not extensive between bacterial RNAP and eukaryotic RNAP. This makes Rif a good drug to combat tuberculosis and to be used for long periods (Villain-Guillot *et al.*, 2007). However, RNAP can acquire mutations, which render the RNAP resistant to Rif. Most of the mutations isolated so far in RNAP appear to be around the active site and are present in a short 81 nt stretch of the  $\beta$ -subunit known as the “rifampicin resistance-determining region” (Jin & Gross, 1988, Malshetty *et al.*, 2010). These mutants can become insensitive to Rif but might decrease their “fitness”. However, if these mutants survive they may later obtain other mutations that might restore their “fitness” level and allow these mutants to persist leading to Rif resistance (Malshetty *et al.*, 2010).

Rif resistance can also be imposed by RNAP binding proteins e.g. DnaA and RbpA. In *E. coli*, DnaA, which is a key initiator of DNA replication, was also found to interact with RNAP at the  $\beta$ -subunit and influence Rif resistance. Nevertheless, it is not known whether DnaA prevents Rif binding, or whether the presence of DnaA stabilises the  $Mg^{2+}$  and keeps RNAP active (Flatten *et al.*, 2009). RbpA is the subject of this thesis, but the mechanism by which it affects Rif resistance in *S. coelicolor* is not known (Newell *et al.*, 2006). Bacteria can also modify Rif making it inactive by phosphorylation, ribosylation and glucosylation (Tupin *et al.*, 2010). Therefore, due to the growing problem with Rif resistance, it is mostly used selectively in treatments of tuberculosis, meningitis and brucellosis. It is also used in a combination therapy with other drugs to reduce its resistance (Villain-Guillot *et al.*, 2007).

## **1.2 Bacterial gene regulation**

Transcription is the process of transferring genetic information from DNA to RNA. This process of RNA synthesis is catalysed by the enzyme DNA-dependent RNA polymerase (RNAP) and is the first step in the production of protein from DNA (Murakami & Darst, 2003). RNAP activity is controlled by

factors that bind to it and exert their effect, either positively or negatively (Section 1.2.5-1.2.9). mRNA is then translated to protein, which can then exert its function in the cell.

### 1.2.1 Bacterial DNA-dependent RNA polymerase

Bacteria have only one form of RNAP, which is a multi-subunit enzyme. The core RNAP is composed of five subunits; two alpha subunits ( $\alpha$ ), one beta subunit ( $\beta$ ), one beta prime subunit ( $\beta'$ ), and an omega ( $\omega$ ) subunit ( $\alpha_2\beta\beta'\omega$ ) (Young *et al.*, 2002). It has a molecular mass of approximately 400 kDa, and the overall structure is well conserved in bacteria, archaea and eukaryotes. However, the mechanism for promoter recognition is distinct, as are the factors that bind to RNAP to modulate its activity (Zhang *et al.*, 1999). Addition of the bacterial sigma factor,  $\sigma$ , to RNAP forms the RNAP holoenzyme, which is the fully functional enzyme that can recognise promoter DNA and initiate transcription.

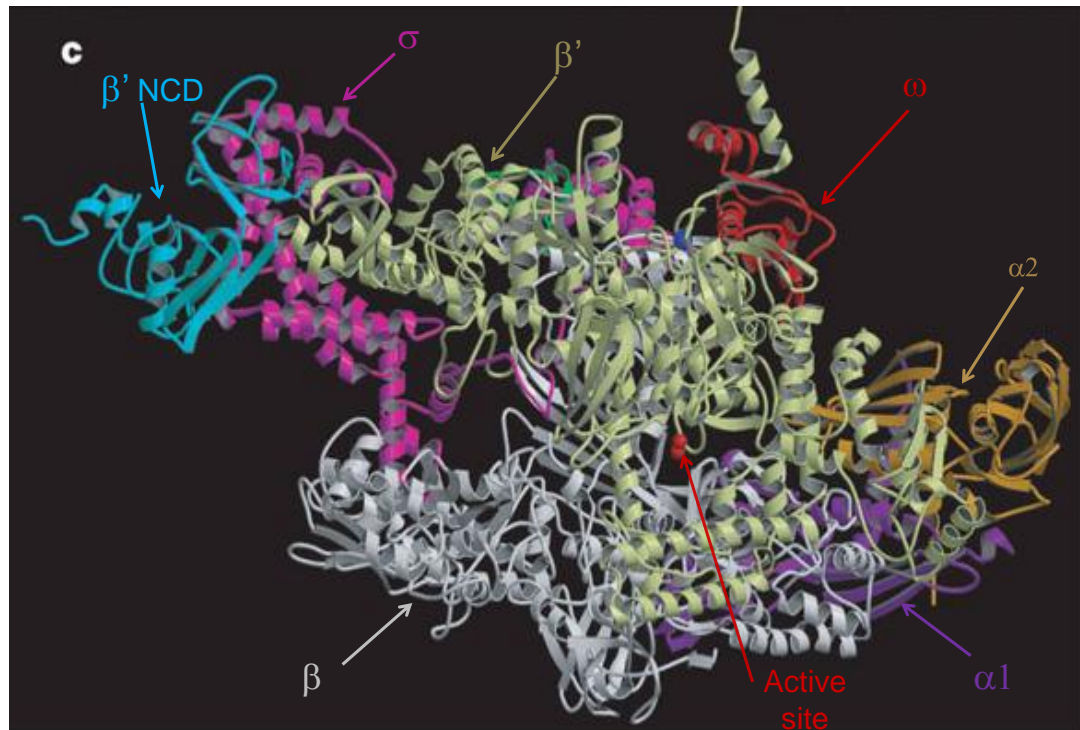
Obtaining pure RNAP crystals for X-ray crystallography proved difficult, therefore, the three dimensional structure of *E. coli* RNAP holoenzyme was first determined using electron microscopy (Darst *et al.*, 1989). The first crystal structure of the core bacterial RNAP was determined at 3.3 Å resolution by Zhang, G., and colleagues and it was that of *Thermus aquaticus* (Zhang *et al.*, 1999). It was shown from this structure that RNAP looks like a 'crab-claw' with one jaw composed mainly of the  $\beta$  subunit and the other of the  $\beta'$  subunit (Zhang *et al.*, 1999, Young *et al.*, 2002). The RNAP molecule is 150 Å long, 115 Å tall and 110 Å wide with a channel width of ~27 Å (Zhang *et al.*, 1999).

The  $\alpha$  subunits dimerise and are situated at the interface of the  $\beta$  and  $\beta'$  jaws. The  $\alpha$  subunit is composed of two domains: the  $\alpha$ -N terminal domain ( $\alpha$ -NTD), which is ~26 kDa and is flexibly linked to the  $\alpha$ -C terminal domain ( $\alpha$ -CTD) that is ~9 kDa. The  $\alpha$ -NTD forms a dimer and acts as a scaffold for the assembly of the  $\beta$  and  $\beta'$  subunits. The  $\alpha$ -CTD however, interacts with transcriptional

regulator proteins and binds non-specifically with sequences of the promoter DNA (Blatter *et al.*, 1994). The  $\omega$  subunit interacts extensively with the C-terminal tail of the  $\beta'$  subunit and this subunit is thought to have a function in the finishing stages of the assembly of the core RNAP (Zhang *et al.*, 1999).

The active site of RNAP is found at the back of the channel where the  $\beta$  and  $\beta'$  subunits form a major contact. It has a chelated  $Mg^{2+}$  ion co-ordinated by three Aspartate (Asp - D) residues found in the  $\beta'$  subunit conserved motif - NADFDGD. This active centre is vital for the catalytic activity of RNAP (Zhang *et al.*, 1999).

The  $\beta$  and  $\beta'$  subunits are together composed of five mobile modules;  $\beta$ -upstream lobe,  $\beta$ -downstream lobe, the  $\beta$  flap, the  $\beta'$  jaw and the  $\beta'$  clamp, which all surround the active site of the enzyme (Ghosh *et al.*, 2010). These mobile domains allow the binding of the sigma factor and the promoter DNA. In transcription initiation, the clamp swings open the channel so DNA can enter the active site but during elongation, the clamp closes. This helps in the outstanding processivity of transcription (Murakami & Darst, 2003). The base of the  $\beta'$  subunit has a Zn (II) ion that is complexed by four cysteine residues and plays an important structural role in the folding of the  $\beta'$  subunit (Zhang *et al.*, 1999).



**Figure 1.3: The crystal structure of *Thermus thermophilus* RNAP holoenzyme at 2.6 Å resolution.**

The subunits are coloured:  $\beta$  (white),  $\beta'$  (tan),  $\beta'$  non-conserved domain (blue),  $\alpha 1$  (purple),  $\alpha 2$  (yellow),  $\sigma$  (pink), and  $\omega$  (red). The figure shows two catalytic  $\text{Mg}^{2+}$  ions as red spheres (active site) and  $\text{Zn (II)}$  ion as a blue sphere (at the base of  $\beta'$ ). Copyright license No.: 2942490346046 (Vassilyev *et al.*, 2002).

### 1.2.2 Sigma factors

Sigma factors are the sixth subunit of the bacterial RNAP and are essential for transcription initiation. The sigma factor recruits RNAP to the promoter DNA, and aids in DNA melting and thus, the formation of the open complex (Section 1.2.4.1). It is also involved in promoter escape and clearance (Section 1.2.4.1). Sigma factors are specific to particular promoters and hence, are involved in the regulation of gene expression.

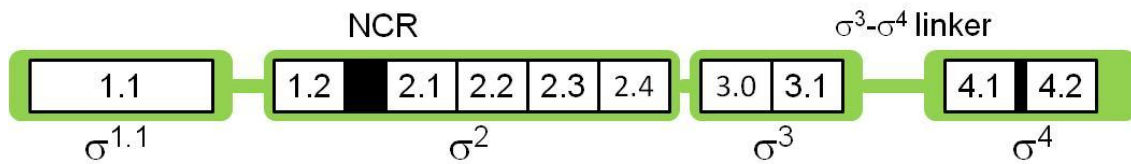


#### 1.2.2.1 The different groups of the $\sigma^{70}$ -related family

There are two families of sigma factors; the  $\sigma^{70}$ -related family and the  $\sigma^{54}$  family. In *E. coli*, the principal sigma factor is  $\sigma^{70}$ , which is homologous to the vast majority of sigma factors and acts as a model for the  $\sigma^{70}$  family (Wilson & Dombroski, 1997). Members of the  $\sigma^{54}$  family are structurally and mechanistically distinct and activate the transcription of genes that are involved in the metabolism of nitrogen, pathogenicity and cytotoxicity (Ghosh *et al.*, 2010). The  $\sigma^{54}$  family appears to be absent from the Actinobacteria and so these are not discussed further here (Sachdeva *et al.*, 2010).

$\sigma^{70}$ -related sigma factors recognise specific sequences on the promoter DNA: the -10 element, with a consensus sequence of TATAAT, which is centred at position -10 from the transcription start site and the -35 element, with a consensus sequence of TTGACA, starts at position -35 from the transcription start site. Some promoter regions have an “extended -10 element”, TGnTATAAT, that is sufficient to direct transcription without the need for the -35 element (Barne *et al.*, 1997). rRNA promoters have a “discriminator region”, a rich GC region, downstream of the -10 element that affects the strength and regulation of the promoter (Haugen *et al.*, 2008). An “UP element”, which is an AT-rich region positioned between 60 and 40 bp upstream of the transcriptional start point, is bound by the RNAP  $\alpha$ -CTD and stimulates the activity of the promoter (Ross *et al.*, 1993).

There are four groups of  $\sigma^{70}$ -related factors. Group I (primary) sigma factors are essential for survival and direct the transcription of the house-keeping genes. Most bacteria have only one primary sigma factor, which comprise four main domains,  $\sigma^{1.1}$ ,  $\sigma^2$ ,  $\sigma^3$  and  $\sigma^4$ , that are flexibly linked. These domains contain conserved sub-regions 1.1, 1.2-2.4, 3.0-3.1 and 4.1-4.2, respectively (Section 1.2.2.2) (Murakami & Darst, 2003).



**Figure 1.4: Schematic diagram of the  $\sigma^{70}$  principal sigma factor.**

It is showing the four distinct domains:  $\sigma^{1.1}$ ,  $\sigma^2$ ,  $\sigma^3$  and  $\sigma^4$  along with its regions 1.1, 1.2-2.4, 3.0-3.1 and 4.1-4.2, respectively. The diagram also shows the domains connected by linkers,  $\sigma^3$ - $\sigma^4$  linker and the NCR between 1.2 and 2.1 (Adapted from Murakami & Darst, 2003).

Group II sigma factors are similar to those of Group I, although they lack most of region 1.1. They are non-essential for bacterial growth and survival nevertheless, they normally control the transcription of genes involved in universal stress responses and genes expressed during stationary phase (e.g.  $\sigma^S$  in *E. coli*) (Loewen & Hengge-Aronis, 1994). They also recognise the -10 element, the -35 element and the “extended -10 element”. The only difference is that in the “extended -10 promoters”, domain 3 contacts a C nucleotide at position -13 (Paget & Helmann, 2003).

The third group of sigma factors, Group III, are more diverse structurally and functionally. They lack region 1.1, however they consist of domains 2-4. They are involved in controlling clusters of genes involved in heat shock, sporulation and flagellar biosynthesis. Their primary promoter elements are different from Group I and Group II sigma factors, although they have some conserved nucleotides in the -10 element (Paget & Helmann, 2003, Rodrigue *et al.*, 2006).

The largest numbers of sigma factors belong to Group IV and they are also known as the Extra-Cytoplasmic Function (ECF) sigma factors (Lonetto *et al.*, 1994). As the name implies, they respond mainly to extra-cellular signals from the environment and might transcribe genes involved in stress responses or cell-

wall homeostasis. They only have conserved regions 2 and 4 and lack domain 1.1 and 3. An example of an ECF sigma factor is  $\sigma^R$  of *S. coelicolor* that controls the thiol-disulphide balance in the hyphae (Paget *et al.*, 2001b, Paget *et al.*, 2002, Paget & Helmann, 2003).

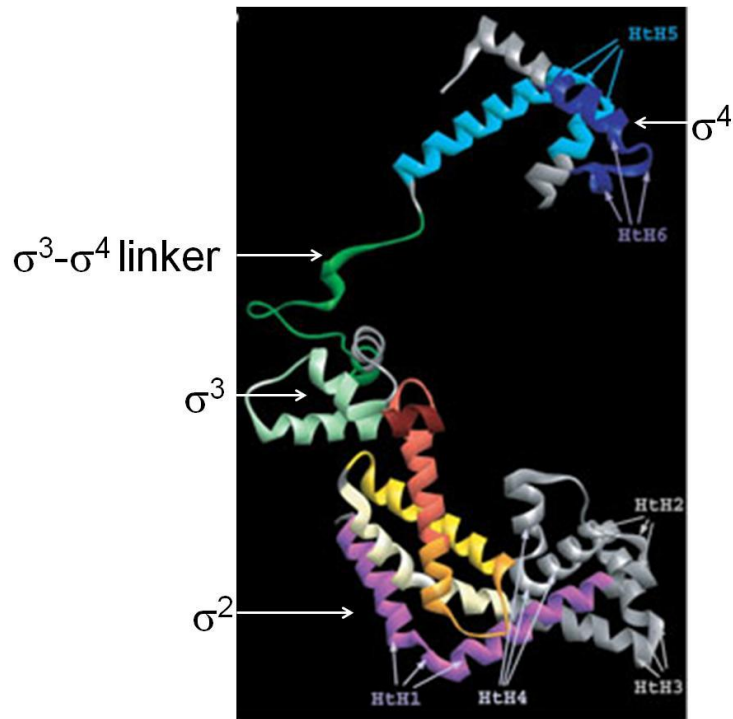
#### 1.2.2.2 The domains of the principal sigma factors

##### $\sigma^{1.1}$

Region 1.1 has a poorly conserved sequence, but examples tend to retain their acidic property (Murakami *et al.*, 2002b). The main function of region 1.1 is to auto-inhibit free sigma factor from recognising promoter sequences. However, when sigma binds to RNAP this auto-inhibition is lost (Murakami & Darst, 2003). Region 1.1 is also thought to enhance the rate of transition from closed complex to open complex at some promoters (Wilson & Dombroski, 1997, Murakami *et al.*, 2002b). In *T. aquaticus*  $\sigma^A$ , region 1.1 is buried in the active site of RNAP (Murakami *et al.*, 2002a). The acidic region of  $\sigma^{1.1}$  is located N-terminally of the  $\alpha$ -helix and is thought to interact with the walls of the channel and thus possibly assist in the widening of the channel to allow DNA entry into the active site. This leads to an increase in the rate of open complex formation (Murakami & Darst, 2003).

##### $\sigma^2$

The 2.6 Å resolution crystal structure of region 1.2-2.4 was first solved in isolation (Malhotra *et al.*, 1996). It showed that this region is composed mainly of helices interconnected by loops forming four helix-turn-helix (HtH) motifs: HtH1, HtH2, HtH3 and HtH4 (Fig. 1.5) (Vassylyev *et al.*, 2002). The sigma domain 2 containing conserved regions 1.2, 2.1, 2.2, 2.3 and 2.4 extensively binds to every nucleotide between -12 to -6 of the melted -10 element, with most interactions occurring in region 2.3 (Feklistov & Darst, 2011).



**Figure 1.5: A ribbon diagram of the principal sigma factor,  $\sigma^{70}$ .**

It is composed mainly of helices connected by turns or loops.  $\sigma^2$  is composed of four HtH with region 1.2 (purple), non-conserved region (grey), 2.1 (tan), 2.2 (yellow), 2.3 (orange), 2.4 (red).  $\sigma^3$  is composed of region 3.0 (dark red) and 3.1 (light green),  $\sigma^3$ - $\sigma^4$  linker composed of region 3.2 (dark green), and  $\sigma^4$  is composed of two HtH with region 4.1 (light blue) and 4.2 (dark blue). Copyright license No.: 2942490346046 (Vassilyev *et al.*, 2002).

Region 1.2 interacts with residues in the  $\beta'$  coiled-coil domain of RNAP and is thought to aid region 2.3 in forming contacts with the -10 element and so cause DNA melting (Fig. 1.5). Region 1.2 also affects the recognition of the -10 element by region 2.4 (Baldwin & Dombroski, 2001, Zenkin *et al.*, 2007). In rRNA and tRNA promoters, region 1.2 contacts the non-template strand at the “discriminator” region and this affects the prolonged existence of the open complex (Haugen *et al.*, 2008). The “discriminator” region is located two base pairs downstream of the -10 element and at this position the nucleotide can be a cytosine or a guanine, depending on the extent of regulation. In the *E. coli rrn P1*

promoter, it was observed that if the nucleotide at this position is a cytosine, which is common in rRNA promoters, then the interaction of region 1.2 with the discriminator element is weak and hence, short-lived open complexes are formed with RNAP. Most promoters, however, appear to have a guanine at the “discriminator” region forming longer-lived complexes with RNAP. Methionine (Met) 102 was the candidate amino acid in sigma region 1.2, which is thought to interact with guanine at this position, thereby enhancing the existence of the open complex (Haugen *et al.*, 2008). The stability of the open complex can have a crucial role on the response of the promoter to the global effector, ppGpp (Section 1.2.5.1).

Some residues in region 2.1 are involved in binding to the  $\beta'$  coiled-coil domain of RNAP. Region 2.1 is comprised of two  $\alpha$ -helices: the C-terminal part of helix 12a and helix 12b have approximately 45° kink between them (Fig. 1.5). Most of the conserved residues in this region are located close to the kink and are involved in strong interactions with RNAP (Malhotra *et al.*, 1996). Residues in region 2.1 also make up 23% of the interactions between sigma domain 2 and the -10 element (Feklistov & Darst, 2011).

Region 2.2 is composed almost entirely of helix 13 (Fig. 1.5) and has most of the hydrophobic residues that are involved in the interaction with the clamp-helix region of the  $\beta'$  subunit of RNAP, which is one of the major interactions between sigma and RNAP (Campbell *et al.*, 2002, Yang & Lewis, 2010). These residues are conserved in 97% of all sigma factors and thus, region 2.2 is the most highly conserved region of the sigma factor (Malhotra *et al.*, 1996). Conserved basic residues in region 2.2 also interact with the non-template strand of the negatively charged DNA (Murakami *et al.*, 2002a). In *T. aquaticus*  $\sigma^A$ , R237 forms hydrogen bonds with the O<sub>4</sub> atom of T<sub>-12</sub> in the -10 element, making T<sub>-12</sub> a very highly conserved nucleotide of the -10 element (Feklistov & Darst, 2011).

There are six conserved aromatic residues in region 2.3 that are thought to be crucial for promoter melting leading to open complex formation: two tyrosines (Y), two tryptophans (W) and two phenylalanines (F). In *E. coli*, four aromatic residues Y425, Y430, W433 and W434 are solvent exposed and located on helix 14 causing promoter melting. The two phenylalanine residues have an important function in the formation of the hydrophobic core to ensure appropriate folding of the protein (Malhotra *et al.*, 1996).

Region 2.4 has an amphipathic helix with hydrophobic residues contributing to the hydrophobic core interaction with RNAP and hydrophilic residues directly interacting with bases of the -10 element promoter. In *T. aquaticus*  $\sigma^A$ , Q260 contacts A<sub>-12</sub> on the template strand only when strand separation is achieved at -11 (Feklistov & Darst, 2011). In *E. coli*, Arg-441 contacts the -13 base of the -10 element (Malhotra *et al.*, 1996, Murakami & Darst, 2003).

### $\sigma^3$

Some promoters have an “extended -10 element”, which is recognised by amino acids in region 3 and removes the need for a -35 element (Campbell *et al.*, 2002). Domain 3 is composed of three helices (Fig. 1.5); the first helix is of particular importance as it interacts with DNA sequences upstream from the -10 element in the “extended -10 promoter” (Murakami & Darst, 2003). In *E. coli* region 3.0, Q485 recognises the extended -10 sequence whilst the H455 non-specifically binds to the phosphate backbone of the non-template strand (Murakami *et al.*, 2002a). This interaction stabilises the open complex formation (Murakami & Darst, 2003).

### $\sigma^4$

$\sigma^4$  is composed of four helices forming two HtH motifs: HtH5 and HtH6 (Fig. 1.5) (Vassilyev *et al.*, 2002). It has a concave pocket containing the hydrophobic residues of region 4.1 and these residues interact with the  $\beta$ -flap of RNAP. This interaction is important to help position region 4 in relation to region 2 and

hence, allow the simultaneous contact with the -35 and -10 elements, respectively (Dove *et al.*, 2003). Region 4.2 recognises and binds to the -35 element inducing a bend in the DNA which brings the upstream DNA closer to RNAP. This kink enhances interaction between the  $\alpha$ -CTD of RNAP with DNA upstream from the -35 element and with activators present upstream of the -35 element (Murakami & Darst, 2003).

#### $\sigma^3$ - $\sigma^4$ linker

Connecting domain 3 and 4 is the 33-residue linker (Fig. 1.5). This linker starts at domain 3, runs near the active site of RNAP, through the RNA exit channel and connects to domain 4 of the sigma factor. This linker is thought to be involved in maintaining the stability of the incoming nucleotide during transcription initiation as it is present near the active site of RNAP. Also, it is thought to help in sigma dissociation from the core RNAP after transcription initiation as it gets displaced by the nascent RNA from the RNA exit channel causing weakening of the  $\sigma^4$  domain and finally the rest of the sigma factor (Section 1.2.4.1- promoter escape and clearance) (Paget & Helmann, 2003).

### 1.2.3 The sigma factors of Actinobacteria

#### 1.2.3.1 *S. coelicolor*

The *S. coelicolor* genome was found to encode 65 sigma factors, all of which belong to the  $\sigma^{70}$  family (Bentley *et al.*, 2002). Prior to this, initial studies found that RNAP is heterogenous and exists in numerous forms containing *S. coelicolor*  $\sigma^{49}$ ,  $\sigma^{35}$  and  $\sigma^{28}$  (Westpheling & Brawner, 1989). An *rpoD* probe derived from ten conserved amino acids in the principal sigma factor from *Myxococcus xanthus* was used to search the *S. coelicolor* genome for sigma factors (Buttner *et al.*, 1990). This probe helped in the identification of four genes that encode sigma factors:  $\sigma^{\text{HrdA}}$ ,  $\sigma^{\text{HrdB}}$ ,  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$  (hrd =homologue of *rpoD* gene) (Buttner *et al.*, 1990).

Three of these genes, *hrdA*, *hrdC*, and *hrdD*, are non-essential and strains comprising deletion mutants of all three genes are viable and are unaffected in terms of morphology, differentiation and antibiotic production (Buttner *et al.*, 1990, Tanaka *et al.*, 1991). However, it proved impossible to delete the *hrdB* gene, which suggested that  $\sigma^{\text{HrdB}}$  is the principal sigma factor in *S. coelicolor* (Buttner *et al.*, 1990). The complete nucleotide sequence of  $\sigma^{\text{HrdB}}$  was determined by Shiina and colleagues (Shiina *et al.*, 1991).  $\sigma^{\text{HrdB}}$  has 511 amino acids and a molecular weight ( $M_w$ ) of 55,902 (StrepDB).

Apart from the principal ( $\sigma^{\text{HrdB}}$ ) and principal-like sigma factors ( $\sigma^{\text{HrdA}}$ ,  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$ ), *S. coelicolor* encodes ten sigma factors that are related to the stress response sigma factor in *Bacillus subtilis*,  $\sigma^{\text{B}}$  (Paget & Helmann, 2003). These include:  $\sigma^{\text{B}}$ , which has a role in osmoprotection and morphological development (Cho *et al.*, 2001) and  $\sigma^{\text{F}}$ , which is involved in spore maturation (Kelemen *et al.*, 1996). WhiG resembles the flagella-related sigma factors such as FliA (Helmann, 1991) and is essential for sporulation (Kelemen *et al.*, 1996). The remaining 51 sigma factors belong to the ECF family (Paget *et al.*, 2002), most with unknown function. Of those that have been investigated,  $\sigma^{\text{E}}$  is involved in maintaining the integrity of the cell envelope (Lonetto *et al.*, 1994),  $\sigma^{\text{R}}$  responds to disulphide stress (Paget *et al.*, 1998) and  $\sigma^{\text{BldN}}$  has a role in the formation of the aerial mycelium (Bibb *et al.*, 2000).

#### 1.2.3.2 *M. tuberculosis*

*M. tuberculosis* has thirteen sigma factors and all are members of the  $\sigma^{70}$  family.  $\sigma^{\text{A}}$  is the principal sigma factor (Group I) and has all the four conserved regions. It is particularly stable and its mRNA has a half-life of more than 40 min (Hu & Coates, 1999). It controls the majority of genes and is essential for the viability of this bacterium. It also modulates expression of virulence genes, which are thought to aid resistance to superoxide and improve intracellular growth (Rodrigue *et al.*, 2006) .



Another important sigma factor in *M. tuberculosis* is  $\sigma^B$ , which is the sole Group II member.  $\sigma^B$  mRNA has a much shorter half-life of 2.4 min, suggesting that it is involved in adaptive responses (Hu & Coates, 1999). The *sigB* gene is positioned at approximately 3 kb downstream from *sigA*. The sigma domains 2, 3 and 4 are well conserved between  $\sigma^A$  and  $\sigma^B$ .  $\sigma^B$  is auto-regulated and its expression can also be induced by  $\sigma^E$  in response to surface stress and by  $\sigma^H$  in response to heat shock or oxidative stress (Sachdeva *et al.*, 2010). Hence,  $\sigma^B$  seems to be involved in transcription of genes expressed as a result of stress or carbon starvation, and stationary phase survival genes (Rodrigue *et al.*, 2006).

$\sigma^F$  belongs to Group III and is induced by cold shock, oxidative stress, depletion of nutrients, presence of a number of antibiotics, and during stationary phase. It controls genes involved in maintaining the cell envelope and in stationary phase (Rodrigue *et al.*, 2006). The remaining ten sigma factors belong to Group IV, e.g.  $\sigma^H$ , an orthologue of *S. coelicolor*  $\sigma^R$ , that is required for survival at elevated temperatures and in oxidative conditions (Rodrigue *et al.*, 2006).

#### **1.2.4 Transcription in bacteria**

The transcription cycle is divided into three main stages; initiation, elongation and termination (Canals & Coll, 2009). During transcription, core RNAP binds to a sigma factor and DNA, undergoes various conformational changes, initiates *de novo* RNA synthesis, loses the sigma factor at some point during the cycle, and binds to other regulatory factors to aid in transcription elongation and/or termination.

##### **1.2.4.1 Initiation**

Initiation is the first stage of transcription and is a major point at which gene expression is regulated. Regulation can be brought about by proteins that bind to DNA and either positively or negatively affect transcription together with the use of alternative sigma factors that recognise different promoters (Paget *et al.*, 2002). Core RNAP interacts with a sigma factor to form the holoenzyme that can

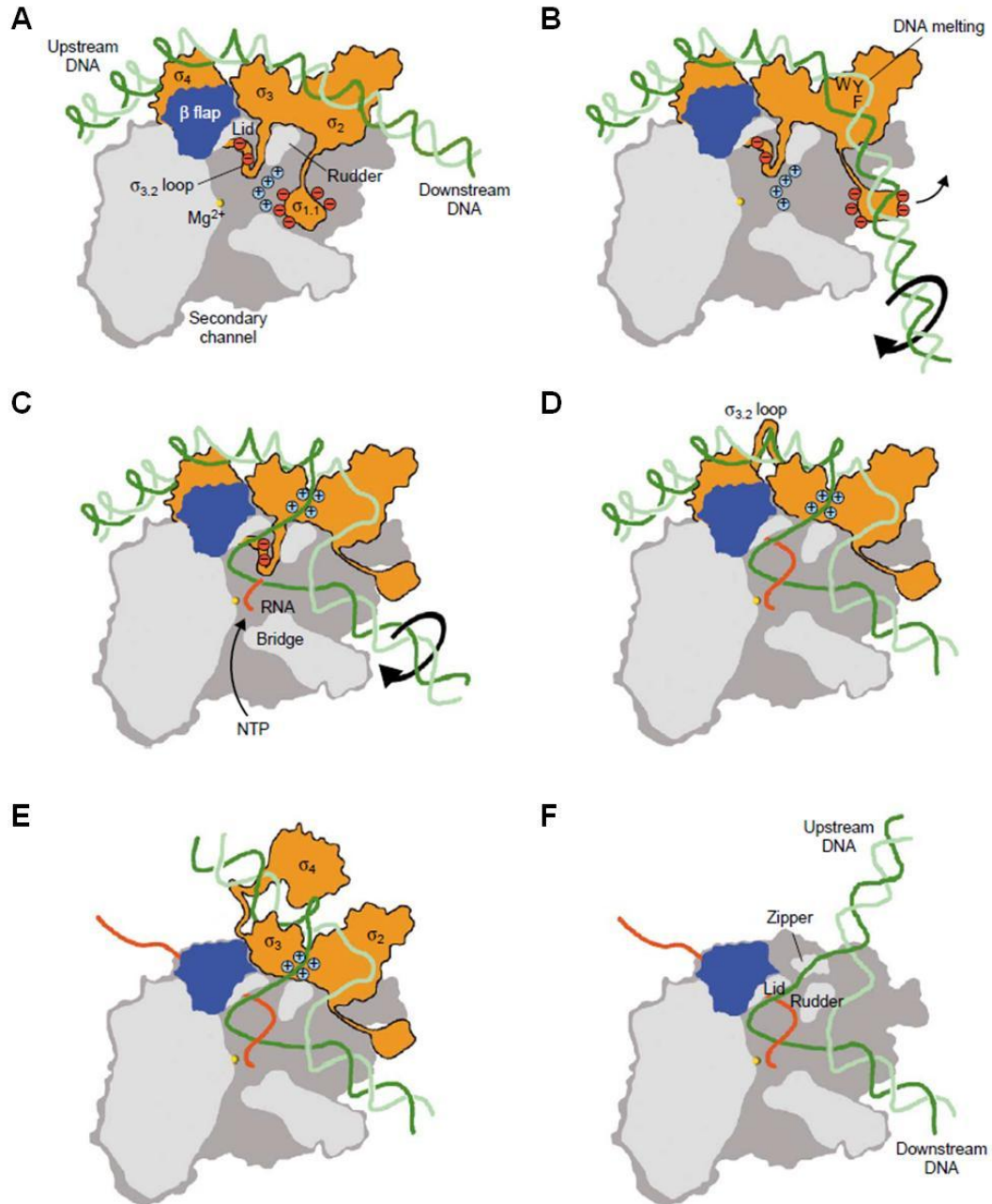
recognise the promoter in the double-stranded DNA and binds to it forming a closed complex. The RNAP melts the double-stranded DNA, around the initiation start site, into single-stranded DNA forming the open complex (transcription bubble) (Zhang *et al.*, 1999). *De novo* RNA synthesis is initiated using mononucleotides in the 5' to 3' direction (Young *et al.*, 2002). The nascent RNA follows the template for ~9 nt and leaves the RNAP *via* the RNA exit channel. RNA synthesis can be abortive until ~12 nt (See below). However, when the RNA finally escapes abortive initiation it proceeds into elongation (Murakami & Darst, 2003).

#### Holoenzyme formation

Core RNAP can bind to DNA non-specifically. However, it requires a sigma factor to recognise specific promoter sequences. The different domains of the sigma factor, together with the linkers, all interact with the core RNAP. This high affinity of sigma to the core RNAP provides stability to the holoenzyme (Murakami & Darst, 2003).

#### Formation of the closed complex

In the RNAP holoenzyme the DNA binding domains, 2 and 4, are solvent-exposed and together with the mobile modules of the holoenzyme can aid in altering the spacing between the sigma regions to allow them to bind to their respective elements on the DNA (Young *et al.*, 2002). The closed complex is formed when the RNAP holoenzyme contacts the promoter DNA at the -35 and -10 elements using region 4.2 and 2.4, respectively (Fig. 1.6A). The DNA lies across the outside of one face of the RNAP (Murakami *et al.*, 2002a). The  $\sigma^3$ - $\sigma^4$  linker sits close to the active centre and passes through the RNA exit channel. The  $\sigma^{1.1}$  domain interacts with the RNAP active site channel by electrostatic charges (Murakami & Darst, 2003).



**Figure 1.6: A cross-section view of RNAP during transcription initiation.**

The whole RNAP (grey),  $\beta$ -flap (blue),  $Mg^{2+}$  ion (yellow), sigma (orange), template strand (dark green), and non-template strand (light green) are shown. The negative charges (red) represent the acidic property of region 1.1 and  $\sigma^3$ - $\sigma^4$  linker. The positive charges (blue) represent the basic regions within the active channel and  $\sigma^2$ . **(A)** Closed complex formation. **(B)** Showing initial DNA melting,

DNA kinking, and displacement of the domain 1.1. **(C)** This representation shows downstream DNA melting, nucleotide entry (NTP), and *de novo* RNA synthesis (red). **(D)** Showing RNA elongation after  $\sigma^3$ - $\sigma^4$  linker is displaced. **(E)** Showing  $\sigma^4$  losing interaction with RNAP and the -35 element. **(F)** Sigma dissociated off RNAP. Copyright license No.: 2937560968109 (Murakami & Darst, 2003).

### Open complex formation

Promoter melting commences when the RNAP holoenzyme contacts the double-stranded DNA and unwinds the DNA from -11 to +3 without requiring any external energy. In *T. aquaticus*  $\sigma^A$ , strand separation starts at -11 when W256 interacts with A<sub>-11</sub> of the -10 element on the non-template strand (Fig. 1.6B) and causes the base to flip out and sit in a hydrophobic pocket in domain 2. This pocket is composed of Y253, R246, F242 and F248 which all interact with A<sub>-11</sub>. When the DNA is absent, the side chain of R246 is unstable and moves, which allows A<sub>-11</sub> to flip into the pocket. However, this side chain becomes stabilised by E243 and the phosphate of the T<sub>-12</sub> when A<sub>-11</sub> is in the pocket. A<sub>-11</sub> is very highly conserved as its pocket can only accommodate an A. This is thought to initiate full DNA melting (Feklistov & Darst, 2011).

Residues in region 2.3 contact the non-template strand of the promoter DNA. This is followed by the unwinding of the DNA and induces a 90° bend in the DNA between -10 and -9 (Fig. 1.6B) (Feklistov & Darst, 2011). In *T. aquaticus*  $\sigma^A$ , T252 is the most conserved residue that interacts with the phosphate of A<sub>-9</sub> and plays a critical role in DNA melting. In addition, T255 interacts with the base of A<sub>-8</sub>. These residues cause the T<sub>-7</sub> base to also flip out into a hydrophilic pocket that is composed of residues from region 1.2, 2.1 and 2.3. This interaction, together with the recognition of the discriminator region, might be a final check before DNA melting continues towards the transcription start site (Fig. 1.6B) (Feklistov & Darst, 2011).

$\sigma^{1.1}$  is then displaced from the active site channel (Fig. 1.6B) and this together with the conserved basic residues of region 2.3 and domain 3 direct the melted template DNA strand through a tunnel into the active site of RNAP (Fig. 1.6C) (Murakami *et al.*, 2002a). The DNA region +5 to +12 is held by the  $\beta$  and  $\beta'$  subunits contributing to the stability of the complex. *De novo* RNA synthesis starts by the incorporation of the incoming nucleotides from the 5' to 3' direction at the active site of the RNAP (Fig. 1.6C). The nucleotides enter the active site through a secondary channel (Fig. 1.6C). A phosphodiester bond is formed between the incoming nucleotides, and the nascent RNA starts to elongate (Murakami & Darst, 2003).

#### Abortive initiation

Abortive initiation is the recurring synthesis of short RNAs and their release from the RNAP. Nascent RNA with 5-6 nt can be supported in the channel of RNAP. However, when the RNA elongates to 12-14 nt, the nascent RNA encounters the  $\sigma^3$ - $\sigma^4$  linker in the RNA exit channel and competes with the linker for binding (Fig. 1.6C). This leads to an arrest in initiation and destabilises the RNA transcript causing the formation of abortive transcripts (Murakami *et al.*, 2002a, Kulbachinskiy & Mustaev, 2006).

The mechanism of abortive initiation is thought to occur by “scrunching”, whereby single-stranded DNA downstream from the transcription start site is pulled into the RNAP 1 bp per phosphodiester bond (Kapanidis *et al.*, 2006). The single-stranded DNA is accommodated within the RNAP. When the nascent RNA is aborted and released, RNAP pushes out the accumulated DNA restoring its initial position ready for another round of transcription initiation (Kapanidis *et al.*, 2006).

#### Promoter escape and clearance

After numerous rounds of abortive initiation, the  $\sigma^3$ - $\sigma^4$  linker is eventually displaced by the ~12 nt long nascent RNA (Fig. 1.6D), which will destabilise the

interaction between domain 4 and the  $\beta$ -flap (Section 1.2.2.2-  $\sigma^4$ ) (Kulbachinskiy & Mustaev, 2006). The first step of sigma dissociation from RNAP involves the  $\sigma^4$  domain losing interaction with RNAP, and consequently the -35 element (Fig. 1.6E) (Yang & Lewis, 2010). The complex can now enter the elongation phase where the rest of the sigma factor is subsequently dissociated from the RNAP (Fig. 1.6F). This is the classical model of promoter escape and clearance (Murakami & Darst, 2003, Yang & Lewis, 2010).

There is still debate as to whether the sigma factor dissociates immediately after promoter escape or whether it is still attached to the RNAP and dissociates during elongation (Murakami & Darst, 2003). It is thought that most sigma factors are stochastically released from the RNAP early on in transcription elongation. However, there is some evidence that a small fraction of sigma factors are present in the elongation complex and either get displaced during elongation, or remain bound throughout the transcription cycle (Yang & Lewis, 2010).

#### 1.2.4.2 Elongation

Transcription elongation occurs directly after RNAP escapes the promoter region of the DNA. The nascent RNA is ~12 nt long and starts elongating by forming phosphodiester bonds with the incoming nucleotides. To ensure efficient transcription elongation, proteins bind to RNAP to control the rate of transcription, avoid premature termination, and act in response to intrinsic pause signals. These regulatory proteins are known as elongation factors (Belogurov *et al.*, 2009).

NusA is an essential elongation factor, which plays a role in anti-termination, stalling of RNAP at RNA hairpin regions, termination, and in *E. coli*, it negatively regulates the activity of toxic foreign genes (Davies *et al.*, 2005, Yang *et al.*, 2009). NusA has an N-terminal domain that interacts with RNAP at the  $\beta$ -flap tip and a C-terminal domain that is thought to non-specifically bind to RNAP to

stabilise the complex. The CTD of NusA is thought to interact with the transcript and can therefore exert control over pausing and transcription termination (Yang *et al.*, 2009).

NusG and RfaH are elongation factors that increase the rate of transcript elongation. NusG is essential and binds to RNAP using its N-terminal domain to optimise transcription of most genes in *E. coli*. RfaH, which is evolutionarily related to NusG, is a non-essential elongation factor that binds to specific operons containing the *ops* element and induces the expression of foreign genes. The *ops* element recruits RfaH by displacing the CTD and allowing the buried region between NTD and CTD to be exposed so as to interact with RNAP. Another major difference between these two proteins is that NusG acts by possibly binding to Rho and thus, is involved in Rho-dependent termination, whereas RfaH does not act through Rho (Belogurov *et al.*, 2009).

Other elongation factors that maintain the processivity of transcription elongation are GreA and GreB. These proteins bind at the tip of the secondary channel of RNAP, where the nucleotides enter into the active site channel, using their CTD. The N-terminal coiled-coil domain extends into the RNAP active centre where it triggers the nucleolytic activity of the RNAP, cleaving a back-tracked RNA molecule. Transcription can then resume when the new 3' end is located at the active site, thereby avoiding transcription arrest (Davies *et al.*, 2005). Gre proteins are also involved in promoter escape and the restraint of promoter-proximal pausing (Stepanova *et al.*, 2007, Kusuya *et al.*, 2011).

#### 1.2.4.3 Termination

Transcription termination can be classified as intrinsic termination or factor-dependent termination. For intrinsic termination, elongation proceeds until a terminator sequence in the nascent RNA, is reached. These sequences are generally a short GC-rich hairpin structure followed by a rich sequence of

thymine repeats. On encountering these sequences, RNAP dissociates leading to termination (Kingsford *et al.*, 2007).

Termination can also be Rho-dependent requiring Rho protein. Rho is an ATP-dependent helicase or translocase that exists as a homo-hexamer and is found in many bacterial species. It facilitates termination by binding to newly synthesised RNA at the *rut* (Rho utilization) site, translocating along the chain and when it reaches the transcription elongation complex causes dissociation (Canals & Coll, 2009, Kalyani *et al.*, 2011).

### **1.2.5 Factors that affect RNAP activity**

RNAP is regulated by many effectors that interact directly or indirectly with it, to either inhibit or stimulate its activity. As the synthesis of proteins, rRNAs and ribosomal proteins utilises energy, these biosynthetic processes are strongly regulated to ensure they are carried out only when conditions are suitable (Lemke *et al.*, 2011). Hence, under stress conditions, or at a different growth phase, factors will affect the activity of RNAP at different promoters to ensure expression of appropriate genes, whilst hindering transcription of others.

#### **1.2.5.1 Guanosine 5'-diphosphate 3'-diphosphate (ppGpp)**

ppGpp is an alarmone molecule that controls RNA synthesis in the cell. It is produced by RelA and SpoT, with RelA being the major ppGpp synthetase. During amino acid starvation, the ratio of uncharged to charged tRNA increases and this switches on ppGpp synthesis. SpoT has both ppGpp synthetase and hydrolase activities. The activity of SpoT is stimulated by carbon, phosphate, iron, and fatty acid starvation and thus, ppGpp responds to nutrient richness. During amino acid starvation, the stringent response causes the build up of ppGpp in the cell, which will lead to an abrupt shutdown of rRNA synthesis and growth arrest (Jin *et al.*, 2011a, Lemke *et al.*, 2011).



At *rrn* promoters, ppGpp seems to have a key role in decreasing the half-life of open complexes and shifting the equilibrium towards the closed complex (Jin *et al.*, 2011a). ppGpp also regulates the expression of amino acid biosynthesis and uptake genes, whereas it down-regulates rRNA genes (Cavanagh *et al.*, 2010). In addition, ppGpp is thought to play a role in transcription elongation by stimulating RNAP pausing on *rrn* operons and other non-*rrn* promoters containing high G/C “discriminator” sequences (Haugen *et al.*, 2008, Jin *et al.*, 2011a).

Gourse and coworkers discovered that ppGpp works co-operatively with DksA to control rRNA synthesis (Section 1.2.6.1) (Lemke *et al.*, 2011). ppGpp is, however, the main effector in the control of rRNA synthesis. Apart from its direct effect on promoters, ppGpp can also act indirectly during the stringent response by redistributing RNAP to non-*rrn* promoters. The number of RNAP in the cell is limited for the activation of transcription from all promoters. In rich media, RNAP is recruited to rRNA promoters and catalyses rRNA synthesis; therefore, RNAP is limited for transcription of non-*rrn* genes. However, during the stringent response rRNA synthesis is shut down, releasing RNAP, which becomes available to transcribe non-*rrn* genes (Jin *et al.*, 2011a).

ppGpp was shown to target RNAP and bind at the substrate entry channel adjacent to the active centre of the *T. thermophilus* RNAP (Artsimovitch *et al.*, 2004). However, this interaction was shown to be incorrect and the exact position of interaction is still under investigation (Vrentas *et al.*, 2008). Therefore, further work in identifying the binding sites of ppGpp to RNAP will provide an insight as to the mechanism of action of ppGpp in transcription regulation during the stringent response.

#### 1.2.5.2 6S RNA

6S RNA is a small untranslated RNA that controls transcription during early and late stationary phase in *E. coli* (Cavanagh *et al.*, 2008). It is a double stranded

RNA that has a large bulge in the centre, resembling the DNA in the open complex conformation and as a result, it can be used as a template for RNAP (Wassarman & Saecker, 2006). The 6S RNA level builds up during late stationary phase and interacts with RNAP- $\sigma^{70}$  holoenzyme at the active site channel (Wassarman & Storz, 2000, Wassarman & Saecker, 2006).

Through late stationary phase, the 6S RNA binds to most of the  $\sigma^{70}$ -RNAP and inhibits RNAP- $\sigma^{70}$ -dependent transcription (Trotochaud & Wassarman, 2004). This RNAP- $\sigma^{70}$ /6S RNA complex sequesters  $\sigma^{70}$  and decreases  $\sigma^{70}$ -dependent transcription, possibly allowing other alternative sigma factors to bind to RNAP and exert their effect. However, when nutrients are encountered, 6S RNA serves as a template and the increase in NTP will direct transcription from the bound 6S RNA and lead to the release of  $\sigma^{70}$ , which can bind to RNAP and direct  $\sigma^{70}$ -dependent transcription (Wassarman & Saecker, 2006). This is the direct effect of 6S RNA.

The main contact point of 6S RNA with RNAP- $\sigma^{70}$  complex is with the  $\sigma^{70}$  region 4.2 (Klocko & Wassarman, 2009). Therefore, 6S RNA competes with promoter DNA for binding to region 4.2 of  $\sigma^{70}$ . The sensitivity of promoters to 6S RNA relies on the strength of the -35 element and the presence of the “extended -10 element” (Cavanagh *et al.*, 2008). A weak -35 element together with the presence of an “extended -10 element” increases the inhibition effect of 6S RNA towards the RNAP- $\sigma^{70}$ -dependent transcription (Cavanagh *et al.*, 2008).

The 6S RNA also has an indirect effect on transcription regulation. During early stationary phase 6S RNA enhances the transcription of *reIA*, which will increase the level of ppGpp and thus, the expression of amino acid biosynthesis genes and the shutdown of rRNA genes. Nonetheless, ppGpp is not involved in 6S RNA regulation of transcription during late stationary phase (Cavanagh *et al.*, 2010).

### 1.2.6 RNAP binding proteins

Apart from sigma factors that interact with RNAP and aid the initiation of transcription, other proteins have been discovered that affect the different stages of the transcription cycle. As mentioned in Section 1.2.4.2, NusA and NusG proteins interact with RNAP and aid transcription elongation. Additionally, the Gre factors, GreA and GreB, are other elongation factors that directly interact with RNAP and aid in the processivity of transcription (Section 1.2.4.2). Another protein in *E. coli* RapA, which is an ATPase, binds to core RNAP and promotes its recycling and this leads to transcription activation (Jin *et al.*, 2011b). An example of a phage-encoded inhibitor of RNAP is T4 AsiA. AsiA has been shown to interact with *E. coli* RNAP holoenzyme at the  $\beta$ -flap and  $\sigma^{70}$  region 4.2 thereby inhibiting transcription initiation (Yuan *et al.*, 2009). Other examples of RNAP binding proteins, DksA and CarD, are discussed below (Section 1.2.6.1 and 1.2.6.2).

Some antibiotics have also been isolated that interact with RNAP and inhibit its function. A key example is Rif that interacts with RNAP and inhibits transcription initiation (Section 1.1.2.3). Other examples of antibiotics that interact with RNAP are: Myxopyronin, which prevents RNAP from interacting with the promoter DNA (Mukhopadhyay *et al.*, 2008), Microcin J25 that prevents NTP uptake as it binds to the secondary channel of RNAP (Mukhopadhyay *et al.*, 2004) and Streptolydigin, which binds to the bridge helix of RNAP and affect RNAP function (Tuske *et al.*, 2005).

#### 1.2.6.1 DksA

In *E. coli*, DksA is a transcription factor that binds to RNAP. It belongs to a set of regulators that interact at the secondary channel of RNAP. It is mainly helical and is composed of two domains; a coiled-coil domain that extends into the secondary channel close to the active site where it can exert its regulation, and a globular domain that lies outside the channel (Perederina *et al.*, 2004, Furman *et al.*, 2011). DksA contains a Zn finger motif with the Zn (II) ion being chelated

by two cysteines in the globular domain and two cysteines in the coiled-coil domain. This Zn is essential for the functional role of DksA (Paul *et al.*, 2004, Blaby-Haas *et al.*, 2011). It is a stable protein with a half-life of 44 min (Chandrangsu *et al.*, 2011).

DksA co-operates with ppGpp and the initiating nucleotide (iNTP) to regulate gene expression during stress and nutrient starvation (Lee *et al.*, 2012). It is autoregulated and its level is kept relatively constant throughout growth *via* a negative feedback loop. Hence, when the level of DksA rises slightly, ppGpp and DksA inhibit transcription at the *dksA* promoter (Chandrangsu *et al.*, 2011).

One of the roles of DksA is to inhibit open complex formation by binding to closed complexes at *rrn* promoters and thus, decreasing rRNA synthesis (Rutherford *et al.*, 2009). It was shown that one of the two conserved aspartate residues located near the tip of the coiled-coil domain, D74, is involved in exerting the role of DksA on transcription initiation (Lee *et al.*, 2012). As DksA works with ppGpp, it can either repress transcription initiation from *rrn* promoters, or activate amino acid biosynthesis, or possibly exert no effect on some promoters (Paul *et al.*, 2005). ppGpp, together with DksA, have also been shown to reduce the stability of promoter-RNAP complexes. This seems to be dependent on the promoter sequence (Haugen *et al.*, 2006, Chandrangsu *et al.*, 2011).

Another role of DksA, that is independent of ppGpp, is its role in DNA replication. DNA replication is halted when the replication fork collapses due to encountering a paused elongation RNAP complex. Thus, during nutrient stress, many stalled elongation RNAP complexes are formed. DksA has been shown to be important in ensuring efficient DNA replication by either preventing RNAP pausing or destabilising the elongation RNAP complex off the DNA to allow elongation of the replication fork (Tehranchi *et al.*, 2010). The exact molecular mechanism is still unknown.

#### 1.2.6.2 CarD

Mycobacteria lack DksA, however, they undergo the stringent response using ppGpp. It was then discovered that CarD, a novel RNAP binding protein, regulates transcription from *rRNA* promoters (Stallings *et al.*, 2009). Unlike DksA, CarD is essential to maintain the viability of Mycobacteria including *M. tuberculosis* and *M. smegmatis*. It has important implications in the DNA damage response and regulating the stringent response (Stallings *et al.*, 2009).

CarD is composed of two regions: the N-terminal motif, which is similar to the RNAP interacting domain (RID) of transcription repair coupling factor (TRCF), binds to the N-terminus of the  $\beta$ -subunit of RNAP and the C-terminal region, which consists of a conserved leucine zipper motif and so far has an unidentified role (Stallings & Glickman, 2011). Nonetheless, it has been shown that CarD C-terminal region is vital for the survival of mycobacteria (Stallings & Glickman, 2011).

CarD is required to repress rRNA synthesis from the one or two *rRNA* promoters found in mycobacteria, and also inhibit the formation of ribosomal proteins (Stallings & Glickman, 2011). As CarD and DksA bind at different regions of RNAP, their mechanism of action is different. CarD was shown to be essential for the pathogenesis and persistence of *M. tuberculosis*. Therefore, a possible anti-tuberculosis drug that aims to bind to CarD and disrupt its interaction with RNAP may be an approach for the treatment of tuberculosis (Stallings *et al.*, 2009). The exact mechanism of action of CarD has not yet been determined.

#### **1.2.7 Anti-sigma factors**

Transcriptional regulation is further controlled by proteins that bind to sigma factors and alter their effect. Proteins that bind to sigma factors can have a negative or a positive effect on transcription. Proteins that exert a negative effect are known as anti-sigma factors (Campbell *et al.*, 2008). An example of an ECF

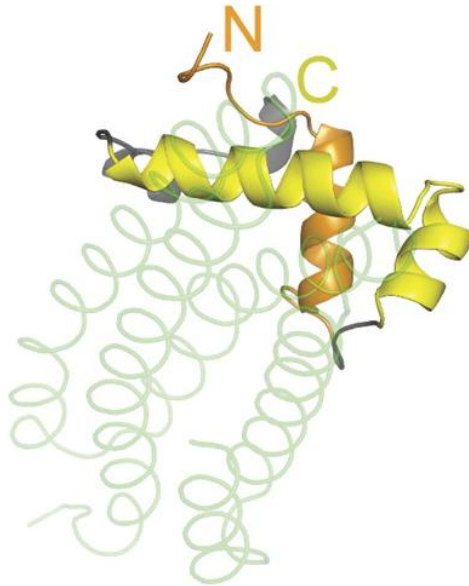
anti-sigma factor that specifically interacts with sigma is FlgM which targets  $\sigma^{28}$  in *Salmonella typhimurium*.

$\sigma^{28}$  is activated during flagellar assembly and activates *late* gene expression e.g. flagellin and motility genes (Hughes *et al.*, 1993). FlgM appears to sense the integrity of the flagellar assembly and if the flagellum is well assembled, FlgM is exported out of the cell by the flagellar-specific export apparatus, whereas, if there is a defect in the flagellar assembly, FlgM interacts with  $\sigma^{28}$  and inactivates it, serving as an anti-sigma factor (Hughes *et al.*, 1993). FlgM- $\sigma^{28}$  complex has a stoichiometry of 1:1 and FlgM makes multiple contacts with region 2.1, 3.1 and 4 (Chadsey & Hughes, 2001). FlgM was shown to interact with free  $\sigma^{28}$  and prevent the formation of the holoenzyme (RNAP- $\sigma^{28}$ ) (Ohnishi *et al.*, 1992). It was also shown to increase the rate of dissociation of RNAP- $\sigma^{28}$  leading to holoenzyme destabilisation (Chadsey *et al.*, 1998). Other examples of anti- $\sigma$  factors, Rsd, RsrA and Gin, are explained in detail below (Section 1.2.7.1-1.2.7.3).

#### 1.2.7.1 Rsd

Rsd is an anti-sigma factor of the *E. coli* principal sigma factor  $\sigma^{70}$  and is produced when the cells enter stationary phase (Jishage & Ishihama, 1998). Rsd was shown to interact with region 4 of  $\sigma^{70}$  (Fig. 1.7) (Patikoglou *et al.*, 2007) and region 2 (Yuan *et al.*, 2008) inhibiting  $\sigma^{70}$ -dependent transcription. The exact mechanism of action of Rsd is yet to be defined. Rsd is thought to interact with  $\sigma^{70}$  forming Rsd- $\sigma^{70}$  complex preventing  $\sigma^{70}$  association with core RNAP (Jishage & Ishihama, 1999). It also appears to inhibit transcription from promoters with/without an “extended -10 element” and causes the dissociation of  $\sigma^{70}$  from RNAP holoenzyme (Ilag *et al.*, 2004, Yuan *et al.*, 2008, Campbell *et al.*, 2008). It was also shown to be involved in sigma competition favouring  $\sigma^{70}$  switching to  $\sigma^S$  (Jishage & Ishihama, 1999, Hofmann *et al.*, 2011), *i.e* during stationary phase Rsd sequesters  $\sigma^{70}$  allowing  $\sigma^S$  to bind to core RNAP. Rsd has

also been shown to stimulate the expression of some  $\sigma^S$ -dependent genes (Mitchell *et al.*, 2007, Hofmann *et al.*, 2011).



**Figure 1.7: Structure of  $\sigma^{70}$  region 4 in complex with anti- $\sigma$  factor, Rsd.**

The green light ribbon represents Rsd and the thick yellow and orange ribbons represent the  $\sigma^{70(4)}$ . Copyright license No.: 2942540160750 (Campbell *et al.*, 2008).

#### 1.2.7.2 RsrA

RsrA, Regulator of SigR, is an anti-ECF  $\sigma$  factor that binds to  $\sigma^R$  and inhibits  $\sigma^R$ -dependent transcription in *S. coelicolor*. During oxidative stress, the cytoplasmic thiol-disulphide balance becomes more oxidising and this leads to the activation of  $\sigma^R$  and increased expression of *sigR*.  $\sigma^R$  binds to RNAP and stimulates the transcription of the thioredoxin (*trxBA*) operon (Paget *et al.*, 1998, Paget *et al.*, 2001a). Thioredoxin reduces the disulphide bonds in oxidised proteins, thereby restoring normal thiol levels, which ultimately switches off expression of the *trxBA* operon as well as *sigR* itself.

RsrA was the first reported member of a large family of anti-sigma factors, all of which are likely to require bound Zn for structure and function - the ZAS (Zn binding anti-sigma factor) family (Paget *et al.*, 2001a). Whereas many members of the ZAS family are large proteins, often with transmembrane domains, RsrA is only 105 amino acids and has a  $M_w$  of 11,680. Unlike RsrA, most anti-ECF  $\sigma$  factors are usually membrane-bound (Staroń *et al.*, 2009). RsrA is encoded by the *rsrA* gene, which is located directly downstream from *sigR* (Kang *et al.*, 1999).

The Zn (II) ion is co-ordinated by three cysteine residues, from the overall seven cysteines present in RsrA, together with a histidine residue. The three conserved residues from the ZAS-specific sequence motif (H37xxxC41xxC44) together with C11 co-ordinate the Zn (II) ion (Li *et al.*, 2003, Zdanowski *et al.*, 2006). RsrA forms up to three intra-molecular disulphide bonds when it is oxidised *in vitro* and this causes a conformational change in its structure and release of the Zn (II) ion. However, it is likely that a single disulphide bond between C11 and C44 is initially formed and that this represents the key redox switch (Li *et al.*, 2003, Zdanowski *et al.*, 2006).

During oxidative stress *in vivo*, RsrA oxidation allows  $\sigma^R$  to dissociate and bind to RNAP allowing it to activate transcription of  $\sigma^R$ -dependent genes (over 50 genes), many of which are involved in thiol-disulphide metabolism and the HspR heat-shock regulon, Clp and Lon ATP-dependent AAA(+) proteases (Kallifidas *et al.*, 2010). One of the target genes encodes thioredoxin, from the *txrBA* operon, that reduces the oxidised RsrA. RsrA in its reduced state can now bind  $\sigma^R$  inhibiting  $\sigma^R$ -dependent transcription. This homeostatic system thereby restores the normal balance of reduced thiols in the cytoplasm (Paget *et al.*, 2001a).

RsrA was shown to interact mainly with region 2.1-2.4 of  $\sigma^R$  in a 1:1 stoichiometry. The mechanism of inhibition of  $\sigma^R$  activity by RsrA is by blocking



$\sigma^R$  from associating with core RNAP as it binds to region 2.1 and 2.2, which make major interactions with core RNAP and also prevents DNA melting and the -10 element recognition as it binds to region 2.3 and 2.4 (Li *et al.*, 2002).

#### 1.2.7.3 Gin

In *Bacillus subtilis*, sporulation is regulated by four sigma factors ( $\sigma^F$ ,  $\sigma^G$ ,  $\sigma^E$  and  $\sigma^K$ ) that stimulate specific sets of genes at precise times throughout morphological differentiation (Hilbert & Piggot, 2004). During sporulation, the axial filament is formed leading to asymmetric division, whereby the cell division apparatus is relocalised to the polar site and consequently, the formation of the sporulation division septum. DNA is transferred from the mother cell to the forespore, followed by complete engulfment of the forespore, the deposition of cortex and protective layer of proteins around the forespore known as the coat. The forespore then matures followed by mother cell lysis releasing the forespore into the environment (Hilbert & Piggot, 2004).

After the asymmetric septum is formed,  $\sigma^F$  is activated in the forespore and this eventually leads to  $\sigma^E$  activation in the mother cell.  $\sigma^G$  and  $\sigma^K$  are later expressed by  $\sigma^F$  and  $\sigma^E$ , respectively.  $\sigma^F$  controls the initial development of the forespore whilst  $\sigma^G$  regulates genes involved in sporulation, germination, and in protecting the spore from DNA damage (Hilbert & Piggot, 2004, Serrano *et al.*, 2011).

There are three regulators of  $\sigma^G$ ; LonA protease, anti-sigma factor SpoIIAB, which controls both  $\sigma^G$  and  $\sigma^F$  activity, and Gin. Gin, also called CsfB, a product of the *csfB* gene, has been identified as the main anti-sigma factor for  $\sigma^G$ . It specifically binds to  $\sigma^G$  and is present early in the forespore compartment.  $\sigma^G$  expression is stimulated in the forespore towards the end of engulfment and it is thought that Gin inhibits its activity until after the complete forespore engulfment when  $\sigma^G$  becomes active (Serrano *et al.*, 2011).

Gin is a small protein and, together with its orthologues, has 44 to 74 amino acids. It contains six conserved residues; four cysteines (C11, C14, C30, C33), a glycine (G21) and a tyrosine (Y47) (Rhayat *et al.*, 2009). Gin is thought to exist as a dimer and to co-ordinate a Zn (II) ion. C11 and C14 on one polypeptide chain are thought to co-ordinate Zn with C30 and C33 on another polypeptide chain (Rhayat *et al.*, 2009).

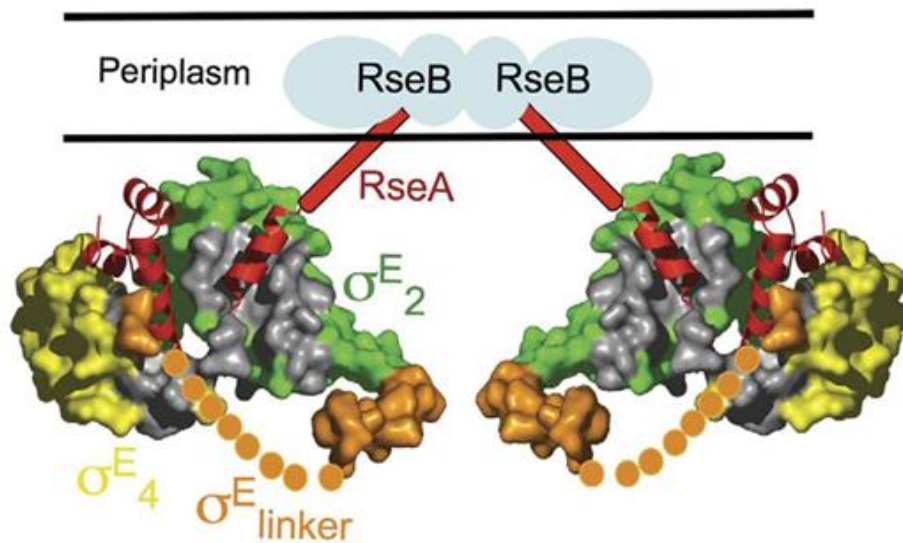
Mutation of five conserved residues of Gin: C11, C14, C30, C33 and Y47 restored  $\sigma^G$  activity. However, C14, C33 and Y47 mutants cannot interact with  $\sigma^G$  showing that these residues are important for  $\sigma^G$  interaction. A mutation in region 2.2 of  $\sigma^G$  substituting asparagine 45 to glutamic acid (N45E) inhibited binding of Gin to  $\sigma^G$  and therefore, increased the activity of  $\sigma^G$ . It was also observed that substitution of this single amino acid in  $\sigma^F$ , E39 to N, allowed the binding of Gin to  $\sigma^F$ . Hence, Gin loses its specificity by replacement of this single amino acid (Serrano *et al.*, 2011).

### 1.2.8 Proteins that bind to anti-sigma factors

Anti-sigma factors are regulated by proteins that bind to them and either inhibit their activity (known as anti-anti-sigma factors) or collaborate with the anti-sigma factors (known as co-anti-sigma factors). Another way in which anti-sigma factors are regulated is through proteolysis (Campbell *et al.*, 2008).

RsfA is an anti-sigma factor that interacts with  $\sigma^F$  in *S. coelicolor*. As mentioned earlier,  $\sigma^F$  is a Group III sigma factor that is involved in later stages of sporulation. It has been shown that there are two anti-anti-sigma factors, SCO0781 and SCO0869, which interact with RsfA (Kim *et al.*, 2008). SCO0869 has a serine residue in its HATPase\_c domain, which is phosphorylated by RsfA whilst SCO0781 lacks the serine residue. The exact mechanism of action is still under investigation (Kim *et al.*, 2008).

RseB is a co-anti-sigma factor of  $\sigma^E$  in *E. coli*.  $\sigma^E$  is involved in the cell-envelope response and is held inactive by RseA, the anti-sigma factor. RseB is a homodimer that is embedded in the periplasm and it secures the interaction of  $\sigma^E$ /RseA complex. It also inhibits RseA degradation by DegS and RseP (Wollmann & Zeth, 2007). Misfolded outer membrane proteins in the periplasm are thought to cause the release of RseB interaction with RseA and lead to RseA cleavage and hence  $\sigma^E$  is released (Campbell *et al.*, 2008, Kulp & Kuehn, 2011). The precise mechanism by which  $\sigma^E$  is released and the role RseB plays is still to be determined (Campbell *et al.*, 2008).



**Figure 1.8: Structure of  $\sigma^E$ /RseA complex.**

$\sigma^E$  region 2 (green),  $\sigma^E$  linker (orange),  $\sigma^E$  region 4 (yellow), regions that interact with RNAP (grey) are shown. RseB homodimer (blue) is found in the periplasm and binds to two separate RseA (red) to form a heterotetramer, 2RseB:2RseA. Only the cytoplasmic domain of RseA is shown. Copyright license No.: 2942540160750 (Campbell *et al.*, 2008).

### 1.2.9 Positive regulators of sigma factors

Most proteins that bind sigma factors have been shown to be negative regulators of transcription - anti-sigma factors. However, one protein, Crl, has

been shown to bind to sigma and positively regulate its activity (Section 1.2.9.1). The main subject of this project, RbpA, has also been shown to bind to the principal sigma factor and activate transcription of its genes in *S. coelicolor* (P. Doughty and M. Paget, personal communication; Section 1.2.9.2).

#### 1.2.9.1 Crl

Crl is a small global regulatory protein found in *E. coli* and *Salmonella* that has been shown to bind to the stationary phase sigma factor,  $\sigma^s$ , with a stoichiometry of 1:1.  $\sigma^s$  is encoded by *rpoS* and its transcription increases during exponential growth under particular stress conditions: low temperature, pH and high osmolarity (England *et al.*, 2008).  $\sigma^s$  associates with core RNAP and activates the transcription of over 300 genes some of which are responsible for the regulation of stationary phase-specific genes and stress response factors, formation of biofilm and stimulating the virulence of *S. enterica* serovar Typhimurium (Bang *et al.*, 2005, Weber *et al.*, 2005, Monteil *et al.*, 2010b). In *E. coli*,  $\sigma^s$  has the lowest affinity for core RNAP *in vitro* and has to compete with the six other sigma factors. This association is facilitated by Crl, which directly binds to  $\sigma^s$  and not to DNA (England *et al.*, 2008).

Crl appears to play a major role in sigma competition by increasing the affinity of  $\sigma^s$  for core RNAP seven-fold, leading to the formation of the holoenzyme, RNAP- $\sigma^s$  (Gaal *et al.*, 2006, Typas *et al.*, 2007, England *et al.*, 2008). Crl was also shown to stimulate to a lesser extent RNAP- $\sigma^{70}$  transcription (Gaal *et al.*, 2006). In addition, it was shown to affect the rate of open complex formation, which leads to activation of  $\sigma^s$ -dependent genes (Monteil *et al.*, 2010a, Monteil *et al.*, 2010b). Monteil *et al.*, determined, using a bacterial adenylate cyclase two-hybrid system that the Crl binds to domain 2 of  $\sigma^s$ . They also showed that four conserved residues of Crl, Y22, F53, W56 and W82, are important for maintaining the activity of Crl (Monteil *et al.*, 2010b).

*S. Typhi* Ty19 contains a naturally occurring mutant allele of *rpoS*, that encodes the  $\sigma^{S}_{Ty19}$  protein that is present in low amounts and has a very weak affinity for core RNAP compared to wild-type  $\sigma^S$ . Monteil, V *et al.*, also showed that Crl increases the activity of this  $\sigma^{S}_{Ty19}$  to wild-type levels. The mutation in *rpoS*<sub>Ty19</sub> is the substitution of glycine 282 to valine (G282V) and this is found in region 4 of  $\sigma^{S}_{Ty19}$ . Therefore, it was suggested that the interaction of Crl with region 2 of  $\sigma^{S}_{Ty19}$  would lead to a conformational change causing  $\sigma^{S}_{Ty19}$  to bind to core RNAP forming the holoenzyme, RNAP- $\sigma^{S}_{Ty19}$ , and activating  $\sigma^S$ -dependent genes (Monteil *et al.*, 2010b).

#### 1.2.9.2 RbpA

RbpA is a newly discovered RNAP-binding protein that was identified when it co-eluted with RNAP during gel filtration (Paget *et al.*, 2001b). Interest in *rbpA* was increased when it was discovered that  $\sigma^R$  induces its expression in response to oxidative stress, suggesting that it might be involved in the oxidative stress response (Section 1.2.7.2) (Paget *et al.*, 2001b). Further studies, however, suggested that RbpA plays a more general role in transcription.

*rbpA* mutants displayed a slow-growth phenotype, with a two-fold longer doubling time, although they appeared to be unaffected in morphological differentiation (Newell *et al.*, 2006). These mutants also showed more sensitivity to rifampicin, an antibiotic that inhibits transcription initiation by binding to a pocket in the  $\beta$  subunit of RNAP located deep within the DNA/RNA channel. Furthermore, mutants produced in excess, the blue pigmented antibiotic actinorhodin (Newell *et al.*, 2006). RbpA was also shown to be a metalloprotein, co-ordinating a Zn (II) ion (P. Doughty, and M. Paget, personal communication).

Unpublished *in vitro* studies suggest that RbpA is a positive effector of transcription and stimulates  $\sigma^{HrdB}$ -dependent transcription from rRNA promoters and all  $\sigma^{HrdB}$ -dependent promoters tested so far. Thus, it appears to be a general regulator of  $\sigma^{HrdB}$ -dependent transcription (P. Doughty and M. Paget,

unpublished observation). The slow growth rate of *rbpA* mutants might therefore be due to the reduced rate of rRNA synthesis (Newell *et al.*, 2006). RbpA appears to be involved in stimulating transcription initiation as it activates single-round transcription assays (P. Doughty, and M. Paget, personal communication). Preliminary work using potassium permanganate (KMnO<sub>4</sub>) assays have shown that RbpA stimulates promoter melting, suggesting a possible role of RbpA in open complex formation assays (P. Doughty, and M. Paget, personal communication). Although RbpA was found to co-elute with *S. coelicolor* RNAP, more recently it has been found that RbpA also interacts directly with  $\sigma^{\text{HrdB}}$  to form a complex independent of RNAP core assays (P. Doughty, and M. Paget, personal communication) (Section 4.1.1).

### **1.3 Project aims**

This project focuses on understanding the structural and functional role of RbpA, a RNAP binding protein that is present only in the Actinobacteria including the pathogenic bacterium *M. tuberculosis*. As mentioned in Section 1.2.9.2, RbpA was shown to be essential for the normal growth of *S. coelicolor* (Newell *et al.*, 2006). It appeared to stimulate promoter melting (P. Doughty and M. Paget, personal communication) suggesting a possible role of RbpA in open complex formation or an earlier role in transcription initiation. RbpA was also suggested to bind to  $\sigma^{\text{HrdB}}$ , the principal sigma factor of *S. coelicolor*, and activate  $\sigma^{\text{HrdB}}$ -dependent promoters (P. Doughty and M. Paget, personal communication). The aims of this thesis are outlined below:

- 1) To investigate the biological function of RbpA through the analysis of an  $\Delta rbpA$  strain.
- 2) To determine whether RbpA is involved in holoenzyme formation.
- 3) To test whether RbpA binds specifically to  $\sigma^{\text{HrdB}}$  or with a variety of sigma factors from groups II, III and IV.
- 4) To identify the region of  $\sigma^{\text{HrdB}}$  to which RbpA binds and *vice versa*.

- 5) To identify key residues that might be involved in the functional role of RbpA.
- 6) To obtain the structure of RbpA and its *M. tuberculosis* orthologue, Rv2050.

## **2- Materials and Methods**



## **2.1 Chemicals and reagents**

- Acrylamide solutions (Severn Biotech Ltd)
- Agarose powder (Melford Laboratory)
- Ammonium persulphate (Sigma)
- Ampicillin (Melford)
- Antifoam (Sigma)
- Apramycin (Duchefa Biochemie)
- Bicinchoninic acid (BCA) Solution (Sigma)
- Bradford Reagent (Sigma)
- Bromophenol blue (Amersham Biosciences)
- Casamino acids (Difco)
- Chloramphenicol (Melford)
- Chloroform (Fisher)
- Deoxyribonucleotide phosphates- dNTPs (New England Biolabs)
- Dimethylsulphoxide (DMSO) (BDH)
- Dithiothreitol (DTT) (Melford)
- Isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) (Melford)
- Kanamycin (Melford)
- Malt extract (Oxoid)
- Nalidixic acid (Duchefa Biochemie)
- N, N-dimethyl-formamide (Fisher)
- Nutrient agar (Difco)
- N-Tris(hydroxymethyl)methyl-2-aminoethane sulfonic acid (TES) (Fisher)
- Phenol (Fisher Scientific)
- Phenylmethylsulfonyl Fluoride (PMSF) (Sigma)
- Radiochemicals: [ $\alpha$ -<sup>32</sup>P] UTP 10  $\mu$ ci/ $\mu$ l, specific activity 800 G/mmol, 0.625  $\mu$ M
- Sigmacote<sup>®</sup> (Sigma)
- Sodium dodecyl sulphate (SDS) (Fisher)
- Spectinomycin (Duchefa Biochemie)

- Ultra Pure Sequagel® Urea Gel™ Complete (national diagnostics)
- Ultra Pure Sequagel® Urea Gel™-6 (national diagnostics)
- Tetramethyl-ethylenediamine (TEMED) (Fisher)
- N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid (TES) (Fisher Scientific)
- Trichloroacetic acid (TCA) (Sigma)
- 2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris) (Fisher Scientific)
- Tryptone (Difco)
- 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-gal) (Melford Laboratories Ltd)
- Yeast extract (Oxoid)

### **2.1.1 Enzymes**

#### 2.1.1.1 Polymerases

- Accuzyme (Bioline)
- Kod™ (Merck Bioscience)
- Phusion™ (New England Biolabs)
- Taq DNA polymerase (New England Biolabs)

#### 2.1.1.2 DNA modifying enzymes

- Shrimp alkaline phosphatase (Promega)
- T4 DNA Ligase (New England Biolabs)

#### 2.1.1.3 DNA/RNA restriction enzymes

- Restriction endonucleases (New England Biolabs)
- RQ RNase-free DNase (Promega)

### 2.1.2 Plasmids and expression vectors

Plasmids and expression vectors used in this project are shown in Table 1.

**Table 1: List of plasmids and expression vectors used in this project.**

Plasmids	Description	Reference
pBlueScript II SK (+)	<i>E. coli</i> cloning vector, <i>ori</i> f1, <i>bla</i> , Amp <sup>R</sup> , <i>ori</i> Co1E1, <i>lacZ</i> $\alpha$ (blue/white selection)	Stratagene (Alting-Mees & Short, 1989)
pBlueScript II SK (+):: <i>rbpA</i>	pBlueScript II SK (+) including <i>rbpA</i> gene cloned at EcoRI site in the multiple cloning site (MCS)	(P. Doughty and M. Paget, unpublished)
pBlueScript II SK (+):: <i>hrdB</i> <sup>(1.1-2)</sup> , <i>hrdB</i> <sup>(2)</sup> , <i>hrdB</i> <sup>(3-4)</sup> , <i>hrdB</i> <sup>(2-4)</sup> , <i>hrdB</i> <sup>(4)</sup>	pBlueScript II SK (+) including <i>hrdB</i> regions: domains 1.1-2, domain 2, domain 3-4, domain 2-4 and domain 4	(P. Doughty and M. Paget, unpublished)
pET15b	<i>E. coli</i> expression vector, N-terminal His-tag <sup>®</sup> sequence, Amp <sup>R</sup>	(Novagen limited)
pET20b	<i>E. coli</i> expression vector, C-terminal His-tag <sup>®</sup> sequence, Amp <sup>R</sup>	(Novagen limited)
pMT3000	<i>E. coli</i> cloning vector, pIJ2925 with modified MCS, <i>bla</i> , Amp <sup>R</sup> , <i>lacZ</i> $\alpha$ (blue/white selection)	(Paget <i>et al.</i> , 1994)
pMT3000:: <i>rbpA</i>	pMT3000 including <i>rbpA</i> gene with its promoter region (735bp) flanked by EcoRI & KpnI	(Newell., 2006)
pIJ6902	<i>PtipA</i> , attP <sup>oC31</sup> , <i>aac(3)IV</i> (apramycin-resistance gene), <i>tsr</i> (thiostrepton-resistance gene), <i>oriT</i> (derived from RK2 for <i>Streptomyces</i> )	(Huang <i>et al.</i> , 2005)
pSET $\Omega$	Shuttle plasmid, bifunctional: replicates in <i>E. coli</i> and integrates in <i>S. coelicolor</i> , derived from pSET152; <i>aac(3)IV</i> <sup>-</sup> , <i>aadA</i> <sup>+</sup> ,	(O'Connor <i>et al.</i> , 2002)

	attP <sup>oC31</sup> , oriT <sub>RK2</sub> ,	
pUT18	Derived from pUC19, Amp <sup>R</sup> , T18 fragment (225-399 aa) of CyaA under control of a <i>lac</i> promoter, MCS inserted at the 5' end of T18	(Karimova <i>et al.</i> , 1998)
pKT25	Derived from pSU40, Kan <sup>R</sup> , T25 fragment (1-224 aa) of CyaA, under control of a <i>lac</i> promoter, MCS inserted at the 3' end of T25	(Karimova <i>et al.</i> , 1998)

### 2.1.3 *S. coelicolor* and *E. coli* strains

#### 2.1.3.1 *E.coli*

JM109 and DH5 $\alpha$  strains were used for routine laboratory cloning and are suitable for blue/white screening (Table 2). BL21 ( $\lambda$ DE3) (pLysS) strain was used for protein over-expression experiments (Table 2). ET12567 containing pUZ8002 (ETZ) is a non-methylating *E. coli* strain (Table 2). It has three mutations on three DNA methyl transferase genes (*dam*<sup>-</sup>, *dcm*<sup>-</sup>, *hsdM*) hence, defective in methylation (MacNeil *et al.*, 1992). ETZ produces unmethylated plasmid or cosmid DNA therefore, efficient in conjugation (Paget *et al.*, 1999). BTH101 is an *E. coli* reporter strain that is deficient in adenylate cyclase (*cya*) (Table 2). It was used for the bacterial adenylate cyclase two-hybrid system (BACTH) to detect protein-protein interactions (Karimova *et al.*, 1998).

**Table 2: A list of *E. coli* strains used in this study**

<i>E. coli</i> Strain	Description	Reference
JM109	F <sup>+</sup> , <i>traD36</i> , <i>proA</i> <sup>+</sup> <i>B</i> <sup>+</sup> , <i>lacI</i> <sup>q</sup> , $\Delta(lacZ)M15/\Delta(lac-proAB)$ , <i>glnV44 e14</i> , <i>gyrA96</i> , <i>recA1</i> , <i>relA1</i> , <i>endA1</i> , <i>thi</i> , <i>hsdR17</i>	(Yanisch-Perron <i>et al.</i> , 1985)
DH5 $\alpha$ <sup>TM</sup>	F <sup>-</sup> , $\phi 80d lacZ \Delta M15$ , <i>endA1</i> , <i>recA1</i> , <i>hsdR17</i> ( <i>r</i> <sub>K</sub> <sup>-</sup> , <i>m</i> <sub>K</sub> <sup>+</sup> ), <i>supE44</i> , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i> , $\Delta(lacZYA-argF)U169$ , $\lambda^-$	Invitrogen Limited
BL21 ( $\lambda$ DE3/pLysS)	F <sup>-</sup> , <i>ompT</i> , <i>hsdS<sub>B</sub></i> ( <i>r</i> <sub>B</sub> <sup>-</sup> <i>m</i> <sub>B</sub> <sup>-</sup> ), <i>dcm</i> , <i>gal</i> , $\lambda$ (DE3) has chromosomal copy of T7 RNAP gene under control of <i>lacUV5</i> , <i>pLysS</i> is a plasmid which carries gene encoding T <sub>7</sub> lysozyme, Cm <sup>r</sup>	(Studier & Moffatt, 1986)
ET12567/pUZ8002	<i>dam-13::Tn9</i> , <i>dcm-6</i> , <i>hsdM</i> , Cm <sup>r</sup> , pUZ is a derivative of RK2 with a mutation in the <i>oriT</i> ( <i>aph</i> )	(MacNeil <i>et al.</i> , 1992, Paget <i>et al.</i> , 1999)
BTH101	F <sup>-</sup> , <i>cya-99</i> , <i>araD139</i> , <i>galE15</i> , <i>galK16</i> , <i>rpsL1</i> ( <i>Str</i> <sup>r</sup> ), <i>hsdR2</i> , <i>mcrA1</i> , <i>mcrB1</i>	(Karimova <i>et al.</i> , 1998)

#### 2.1.3.2 *S. coelicolor*

A list of the *S. coelicolor* A3 (2) strains used is shown in Table 3. M145 is plasmid free and is a prototrophic derivative of the wild-type strain A3 (2). J1981 was created from M145 and used as wild-type models through this study.

**Table 3: List of *S. coelicolor* strains used in this study**

<i>S. coelicolor</i> strains	Description	Reference
M145	SCP1 <sup>-</sup> SCP2 <sup>-</sup>	(Hopwood <i>et al.</i> , 1985)
J1981	M145 $\Delta rpoC::rpoC^{HIS}$	(Babcock <i>et al.</i> , 1997)
S129	J1981 ( $\Delta rbpA::aac(3)/V$ )	(Newell <i>et al.</i> , 2006)
J1915	M145 $\Delta glkA$ SCP1 <sup>-</sup> SCP2 <sup>-</sup>	(Kelemen <i>et al.</i> , 1995)
S101	J1915 ( $\Delta rbpA::aac(3)/V$ )	(Newell., 2006)

## 2.1.4 Growth media and selection

### 2.1.4.1 *E.coli* growth media

#### **Lennox broth (LB)**

This is a liquid media used for general growth of *E. coli* strains. Per litre: 10 g Difco Bacto-tryptone, 5 g Difco yeast extract, 5 g NaCl and 1 g glucose. 50 ml aliquots in universal bottles were sterilised by autoclaving. LB media was prepared with no glucose for BACTH assays.

#### **Lennox agar (LA)**

This was prepared as LB, but with the addition of 1.5 g of agar per 100 ml in 250 ml Erlenmeyer flasks. The flask was stoppered with foam bung, covered with foil and autoclaved.

#### **2xYT**

A rich liquid medium used in the conjugation procedure and heat shocking of *Streptomyces* spores. Per litre: 16 g Difco Bacto-tryptone, 10 g Difco Yeast Extract, and 5 g NaCl. 10 ml aliquots in universal bottles were sterilised by autoclaving.

#### **5x M63 minimal media**

This liquid media was used for growth of BTH101 cells for BACTH assays. Per litre: 10 g  $(\text{NH}_4)_2\text{SO}_4$ , 68 g  $\text{KH}_2\text{PO}_4$ , 2.5 mg  $\text{FeSO}_4 \cdot 6\text{H}_2\text{O}$ , and 5 mg Vitamin B1. The pH was adjusted to 7.0 with KOH and then sterilised by autoclaving.

#### **Optimized high-cell density IPTG-induction minimal media**

This liquid media was used to label proteins with  $^{13}\text{C}$  and  $^{15}\text{N}$  for Nuclear Magnetic Resonance (NMR) Spectroscopy. 50 mM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 25 mM  $\text{KH}_2\text{PO}_4$  pH 8-8.2, 10 mM NaCl, 5 mM  $\text{MgSO}_4$ , 0.2 mM  $\text{CaCl}_2$ , 0.25 x BME vitamins (Sigma), 0.25 x Trace Metals, 0.1%  $^{15}\text{NH}_4\text{Cl}$  and 1%  $^{13}\text{C}$ -Glucose.

1000 x Trace metal solution: 1 ml 0.1 M  $\text{FeCl}_3$  in 0.12 M HCl to 1 ml Trace metal mixture (50 mM  $\text{FeCl}_3$ , 20 mM  $\text{CaCl}_2$ , 10 mM  $\text{MnCl}_2$ , 10 mM  $\text{ZnSO}_4$ , and 2 mM of each  $\text{CoCl}_2$ ,  $\text{CuCl}_2$ ,  $\text{NiCl}_2$ ,  $\text{Na}_2\text{MoO}_4$ ,  $\text{Na}_2\text{SeO}_3$ , and  $\text{H}_3\text{BO}_3$  in 60 mM HCl). Stock solutions were made and all were autoclaved separately and kept at RT (Sivashanmugam *et al.*, 2009).

#### 2.1.4.2 *S. coelicolor* growth media

##### **NMMP**

NMMP is a liquid medium used for the dispersed growth of *S. coelicolor* to stationary phase. Per litre: 2 g  $(\text{NH}_4)_2\text{SO}_4$ , 5 g Difco Casamino acids, 0.6 g  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 50 g PEG 6000 and 1 ml NMMP Minor Elements. NMMP Minor Elements were made up: 1 g/L  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ , 1 g/L  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ , 1 g/L  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ , and 1 g/L anhydrous  $\text{CaCl}_2$  and kept at 4°C. 80 ml of NMMP media was aliquoted out into 100 ml bottles and autoclaved.

Prior to use, 15 ml 0.1 M  $\text{NaH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$ , pH 6.8 and 2.5 ml 20% (w/v) glucose was added to 80 ml NMMP media. 0.1 M  $\text{NaH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$  buffer was prepared by mixing 0.1 M  $\text{NaH}_2\text{PO}_4$  with 0.1 M  $\text{K}_2\text{HPO}_4$  and pH adjusted to 6.8.

##### **Mannitol soya flour medium (MS)**

This is a solid media used for general growth of *Streptomyces* and production of spores. Per litre: 20 g mannitol and 20 g soya flour. 100 ml MS was aliquoted into 250 ml Erlenmeyer flasks containing 1.5 g agar. These flasks were stoppered with foam bung, covered with foil and autoclaved. When it was ready to use, it was heated for 2-3 min to melt the agar.

##### **2 x Pre-germination medium (2 x PG)**

This is a liquid medium used for germination of *Streptomyces* spores before inoculation for growth. Per 100 ml: 1 g Difco yeast extract and 1 g Difco Casamino acids. 1 ml aliquots were made in 2 ml universal bottles and autoclaved.

### Yeast extract-malt extract medium (YEME)

This is a rich liquid medium used to grow *S. coelicolor* and obtain the growth curves of the different strains. Only 10% (w/v) sucrose was added for these experiments. Per litre: 3 g Difco yeast extract, 5 g Difco Bacto-peptone, 3 g Oxoid malt extract, 10 g glucose and 100 g sucrose. 200 ml aliquots were measured into 250 ml bottles and autoclaved. Prior to use: 0.4 ml 2.5 M  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$  and 2 ml of 1% (v/v) Antifoam was added.

#### 2.1.4.3 Antibiotics and additives

A plasmid with a known antibiotic resistance is selected for by addition of the appropriate level of antibiotics. Blue/White Screening is achieved by the addition of X-gal which gets broken down to a blue dye. Below is a table of the different antibiotics and additives used in this study.

**Table 4: A list of antibiotics and additives added to growth media for selection.**

Name	Stock concentration	Stock made up in	Concentration used in liquid media	Concentration used in solid media
Ampicillin	100 mg/ml	60% ethanol	100 $\mu\text{g/ml}$	100 $\mu\text{g/ml}$
Apramycin	50 mg/ml	dH <sub>2</sub> O (Filter-sterilised)	20-50 $\mu\text{g/ml}$	20 $\mu\text{g/ml}$
Choramphenicol	34 mg/ml	100% ethanol	34 $\mu\text{g/ml}$	25 $\mu\text{g/ml}$
IPTG	1 M	dH <sub>2</sub> O (Filter-sterilised)	1 mM	0.5 mM
Kanamycin	50 mg/ml	dH <sub>2</sub> O (Filter-sterilized)	50 $\mu\text{g/ml}$	25 $\mu\text{g/ml}$
Nalidixic Acid	25 mg/ml	0.15 M NaOH	25 $\mu\text{g/ml}$	25 $\mu\text{g/ml}$
Spectinomycin	50 mg/ml	dH <sub>2</sub> O (Filter-sterilised)	50 $\mu\text{g/ml}$	100 $\mu\text{g/ml}$
Thiostrepton	50 mg/ml	Dimethyl-sulphoxide	N/A	10-15 $\mu\text{g/ml}$
X-gal	40 mg/ml	N, N-dimethyl-formamide	N/A	40 $\mu\text{g/ml}$



#### 2.1.4.4 Growth and storage of *E. coli* and *S. coelicolor* strains

*E. coli* strains were grown at 37°C on solid or in liquid media. Colonies formed on solid media were kept for 1 week at 4°C before being discarded. For long-term storage, overnight LB cultures were mixed with an equal volume of 40% (v/v) glycerol then stored at -80°C.

*Streptomyces coelicolor* was grown at 30°C on solid or in liquid media. To prepare spores for long-term storage, confluent lawns were incubated for 4-5 days to allow sporulation, collected and suspended in 20% (v/v) glycerol (Section 2.2.5.2). Spores were stored at -20°C. Prior to growth in liquid media, spores were pre-germinated (Section 2.2.5.3). *S. coelicolor* was cultured in 60ml broth in 250 ml Erlenmeyer flasks. The flasks contained stainless steel springs (1.3 cm diameter coil, 19 sw gauge, marine grade) around the base of the flasks to improve aeration and prevent the formation of compact mycelial sediments.

### 2.1.5 Primers

#### 2.1.5.1 Bacterial two-hybrid

Primers designed for the bacterial two-hybrid system and used in this project are shown in Table 5.

**Table 5: List of primers designed for bacterial two-hybrid system analysis.**

Name		Primer sequence (5' to 3')	Chapter used in:
<i>T18_rbpA</i>	F- R-	GGGGATCCGATGAGTGAGCGAGCTCTTCG GGGGAATTCCCCGCACTCTTGCGGCTGTC	4.1.1
<i>T18_rbpA</i> <sub>ΔH1-AMP</sub>	F- R-	GGGAATTCCCCAGCCGCCAGC CTCAGGGCCGTCGCCGTCAAC	5.1.1
<i>T18_rbpA</i> <sub>ΔAMP</sub>	F- R-	GGGAATTCCCCAGCCGCCAGC GGTGCGTCGCTCCATCAGCATGTC	5.1.1
<i>T18_rbpA</i> <sub>Δβ</sub>	F- R-	GGGGATCCGGAGAAGAAGGCCAAGCCCG GGGGAATTCCCCGCACTCTTGCGGCTGTC	5.1.1
<i>T18_Rv2050</i>	F- R-	GGGGATCCGATGGCTGATCGTGTCCTGAG GGGGAATTCCCGCCGCGCCGACGTGACCGAATG	4.3.1
<i>T25_hrdB(2-4)</i>	F-	GGTCTAGAGACCGCCGACCCGGTCAAGGAC	4.1.1

	R-	GGG <b>GAATT</b> CCTAGTCGAGGTAGTCGCGCAG	
<i>T25_sigAMt</i>	F- R-	GGT <b>CTAG</b> AGGTGGCAGCGACCAAAGCAAG GGG <b>GAATT</b> CTCAGTCCAGGTAGTCGCG CAG	4.3.1
<i>T25_sigBMt</i>	F- R-	CCT <b>CTAG</b> AGGCCGATGCACCCACAAGGGCCA CC <b>GAATT</b> CCTGGCTCAGGATGTCCAGCT	4.3.1
<i>T25_hrdB(1.1-2)</i>	F- R-	GGT <b>CTAG</b> AGGTGTGCGGCCAGCACATCCCGTAC GGG <b>GAATT</b> CCTAGCGCGCCTGGTCGGCCATCGC	4.2.1
<i>T25_hrdB(2)</i>	F- R-	GGT <b>CTAG</b> AGACCGCCGACCCGGTCAAGGAC GGG <b>GAATT</b> CCTAGCGCGCCTGGTCGGCCATCGC	4.2.1
<i>T25_hrdB(3-4)</i>	F- R-	GGT <b>CTAG</b> AGAGCTTCACACTGCTGCAGGAGC GGG <b>GAATT</b> CCTAGTCGAGGTAGTCGCGCAG	4.2.1
<i>T25-hrdB(4)</i>	F- R-	GGT <b>CTAG</b> AGAGCTTCACACTGCTGCAGGAGC GGG <b>GAATT</b> CCTAGTCGAGGTAGTCGCGCAG	4.2.1
<i>T25_sigA(2)</i>	F- R-	GGT <b>CTAG</b> AGTCCGCCGACTCGGTTTCGCGCC GGG <b>GAATT</b> CTCAGCGGGCCTGGTCGGCCATGGCGC	4.3.2.1
<i>T25_hrdA(2-4)</i>	F- R-	GGT <b>CTAG</b> AGTCCTCCGACCTGTTCCGGCAG GGG <b>GAATT</b> CTCAGTCCAGGTAGCCCCTCAG	4.1.2
<i>T25_hrdC(2-4)</i>	F- R-	GGT <b>CTAG</b> AGGAACCCGACCTGCTCGGC GGG <b>GAATT</b> CTCAGCTCGCCCAGTCCAGC	4.1.2
<i>T25_hrdD(2-4)</i>	F- R-	GGT <b>CTAG</b> AGGACCGCGATCTGGTCGGCATG GGG <b>GAATT</b> CTCAGGCCGCGCCTCGAAGC	4.1.2
<i>T25_sigBSc</i>	F- R-	GGT <b>CTAG</b> AGATGACGACGACGACCGCGCGAGCCAC GGG <b>GAATT</b> CTCAGGTGGTGCTGAGCATGCCTTCC	4.1.3
<i>T25_sigR</i>	F- R-	GGT <b>CTAG</b> AGGTGGGTCCGGTCACTGGGAC GGG <b>GAATT</b> CTCATGACCCCGAGCCTTTTCG	4.1.3
<i>T25_sigE</i>	F- R-	GGT <b>CTAG</b> AGATGGGCGAGGTGCTCGAGTTTCG GGG <b>GAATT</b> CTCAGGCCGCGCAACGCTCC	4.1.3
<i>T25_whoG</i>	F- R-	GGT <b>CTAG</b> AGATGCCCCAGCACACCTCCG GGG <b>GAATT</b> CTCAGCGGCCGAAACCCGCAAG	4.1.3
<i>pUT18</i>		CAAGTCGATGCGTTTCGCGATC	5.3

The bold letters represent the restriction enzyme sites introduced. F- and R- represent forward and reverse primers, respectively.

### 2.1.5.2 Site-directed mutagenesis

Primers designed and used in this project for RbpA site-directed mutagenesis are shown in Table 6.

**Table 6: List of primers designed for *rbpA* site-directed mutagenesis.**

Name		Primer sequence (5' to 3')	Chapter used in:
<i>rbpA</i> <sub>Ndel</sub>	F- R-	<b>ATG</b> AGTGAGCGAGCTCTTCG <b>ATG</b> TCGTGCCTCCCGGGCTTG	5.2.1
<i>rbpA</i> <sub>R80A</sub>	F- R-	<b>GCC</b> ACGCACTGGGACATGCTG CGCGGGCTTGGCCTTCTTCTCC	5.2.2
<i>rbpA</i> <sub>T81A</sub>	F- R-	<b>GCCC</b> ACTGGGACATGCTGATGG ACGCGCGGGCTTGGCCTTCTTCTCC	5.2.2
<i>rbpA</i> <sub>H82A</sub>	F- R-	<b>GCCT</b> GGGACATGCTGATGGAGCG CGTACGCGCGGGCTTGGCCTTC	5.2.2
<i>rbpA</i> <sub>W83A</sub>	F- R-	<b>GCCG</b> ACATGCTGATGGAGCGACG GTGCGTACGCGCGGCTTGGCC	5.2.2
<i>rbpA</i> <sub>D84A</sub>	F- R-	<b>GCC</b> ATGCTGATGGAGCGACGC CCAGTGCGTACGCGCGGGGCTTG	5.2.2
<i>rbpA</i> <sub>M85A</sub>	F- R-	<b>GCCCT</b> GATGGAGCGACGCACC GTCCCAGTGCGTACGCGCGGGCTTG	5.2.2
<i>rbpA</i> <sub>L86A</sub>	F- R-	<b>GCC</b> ATGGAGCGACGCACCCGCGAG CATGTCCCAGTGCGTACGCGCGG	5.2.2
<i>rbpA</i> <sub>M87A</sub>	F- R-	<b>GCCG</b> AGCGACGCACCCGCGAGG CAGCATGTCCCAGTGCGTACG	5.2.2
<i>rbpA</i> <sub>E88A</sub>	F- R-	<b>GCCCG</b> ACGCACCCGCGAGGAAC CATCAGCATGTCCCAGTGCGTA	5.2.3
<i>rbpA</i> <sub>E94A</sub>	F- R-	<b>GCCCT</b> CGAAGAGGTCCTCGAGG CTCGCGGGTGCGTCGCTCCATC	5.2.4
<i>rbpA</i> <sub>E97A</sub>	F- R-	<b>GCCG</b> TCCTCGAGGAGCGGCTGG TTCGAGTTCCTCGCGGGTGCGT	5.2.4
<i>rbpA</i> <sub>E101A</sub>	F- R-	<b>GCCCG</b> GCTGGCCGTTCTGCGCT CTCGAGGACCTCTTCGAGTTCC	5.2.4
<i>rbpA</i> <sub>3EA</sub>	F- R-	<b>GCCG</b> TCCTCGAG <b>GCCCG</b> GCTGGCCGTTCTGCGCT TTCGAG <b>GGCCT</b> CGCGGGTGCGTCGCTCC	5.2.4
<i>rbpA</i> <sub>R89A</sub>	F- R-	<b>GCCCG</b> CACCCGCGAGGAACTCG CTCCATCAGCATGTCCCAGTGC	5.2.3
<i>rbpA</i> <sub>R90A</sub>	F- R-	<b>GCC</b> ACCCGCGAGGAACTCGAAG TCGCTCCATCAGCATGTCCCAG	5.2.3
<i>rbpA</i> <sub>2RA</sub>	F- R-	<b>GCC</b> ACCCGCGAGGAACTCGAAG <b>GGCCT</b> CCATCAGCATGTCCCAG	5.2.3

<i>rbpA</i> <sub>R102A</sub>	F- R-	<b>GCC</b> CTGGCCGTTCTGCGCTCCG CTCCTCGAGGACCTCTTCGAG	5.2.4
<i>rbpA</i> <sub>R107A</sub>	F- R-	<b>GCCT</b> CCGGCGCGATGAACATCG CAGAACGGCCAGCCGCTCCTCGAG	5.2.4

NdeI site (CATATG) was introduced into *rbpA* by using *rbpA\_NdeI* primer which introduced CAT upstream of ATG using the reverse primer. The bold letters represent the alanine mutations introduced. F- and R- represent forward and reverse primers, respectively.

#### 2.1.5.3 Protein expression

Primers used in this project for protein expression are shown in Table 7.

**Table 7: List of primers designed for protein over-expression.**

Name		Primer sequence (5' to 3')	Chapter used in:
<i>Rv2050</i>	F- R-	GG <b>CATATG</b> GCTGATCGTGTCCTGAGG CC <b>GGATCCC</b> GGGTCAGCCGCGCCGACGTG	4.3.1
<i>sigA(2)</i>	F- R-	GG <b>CATATG</b> TCCGCCGACTCGGTTCGCGCC CC <b>GGATCCC</b> CATCAGCGGGCCTGGTCGGCCATGGC	4.3.2.1

Bold letters represent restriction sites introduced at the N-terminal and C-terminal of the genes. F- and R- represent forward and reverse primers, respectively.

#### 2.1.5.4 *In vitro* transcription

Primers used in this project to obtain templates used for *in vitro* transcription assays are shown in Table 8.

**Table 8: Primers used to amplify genes used as a template for *in vitro* transcription assays.**

Name		Primer sequence (5' to 3')	Chapters used in:
<i>sco4652</i>	F- R-	GGAATTCGCGCCCGGCCCGCTCCGGTCGCCG GGAAGCTTGTCTCTTTTGAACACACGGCAACG	5.6
<i>sco0527</i>	F- R-	GGAATTCGTCGGCAACAAGGTTCCCGCTCAC GGAAGCTTCCCGCACTCCTGGGCGGGCGGTCT	5.6
<i>rpoB</i>	F- R-	GGAATTCCAACGAGGAGCGAACACGGTCCCCGA GGAAGCTTGTAGACCCCTGGTGACGGGCAGGG	5.6
<i>M13/pUC</i> (-47)	F-	CGCCAGGGTTTTCCCAGTCACGAC	(Chapter 5)
<i>M13/pUC</i> (-48)	R-	AGCGGATAACAATTCACACAGGA	(Chapter 5)

F- and R- represent forward and reverse primers, respectively.

## 2.1.6 Solutions/buffers

### 2.1.6.1 DNA manipulation

#### Buffers for basic cloning

- TE buffer: 1 mM EDTA, 10 mM Tris-HCl pH 8.0, autoclaved.
- 6 x DNA loading dye: 25 mg Bromophenol blue and 30% (v/v) glycerol in 10 ml and filter-sterilised.
- 10 x TBE buffer: 107.8 g Tris base, ~ 55 g boric acid pH 8.3 and 7.44 g EDTA in 1 L.
- Gel Red Stain: 1 x TBE and 10 µl GelRed™ (Cambridge Bioscience). Stored in the dark.

#### Buffers for preparation of competent cells

- Tfb1 buffer: 100 mM RbCl<sub>2</sub>, 50 mM MnCl<sub>2</sub>·4H<sub>2</sub>O, 30 mM potassium acetate, 10 mM CaCl<sub>2</sub>·2H<sub>2</sub>O, 15% (v/v) glycerol. Filter-sterilised, stored at 4°C and protected from light.
- Tfb2 buffer: 10 mM MOPS, 10 mM RbCl<sub>2</sub>, 10 mM CaCl<sub>2</sub>·2H<sub>2</sub>O, 15% (v/v) glycerol. Filter-sterilised, stored at 4°C and protected from light.

### **Plasmid purification by alkaline lysis buffers**

- Solution I: 50 mM Tris-HCl, pH 8; 10 mM EDTA. Stored at RT.
- Solution II: 200 mM NaOH; 1% (v/v) SDS, prior to use add 500  $\mu$ l SDS per 10 ml.
- Solution III: 3 M potassium acetate, pH 5.5. Stored at 4°C.
- Phenol/Chloroform: 20 ml phenol added to 20 ml of chloroform and stored at 4°C away from light.

#### 2.1.6.2 Protein purification

##### **Ni-column affinity purification buffers for soluble proteins**

- Charge buffer: 50 mM  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ .
- Binding buffer: 5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9.
- Wash buffer: 60 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9.
- Elution buffer: 1 M imidazole, 0.5 M NaCl, 20 mM Tris pH-HCl 7.9.
- Strip buffer: 100 mM EDTA, 0.5 M NaCl, 20 mM Tris-NaOH pH 7.9.

All were made with  $\text{dH}_2\text{O}$ , filter-sterilised and stored at RT.

##### **Ni-column affinity purification buffers for insoluble proteins**

- Binding buffer: 5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9, 6 M urea.
- Wash buffer: 60 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9, 6 M urea.
- Elution buffer: 1 M imidazole, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9, 6 M urea.
- Urea Renaturation buffer: 50 mM Tris-NaOH pH 7.9, 1 mM EDTA, 1 mM DTT, 20% (v/v) glycerol, 0.05% (v/v) Triton X-100, 0.5 M proline and urea ranging from 4 M to 0 M.

All were made up with  $\text{dH}_2\text{O}$ , filter-sterilised and stored at RT.

##### **Anion-exchange purification buffers**

- Binding buffer: 50 mM Tris-HCl pH 7.9, 50 mM NaCl, 5% (v/v) glycerol and 5 mM  $\beta$ -mercaptoethanol.

- Anion-exchange Elution buffer: 50 mM Tris-HCl pH 7.9, 1 M NaCl, 5% (v/v) glycerol and 5 mM  $\beta$ -mercaptoethanol.

They were made up with dH<sub>2</sub>O, filter-sterilised and stored at RT.

#### **Gel filtration and NMR buffer**

- Gel filtration buffer: 20 to 50 mM Tris-HCl pH 7.9, 50 to 200 mM NaCl, 5% (v/v) glycerol and 5 mM  $\beta$ -mercaptoethanol and filter-sterilised.
- NMR buffer: 25 mM phosphate buffer pH 6.8, 50 mM NaCl and 1 mM DTT.

#### **Other protein purification buffers**

- Coomassie Brilliant Blue stain solution: 0.25% (w/v) Coomassie Brilliant Blue<sup>TM</sup> R-250 (Sigma), 50% (v/v) methanol, 10% (v/v) glacial acetic acid.
- Destainer solution: 20% (v/v) methanol, 10% (v/v) glacial acetic acid.
- 2 x Protein loading dye with DTT: 250 mM Tris-NaOH pH 6.8, 2% (v/v) SDS, 10% (v/v) glycerol, 0.01% (w/v) Bromophenol Blue and 20 mM DTT.
- Phenylmethylsulfonyl fluoride (PMSF): 10 mg/ml in isopropanol and stored at -20°C.
- 10 x SDS polyacrylamide gel running buffer: 100 ml 10% (v/v) SDS, 30.9 g Tris-HCl, pH 8.3, 144.1 g glycine in 1 L.
- 10 x Thrombin cleavage buffer: 500 mM Tris-HCl pH 8.0, 1.5 M NaCl and 25 mM CaCl<sub>2</sub>.

#### **2.1.6.3 Western blotting**

- Transfer buffer: 14.4 g glycine, 3 g Tris-HCl, 200 ml methanol, 2 ml 10% (v/v) SDS and 800 ml dH<sub>2</sub>O.
- 10 x TBS-Tween buffer: 12 g Tris-HCl, pH 7.6, 40 g NaCl and 5 ml Tween 20<sup>TM</sup> in 500 ml.
- Blocking solution: 5% (w/v) skimmed milk in 1 x TBS-Tween, made fresh before use.

#### 2.1.6.4 *In vitro* transcription assays

- 10 x transcription buffer: 0.4 M Tris-HCl pH 7.9, 0.1 M MgCl<sub>2</sub> and autoclaved. Prior to use add: 50 mM  $\beta$ -Mercaptoethanol and 0.1 mg/ml BSA (Hsu, 2009).
- Formamide loading buffer: 80% (w/v) deionized formamide, 1 x TBE buffer, 10 mM EDTA, 0.08% (w/v) xylene cyanol, 0.08% (w/v) Amaranth. The deionized formamide was divided into 0.8 ml aliquots and stored at -20°C.
- RNAP dilution buffer: 10 mM Tris-HCl, pH 8.0, 10 mM KCl, 5% (v/v) glycerol, 0.4 mg/ml BSA, 0.1% Triton X-100 and 10 mM  $\beta$ -mercaptoethanol (added prior to use).
- 10 x NTP mix: 100  $\mu$ M ATP, 100  $\mu$ M GTP, 100  $\mu$ M CTP, 10  $\mu$ M of UTP and 50  $\mu$ M  $\gamma$ UTP in 15  $\mu$ l.

#### 2.1.6.5 $\beta$ -galactosidase assay

- PM2 medium: 70 mM Na<sub>2</sub>HPO<sub>4</sub>·12H<sub>2</sub>O, 30 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM Mg<sub>2</sub>SO<sub>4</sub>, 0.2 mM MnSO<sub>4</sub>, pH 7.0, autoclaved and 100 mM  $\beta$ -mercaptoethanol (added prior to use)
- 1 M Na<sub>2</sub>CO<sub>3</sub>: autoclaved.
- O-nitrophenol- $\beta$ -galactosidase (ONPG): 4 mg/ml in PM2 medium without  $\beta$ -mercaptoethanol. Stored at -20°C (equilibrated to 30°C prior to use).

#### 2.1.6.6 Others

- TES buffer: 0.05 M TES, pH 8.0 and autoclaved
- Kirby mix (Kieser *et al.*, 2000): 1% (w/v) sodium-triisopropyl naphthalene sulphonate, 6% (w/v) sodium 4-amino salicylate, 6% (v/v) phenol made up in 50 mM Tris-HCl pH 8.3.



## **2.2 DNA manipulation and cloning**

### **2.2.1 DNA manipulation**

#### **2.2.1.1 General Polymerase chain reaction (PCR)**

Basic PCR reactions were set up in 0.2 ml PCR tubes as follows:

- 20-50 pmol/μl forward and reverse primers,
- 200 μM dNTPs,
- 1 x reaction buffer,
- 2 ng/μl plasmid DNA or 25 ng/μl chromosomal DNA,
- Presence of either 5% (w/v) glycerol, 5% (v/v) DMSO or 5x Hi-Specificity Additive (Bioline) and made up to 50 μl with autoclaved dH<sub>2</sub>O

The tubes were placed in the thermal cycler at 97°C for 5 min, held at 80°C and then DNA polymerase was added. The PCR cycle:

- Melting at 96°C for 1 min
  - Annealing at 55-65°C for 45 s
  - Extension at 72°C for 0.5-1 min per 1 kb
  - Final Extension at 72°C for 5 min
  - Held at 4°C
- } x 28-30

The PCR cycle was altered by increasing or decreasing the annealing temperature depending on the GC content of the primers. High GC content requires an increase in annealing temperature and *vice versa*.

#### **2.2.1.2 Inverse PCR**

Primers were phosphorylated by combining 200 pmol oligonucleotides, 1 x T4 DNA kinase buffer, 2 mM ATP and 0.005 units T4 DNA kinase. The reaction was left at 37°C for 30 min and then heated at 90°C for 2 min to inactivate the enzyme. The primers were used to set up a PCR reaction as stated in Section 2.2.1.1, using Accuzyme (Bioline) as the thermostable DNA polymerase.

For large products (>3kb), the following PCR cycle was used:

- Melting at 96°C for 1 min
  - Annealing at 55-65°C for 45 s
  - Extension at 72°C for 2.5 min per 1 kb
  - Melting at 96°C for 1 min
  - Annealing at 55-65°C for 45 s
  - Extension at 72°C for 3.5 min per 1 kb
  - Final Extension at 72°C for 15 min
  - Held at 4°C
- } x 10  
 } x 10

The PCR products were then separated on a 0.8 to 1% of agarose gel and the gene of interest was purified using QIAquick™ PCR Purification Kit (Qiagen) as per manufacturer's protocol.

#### 2.2.1.3 Restriction digestion

DNA digestion was performed by adding ~1 µg DNA, 1 x restriction buffer and ~10 units of each enzyme. Digestion mixture was left at 37°C water bath for 2-3 h.

#### 2.2.1.4 DNA gel purification

The gel extraction and purification was performed using QIAquick™ or MinElute™ Gel Extraction Kit for larger and smaller fragments, respectively, according to manufacturer's instruction (Qiagen).

#### 2.2.1.5 Dephosphorylation of DNA

Digested vectors were inhibited from self-ligation by dephosphorylation. 0.5 units shrimp alkaline phosphatase (SAP) was added to the digested vector and the reaction mixture left at 37°C water bath for 0.5-1 h. The enzyme was inactivated by heating at 65°C for 20 min.

#### 2.2.1.6 DNA ligation

Ligations were prepared in a final volume of 15  $\mu$ l using an insert DNA to vector DNA ratio of 4:1 together with 1 x T4 DNA ligase buffer, 0.2 units of T4 DNA Ligase (New England Biolabs), and incubated at 4°C overnight.

### **2.2.2 *E. coli* DNA transformation**

#### 2.2.2.1 Heat shock

##### **Preparation by $\text{CaCl}_2$ method**

An overnight culture was used to inoculate 50 ml LB (1:100) and grown to an  $\text{OD}_{600}$  of 0.4-0.5. The cells were harvested by centrifugation at 2,000 x g for 10 min and the pellet re-suspended in 20 ml ice-cold 0.1 M  $\text{MgCl}_2$  10% glycerol. The cell suspension was centrifuged at 2,000 x g for 10 min and the pellet re-suspended in 10 ml ice-cold 0.1 M  $\text{CaCl}_2$  10% glycerol. The solution was incubated on ice for 30 min before being centrifuged for a further 10 min at 2,000 x g. The cell pellet was re-suspended in the residual volume, approximately 3 ml, and 150  $\mu$ l aliquots were frozen in dry ice and stored at -80°C.

##### **Preparation by $\text{RbCl}_2$ method**

2 x 50 ml of selective LB were inoculated with 500 ml of overnight culture and grown to an  $\text{OD}_{600}$  of 0.5. The cells were then chilled on ice for 15 min. They were harvested by centrifugation at 2,000 x g for 5 min and the pellet re-suspended in 6 ml Tfb1 buffer (Section 2.1.6.1). The 2 x 50 ml cell cultures were combined and chilled on ice for 15 min. The 12 ml cells were then centrifuged at 2,000 x g for 5 min. The pellet was re-suspended in 2 ml TfbII buffer (Section 2.1.6.1) and kept on ice. The cells were aliquoted into 100  $\mu$ l fractions and frozen on dry ice ethanol bath. They were stored at -80°C.

**Transformation procedure**

The ligation mixture or plasmid of interest was added to competent cells, placed in a 42°C water bath for 60 s and immediately after, 900 µl of LB was added. The cells were recovered at 37°C at 250 rpm for 0.5-1 h. They were then plated on LA plates containing appropriate selection. The plates were left at 37°C overnight. The next day the colonies were selected and checked for recombinants or colonies grown for protein over-expression.

**2.2.2.2 Electroporation****Preparation of electro-competent cells**

50 ml of selective LB were inoculated with 500 µl of overnight culture and grown to an OD<sub>600</sub> of 0.4. The cells were harvested by centrifugation at 2,000 x g for 10 min and the pellet re-suspended with 50 ml of 10% chilled glycerol. This was centrifuged again at 2,000 x g for 10 min and pellet re-suspended with 25ml of 10% chilled glycerol. This was centrifuged again as above, and re-suspended with 10 ml of 10% chilled glycerol. Centrifugation was again performed as above, pellet was re-suspended in the residual volume, ~1.5 ml. 50 µl aliquots were frozen on dry ice and stored at -80°C.

**Electroporation procedure**

Electroporation was done in an Eppendorf Electroporator 2510 system, with a voltage of 2,500 V, 400 Ω and 330 µF. To 50 µl of competent cells approximately 2 ng of the plasmids were added and were immediately transferred to an ice-cold electroporation cuvette, immediately shocked, and 900 µl of LB was added. This was left for 1 h at 37°C at 250 rpm for the cells to recover. These cells were plated on LA plates with appropriate selection and left to grow overnight at 30°C.

### **2.2.3 DNA extraction and purification**

#### 2.2.3.1 Small scale Wizard® Miniprep plasmid purification

The plasmid purification was performed as per the Promega kit and following the manufacturer's instructions for "Wizard® Plus SV Minipreps DNA purification systems". The plasmid was eluted in a total volume of 100 µl using TE buffer.

#### 2.2.3.2 Small scale plasmid purification by alkaline lysis

3 ml of overnight inoculants were harvested by centrifugation at 16,100 x g. The cell pellet was resuspended into 200 µl Solution I (Section 2.1.6.1). 400 µl Solution II (Section 2.1.6.1) was added, mixed by inverting, followed by 300 µl Solution III (Section 2.1.6.1). 1 µl 10 mg/ml RNase was added, incubated at RT for 10 min and centrifuged at 16,100 x g for 5 min. The supernatant was extracted, added to 150 µl phenol/chloroform, vortex mixed for 2 min and centrifuged again as above. The upper phase was transferred to 600 µl isopropanol, chilled on ice for 10 min and centrifuged as above. The pellet was washed with 200 µl 70% ethanol, centrifuged again at 16,100 x g for 1 min and all the supernatant was removed. The pellet was resuspended in 30 µl TE buffer.

#### 2.2.3.3 Large scale Midiprep plasmid purification (Qiagen)

For large scale purifications, 50 ml of culture was extracted by the Qiagen Plasmid Midi kit and the protocol followed as per manufacturer's instructions. The pellet was finally resuspended in 100 µl of TE buffer and stored at -20°C.

### **2.2.4 RNA extraction and purification**

#### 2.2.4.1 Small scale RNA miniprep

30 ml of *S. coelicolor* culture was centrifuged at 2,000 x g for 1 min. The pellet was resuspended in 1 ml Kirby mix solution (Section 2.1.6.6) and sonicated twice for 3 s. 700 µl phenol /chloroform was added, vortex mixed for 2 min and

centrifuged at 16,100 x g for 5 min at 4°C. The upper phase was extracted with 700 µl phenol/chloroform and method repeated (see above). The upper phase was then added to 90 µl sodium acetate, 900 µl isopropanol, mixed by inverting and stored at -20°C for 1 h. The suspension was then centrifuged at 16,100 x g for 10 min at 4°C, pellet washed with 180 µl 70% ethanol, and then resuspended in 200 µl 1 x DNase buffer. It was placed on ice for 5 min, 0.5 µl DNase was added and incubated at 37°C for 1 h. 200 µl phenol/chloroform was added, vortex mixed for 2 min, centrifuged at 16,100 x g for 5 min at 4°C. The upper phase was again extracted, 40 µl sodium acetate and 400 µl isopropanol were added and stored at -20°C for 1 h. This was finally centrifuged at 16,100 rpm for 10 min at 4°C and pellet washed with 100 µl 70% ethanol. The pellet was resuspended in 100 µl RNase-free water. The RNA was cleaned by using the RNeasy mini Kit (Qiagen) and the manufacturer's instructions were followed.

## 2.2.5 Microbiological methods

### 2.2.5.1 Conjugation of DNA from *E. coli* to *S. coelicolor*

Transformed ETZ cells were used to inoculate LB containing 25 µg/µl Chloramphenicol and Kanamycin and appropriate selection for the incoming plasmid. This was grown overnight at 37°C with vigorous shaking. The overnight culture was diluted 1:100 in fresh LB with antibiotic selection and grown at 37°C to an OD<sub>600</sub> of 0.4-0.6. The cells were washed twice with an equal volume of chilled LB and resuspended in 0.1 volume of LB. 10<sup>8</sup> *Streptomyces* spores were added to 500 µl of 2xYT (Section 2.1.4.2) and heat shocked at 50°C for 10 min. 500 µl of *E. coli* cells together with 500 µl of heat-shocked spores were mixed together and briefly centrifuged. Most of the supernatant was removed and the pellet was resuspended in residual volume. These were plated on MS agar plates with 10 mM MgCl<sub>2</sub> and incubated at 30°C for 16-20 h. Overlay was performed by adding 1 ml of sterile dH<sub>2</sub>O containing 0.5 mg of nalidixic acid with the appropriate plasmid antibiotic selection. The plates were incubated for another 4-5 days at 30°C.

#### 2.2.5.2 Harvesting *S. coelicolor* spores

A single colony of *S. coelicolor* was chosen and streaked into a spore plate. The spore plate was incubated at 30°C for 5-7 days allowing the mycelium to sporulate forming a uniform grey surface. 10 ml sterile dH<sub>2</sub>O was added then spores were scrapped off using a loop. The spore content was poured into a 15 ml centrifuge tube and vortex mixed for 2 min. The mixture was filtered from the agar suspension by adding it to an autoclaved syringe stuffed with cotton wool and collecting the filtered suspension. This was centrifuged at 890 x g for 5 min. The spore pellet was resuspended in 1 ml 20% glycerol and stored at -20°C.

#### 2.2.5.3 Determining the growth curve of *S. coelicolor* in liquid media

Spores were added to 1 ml TES buffer (Section 2.1.6.6) to give a starting OD<sub>450</sub> of 0.05-0.06 in 50-60 ml liquid media. The spores were centrifuged briefly and spore pellet resuspended in 1 ml TES buffer. They were heat shocked at 50°C for 10 min, added to 1 ml 2 x PG media (Section 2.1.4.2) and 2 µl 5 M CaCl<sub>2</sub>. This was transferred into 20 ml glass universal with springs and incubated at 37°C with vigorous shaking for 3-5 h to pre-germinate. The cell broth was then transferred into 2 ml Eppendorf tube, centrifuged briefly, the pellet resuspended in 1 ml *S. coelicolor* liquid media broth YEME (Section 2.1.4.2) and left to grow at 30°C with shaking at 200 rpm. When the cultures reach early exponential phase, 0.25-1 ml of samples were taken at 2 h interval and absorbance at OD<sub>450</sub> was measured.

### **2.3 Protein expression and purification**

#### **2.3.1 Protein expression**

##### 2.3.1.1 Unlabelled proteins

BL21 (pLysS) transformants were inoculated in LB with 100 µg/µl Ampicillin and 34 µg/µl Chloramphenicol and left overnight at 37°C with vigorous shaking. Using the overnight inoculants, 20-500 ml culture was grown in the presence of

Ampicillin and Chloramphenicol at 37°C with vigorous shaking to an OD<sub>600</sub> of 0.4-0.7. The cultures were cold-shocked on ice for 15 min and protein expression induced with 1 mM IPTG for 3 h at 30-37°C with vigorous shaking. The cells were then harvested and the cell pellet stored at -80°C until required for protein purification.

#### 2.3.1.2 <sup>13</sup>C <sup>15</sup>N labelling of protein

BL21 (pLysS) transformants were grown in LB in the presence of 100 µg/µl Ampicillin and 34 µg/µl Chloramphenicol to an OD<sub>600</sub> of 3-5. The cells were centrifuged at 1500 x g for 5 min and the cell pellet resuspended in equal volume of optimized high-cell density IPTG-induction minimal media (Section 2.1.4.1). The culture was grown in minimal media for 1.5-2 h at 37°C with vigorous shaking, induced with 1 mM IPTG at 30°C for 3 h. The culture was harvested and the cell pellet stored at -80°C until required for protein purification (Studier, 2005, Sivashanmugam *et al.*, 2009)

### **2.3.2 Protein purification**

#### 2.3.2.1 Preparing cell lysate

The cell pellet was resuspended in Binding buffer (composition dependent on the purification method - Section 2.1.6.2), 25 µg/µl PMSF and 1 protease inhibitor cocktail tablet (Complete mini EDTA-free, Roche) were added. The cell suspension was sonicated 6 times for 10 s with 50 s intervals. The cell suspension was centrifuged at 20,000 x g for 10 min at 4°C. If the protein was soluble, the cleared cell lysate (supernatant) was loaded on the column for purification. If the protein was however insoluble, the whole cell lysate pellet was resuspended in urea binding buffer, sonicated twice for 5 s with 50 s intervals and centrifuged at 20,000 x g for 10 min at 4°C. The cleared cell lysate was then loaded on the column for insoluble protein purification in the presence of urea (Dombroski, 1996).



#### 2.3.2.2 Ni-NTA sepharose hand-made column

Iminodiacetic acid (IDA) Sepharose fast-flow resin was added into a syringe barrel (sealed using glass wool) to give a column volume (CV) of 1-4 ml. The buffers were pumped into the column using a peristaltic pump. The resins were washed with 3 CV of dH<sub>2</sub>O followed by 5 CV of Charge buffer (Section 2.1.6.2) and 3 CV of dH<sub>2</sub>O. The column was equilibrated with 5 CV of Binding buffer (Section 2.1.6.2) prior to the addition of the cleared cell lysate. The column was then washed with 10 CV of Binding buffer, 5 CV of Wash buffer and 1-2 CV of Elution buffer (Section 2.1.6.2). Samples collected from the flow-through and elution were analysed by SDS polyacrylamide gel (Section 2.3.6). The column was stored by washing with 6 CV of Strip buffer (Section 2.1.6.2) followed by 3 CV of dH<sub>2</sub>O before being stored with 20% ethanol.

#### 2.3.2.3 Ni-NTA spin column

Ni-NTA spin column (Qiagen) was used for fast and small scale His-tag protein purification. Refer to the Manufacturer's Ni-NTA Spin handbook for a detailed protocol on Protein Purification under Native Conditions. The Lysis buffer, Wash buffer and Elution buffers were all made up as per manufacturer's instruction.

#### 2.3.2.4 Dynabeads<sup>®</sup> His-tag isolation and pull-down

Dynabeads<sup>®</sup> (Invitrogen) contain immobilised cobalt ions which allows His-tagged proteins to bind to them with a high affinity. 2 x Binding/Wash buffer and His Elution buffer were made as per manufacturer's instructions (Invitrogen). The manufacturer's His-tag protein purification protocol was followed with few alterations. The two proteins of interest were mixed on ice, then added to 20 µl of magnetic dynabeads and left to incubate for 30 min at RT with vigorous shaking. This was then placed on a magnet, washed 4 times with 1 x Binding/Wash buffer, incubated for 15 min with 100 µl His Elution buffer with vigorous shaking prior to eluate collection.

#### 2.3.2.5 Anion-exchange chromatography

##### **HiTrap Q Sepharose™ Fast Flow (QFF) column**

The QFF column (GE Healthcare) has a CV of 1 ml, a flow rate of 1 ml/min, a maximum pressure of 0.3 MPa and the purification was done on an AKTA prime (Amersham Biotech). The pumps were first washed with dH<sub>2</sub>O and then the column was connected and 5 CV dH<sub>2</sub>O was passed through. The pumps and column were then rinsed with 5 CV Anion-exchange binding buffer (Section 2.1.6.2) and the protein was injected into the column. A gradient of 0-100% of Elution buffer (Section 2.1.6.2), was passed through the column and the protein was eluted and collected in 1 ml fractions in the fraction collector. A curve of UV absorption against fractions collected was obtained that gave peaks at 280 nm where the protein eluted. From this curve, fractions were chosen and analysed on SDS polyacrylamide gel. The column was stored by pumping 5 CV dH<sub>2</sub>O through the column and then 20% ethanol as a storage buffer.

##### **Mono Q 5/50 GL column**

Mono Q™ 5/50 GL column (GE Healthcare) has a CV of 1 ml, a flow rate of 0.5-3 ml/min, a maximum pressure of 4.2 MPa and the purification was performed on the AKTA purifier Box-900 (Amersham pharmacia) using UNICORN Software. The protocol is as explained for the QFF (see above). However, the gradient used ranged from 30-100%. Using the UV absorption against fractions collected curve, fractions were chosen and analysed on SDS polyacrylamide gel. The column was stored as above.

#### 2.3.2.6 Gel filtration chromatography

The column used for the gel filtration was the HiLoad™ 16/60 Superdex™ 200 prep grade (GE Healthcare). It has a CV of 124 ml, a flow rate of 0.3-1.6 ml/min, a pressure limit of 0.5 MPa and the purification was performed on the AKTA purifier Box-900 (Amersham pharmacia) using the UNICORN software. The pumps and column were washed with 1 CV dH<sub>2</sub>O followed by 1 CV of the gel filtration buffer (Section 2.1.6.2). The protein was subsequently injected into the

column and 1 ml fractions were collected in the fraction collector. A curve was generated of the UV absorption against the fractions collected, which was used to select fractions to analyse on SDS polyacrylamide gel. The column was stored by pumping 1 CV dH<sub>2</sub>O and then 20% ethanol, which serves as the storage buffer.

### **2.3.3 Protein buffer exchange**

To exchange protein buffer, it was placed in a dialysis tubing cellulose membrane, (MWCO 12,000 Daltons, Sigma), which was sealed on both ends and placed in 0.5-1 L of dialysis buffer (composition varied – Refer to results) and stirred overnight at 4°C.

### **2.3.4 Concentrating protein**

VIVASPIN-6 membrane 5,000 MWCO PES (Sartorius stedim biotech) and VIVASPIN-500 membrane 3,000 MWCO PES (Sartorius stedim biotech) were used to concentrate protein samples. The manufacturer's instructions were followed. Prior to use, dH<sub>2</sub>O was added to the column and centrifuged at 2,000 x g, for 5 min at 10°C. The protein sample was then added and centrifuged at 10,000 x g at 4°C until the required concentration or volume was achieved.

### **2.3.5 Determining the protein concentration**

Protein concentrations were determined using two alternative procedures:

#### **2.3.5.1 Bradford assay**

The Bradford assay is based on the absorbance shift of the Brilliant Blue G dye in the Bradford Reagent (Sigma) from 465 to 595 nm when the protein binds. To quantify the protein concentration, a standard curve was prepared using known concentrations of Bovine Serum Albumin (BSA). 0-5 µl of 1 mg/ml BSA was added to 500 µl dH<sub>2</sub>O, 500 µl Bradford Reagent and absorbance at 595 nm were measured. To measure the concentration of unknown protein, X µl of protein was added to 500 µl dH<sub>2</sub>O and 500 µl Bradford reagent and absorbance

measured. The standard curve was plotted and the equation obtained was used to calculate the protein concentration.

#### 2.3.5.2 Bicinchoninic acid assay (BCA)

The BCA assay is based on the reduction of  $\text{Cu}^{2+}$  to  $\text{Cu}^{1+}$  by the protein, which forms a complex with BCA reagent (Sigma) and the absorbance measured at 562 nm. The procedure is followed as per manufacturer's recommendation. The protein determinant reagent was prepared by adding 1 ml of 4% (w/v)  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  solution to 49 ml of the BCA solution. A standard curve was first generated adding known concentrations of BSA, samples vortex mixed, incubated at 37°C for 30 min and then the absorbance measured at 562 nm. To determine the unknown protein concentration, 0.01-0.1 ml of the protein sample was added to 2 ml of the protein determination reagent and the same procedure was repeated. Using the absorbance reading, the concentration was determined by extrapolation from the standard curve.

### **2.3.6 Protein sample analysis by SDS polyacrylamide gel**

#### 2.3.6.1 NuPAGE® 4-12% Bis-Tris gels

Precast NuPAGE® 4-12% Bis-Tris gels (Invitrogen) were used for analysis of protein samples. The protein samples were prepared to load on the gel as per manufacturer's instruction. SeeBlue® Plus2 Pre-stained standard marker (Invitrogen) was also loaded on the gel. The gels were run on Hoefer SE260, a mini-vertical gel electrophoresis unit (Amersham Biosciences) at 120 V for 1 h 20 min. The running buffer used was NuPAGE® MES SDS Running Buffer (Invitrogen).

#### 2.3.6.2 SDS polyacrylamide gel preparation

SDS polyacrylamide gel were prepared with a 15% resolving gel and a 5% stacking gel. The gel plates were washed with detergent, wiped with 70% ethanol and the plates assembled using a Hoefer Mighty Small dual gel caster

(Hoefer miniSE 260 unit– Amersham Biosciences) as per manufacturer's recommendation. The resolving and stacking gels were made as follows:

15% resolving gel: 5 ml 30% (v/v) acrylamide/bis-acrylamide mix solution (Severn Biotech), 2.5 ml 1.5 M Tris pH 8.8, 0.1 ml 10% (v/v) SDS, 50  $\mu$ l 10% (w/v) Ammonium Persulphate (APS), 20  $\mu$ l TEMED and 2.33 ml dH<sub>2</sub>O.

5% stacking gel: 0.75 ml 30% (v/v) acrylamide/bis-acrylamide mix solution (Severn Biotech), 1.25 ml 0.5 M Tris pH 6.8, 50  $\mu$ l 10% (v/v) SDS, 50  $\mu$ l 10% APS, 8  $\mu$ l TEMED and 2.89 ml dH<sub>2</sub>O.

The resolving gel was first poured into the gel cassette to cover  $\frac{3}{4}$  of the gel plates and 70% ethanol was added. After 15 min, the resolving gel had polymerized, the ethanol was poured out and the gel was rinsed with dH<sub>2</sub>O. The stacking gel was added and immediately the comb was inserted carefully ensuring no air bubble was introduced. This was left to stand for about 30 min. The protein samples were then prepared to load on the gel by adding 2 x protein loading dye with DTT (Section 2.1.6.2). The samples were heat denatured at 100°C for 5 min and ran on the SDS polyacrylamide gel on Hoefer SE260, a mini-vertical gel electrophoresis unit (Amersham Biosciences) at 200 V for 70 min using 1x SDS running buffer (Section 2.1.6.2). The protein bands were visualized by staining with Coomassie Brilliant Blue (Section 2.1.6.2) and the background removed by the destainer solution (Section 2.1.6.2).

### **2.3.7 Hanging drop vapour diffusion**

The “hanging drop vapour diffusion” method was used to check for appropriate buffer conditions for the formation of crystals. 1 ml of different buffers was added into each well of a 24-well plate. To the coverslip, 1  $\mu$ l of the buffer and 1  $\mu$ l of the protein sample were mixed in a single drop and sealed with grease on the wells. These plates were left at 4°C for 2-5 days.

### **2.3.8 Protein preparation for ICP-MS**

A quick buffer exchange into Charge buffer (Section 2.1.6.2) was performed on 0.5 mg purified protein. This was then dried at 105°C for 2-4 h and allowed to cool. 70% metal analysis grade nitric acid was added to digest the protein for 10 min at 105°C and metal analysis grade H<sub>2</sub>O was added to a volume of 3 ml. This was sent for inductively coupled plasma mass spectrometry (ICP-MS) together with a control to serve as background (University of Sussex).

## **2.4 Western blotting**

### **2.4.1 Preparing samples**

At specific stages of growth, *S. coelicolor* samples were collected to test for the level of protein expression. 200-500 µl samples were taken at specific time points, centrifuged at 16,100 x g, for 10 sec at 4°C, 240 µl ice-cold 0.9% (w/v) NaCl was added followed by 60 µl of 50% (v/v) TCA and left on ice for 15 min. This was then centrifuged at 16,100 x g for 1 min at 4°C, pellet resuspended in 100 µl of 1 x loading buffer with reducing agent (Invitrogen), boiled at 100°C for 5 min and centrifuged at 16,100 rpm for 8 min at 4°C. The supernatant was then loaded on 4-12% Bis-Tris polyacrylamide gel for electrophoresis followed by western blotting.

### **2.4.2 Determining protein concentration**

The samples were centrifuged and TCA-precipitated as in Section 2.4.1. However, the pellets were resuspended in 0.1 ml of 5% (v/v) SDS/0.1 M Tris. The samples were vortex mixed, boiled at 100°C for 5 min and added to 2 ml BCA reagent for protein concentration quantification (Section 2.3.5.2).

### **2.4.3 Immunoblotting protocol**

After electrophoresis, the Protran<sup>®</sup> nitrocellulose membrane (0.2µm- Whatman) was cut at the desired size and soaked together with the sponge pads and chromatography paper sheets (3mm- Whatman) in chilled Transfer buffer (Section 2.1.6.3). The cassette was placed with the dark side down and the order was: sponge pad - filter paper - gel - membrane - filter paper - sponge pad. The cassette was closed and placed in the transfer tank together with Transfer buffer and transferred at 100 V, 4°C, for 1-1.5 h depending on the size of the protein of interest.

### **2.4.4 Washes and antibodies**

After the proteins were transferred onto the nitrocellulose membrane, it was rinsed in Transfer buffer and incubated in blocking solution (Section 2.1.6.3) at RT for 1-2 h with continuous agitation. The membrane was then washed 3 times for 10 min with 1 x TBS-Tween (Section 2.1.6.3) and incubated overnight at 4°C with gentle rocking in the appropriate dilution of primary antibody, which was diluted in freshly made blocking solution. The membrane was again washed 3 times with 1 x TBS-Tween for 10 min and then incubated with gentle rocking at 4°C for 1 h in an appropriate dilution of secondary antibody HRP conjugate, which was made in fresh blocking solution. The membrane was again washed 3 times for 10 min and then subjected to ECL detection.

### **2.4.5 ECL detection and developing**

The Amersham<sup>™</sup> ECL<sup>™</sup> Prime Western Blotting Detection Reagent (GE Healthcare) was used as per manufacturer's instructions. 500 µl Solution I and 500 µl Solution II (GE Healthcare) were mixed together, added to the membrane and left to stand for 1 min. The membrane was then removed with forceps and gently touched on the filter paper to remove any excess solution. The soaked membrane was then wrapped in Saran<sup>™</sup> wrap and then exposed in the DNR Bio-Imaging Systems F-chemiBlis 3.2 M machine using the Gel capture software for 20-120 s depending on the intensity of the protein bands.

## **2.5 *In vitro* transcription assays**

### **2.5.1 Preparing the gel**

The plates and comb were washed with detergent and wiped with 100% ethanol. To the top plate, 100  $\mu$ l Sigmacote<sup>®</sup> (Sigma) was added and wiped over the face of the plate to ensure that the gel sticks to the bottom plate. This was again wiped with 100% ethanol. The gel was assembled onto the AE-6141 MiniQuencer (31 x 18 x 32 cm) (Atto corporation- Japan). 10 ml Urea Ultra Pure Sequagel<sup>®</sup> Urea Gel™ Complete (national diagnostics) was added to 20 ml Ultra Pure Sequagel<sup>®</sup> Urea Gel™-6 (national diagnostics) together with 200  $\mu$ l 10% APS. This was quickly poured into the assembled plates, comb inserted and the cassette lied down to set for 2 h. The gel was then pre-run on AE-6141E MiniQuencer (Atto) at 600 V, 50 W for 30 min at 55°C. The samples were then loaded on gel and ran at 600 V, 50 W for 1 h at 55°C.

### **2.5.2 *In vitro* transcription assays**

1  $\mu$ l 5-25  $\mu$ M RbpA was first added to 1  $\mu$ l 2.5  $\mu$ M  $\sigma^{\text{HrdB}}$ , incubated on ice for 15 min, followed by addition of 1  $\mu$ l 500 nM RNAP and again incubated on ice for 15 min. 1  $\mu$ l 50 nM template was then added and left on ice for 15 min. 2  $\mu$ l 10 x transcription buffer/2 M KCl (Section 2.1.6.4) was added and RNase-free H<sub>2</sub>O used to make up the reaction mixture to 10  $\mu$ l. The reaction was started by adding 1  $\mu$ l radioactive NTP mix and placing the reaction at 30°C for 15 min. The reaction was stopped by adding 10  $\mu$ l Formamide loading buffer (Section 2.1.6.4), heated at 80°C for 5 min, loaded on the gel and ran at 600 V, 50 W for 1 h at 55°C. The gel was then transferred to a piece of Whatmann 3 MM paper and dried at 65°C for 1 h using a gel vacuum drier. The dried gel was then exposed to a phosphoimaging plate (Amersham Biochemia) at RT for 16-48 h. The bands were visualized using the typhoon scanner.



## **2.6 $\beta$ -galactosidase assay**

*In vivo* protein-protein interactions were detected using Bacterial adenylate cyclase two-hybrid system (BACTH). The amplified genes were cloned into pBlueScript II SK (+) (Table 1) and then subcloned into either pUT18 or pKT25 vector (Table 1). The recombinant pUT18 and pKT25 vectors were transformed into BTH101 (Table 2) competent cells by heat shock. The BTH101 cells used for the assay were grown in 5 ml LB (without glucose) in the presence of 50  $\mu\text{g}/\mu\text{l}$  Kanamycin, 100  $\mu\text{g}/\mu\text{l}$  ampicillin and 0.5 mM IPTG at 30°C overnight with vigorous shaking. These were diluted 1:5 with M63 minimal medium (Section 2.1.4.1) and the  $\text{OD}_{600}$  was recorded. The cells were permeabilized with 30  $\mu\text{l}$  chloroform, 30  $\mu\text{l}$  0.1% (v/v) SDS, vortex mixed for 10 s and lightly plunged with cotton wool. These were vigorously agitated at 37°C for 30-40 min to allow evaporation of chloroform.

For the enzymatic reaction, 0.1-0.2 ml of permeabilised cells was added to PM2 medium (Section 2.1.6.5) to a volume of 1 ml. These tubes were placed in a 28°C water bath for 5 min. A control tube containing 1 ml PM2 assay buffer was also prepared to serve as the blank. The enzymatic reaction was started by adding 250  $\mu\text{l}$  pre-equilibrated 4 mg/ml ONPG and the time added was noted. When a sufficient yellow colour was observed, 500  $\mu\text{l}$  1 M  $\text{Na}_2\text{CO}_3$  was added to stop the reaction and the time was also noted to calculate the reaction time of the enzyme. The  $\text{OD}_{420}$  was recorded and the enzymatic reaction was done in triplicate to ensure accuracy.

The calculation was done based on the formula:

$$\text{Activity units/ml} = \frac{1000 \times \text{absorbance at 420nm}}{\text{Reaction time (min)} \times \text{volume of culture (ml)} \times \text{absorbance at 600nm}}$$

### **3- RbpA Bioinformatics and its Biological Role**

### **3.0 Overview**

RbpA is a newly discovered RNAP binding protein and thus, there is little information available in the literature regarding its structure and function. Therefore, the first section of this chapter presents a bioinformatics study on RbpA, particularly its conservation among the Actinobacteria, some structural predictions and also the presence of its paralogue, RbpB, in some Actinobacteria. The second section presents some experimental attempts to determine the biological role of RbpA, in particular, transcriptome analysis and the role of RbpA in determining RNAP composition.

### **3.1 Bioinformatic analysis of RbpA**

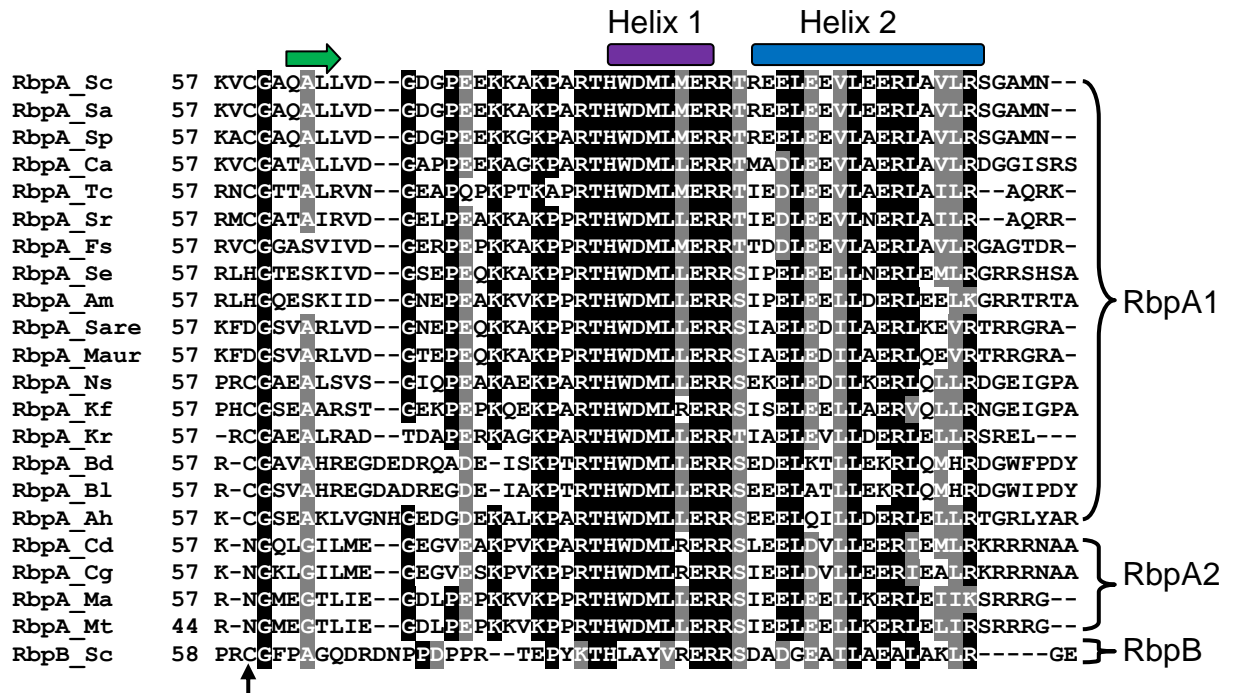
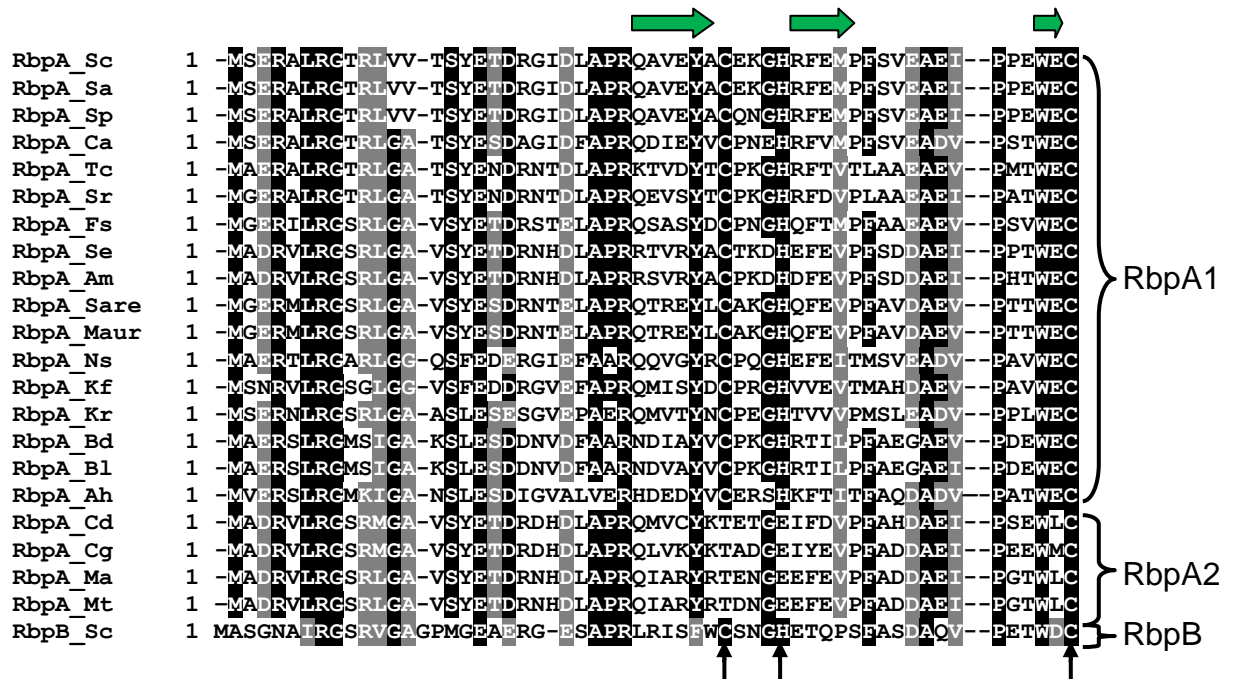
The recent increase in the number of available bacterial genomes allows an assessment of the distribution of RbpA homologues in bacteria as well as useful insights into the conservation of certain amino acid residues. In *S. coelicolor*, RbpA was shown to have a paralogue, RbpB. The distribution of this protein was also assessed.

#### **3.1.1 RbpA in the Actinobacteria**

RbpA (SCO1421) is encoded by *rbpA*, a monocistronic gene that was shown to be partially controlled by  $\sigma^R$  (Paget *et al.*, 2001b). The *rbpA* gene is 375 bp long, and encodes a protein of 124 amino acids with a molecular weight of approximately 14 kDa. To identify the level of conservation and distribution of RbpA, BLASTP and TBLASTN searches using the Integrated Microbial Genome (IMG) database were performed with the amino acid sequence of RbpA. RbpA homologues were present in 255 genomes, all belonging to the Actinobacteria including the following genera: *Streptomyces*, *Frankia*, *Mycobacteria*, *Bifidobacterium*, and *Corynebacterium* (Table A1- Appendices). No RbpA-related protein was identified outside of the Actinobacteria.

There are three distinct classes of RbpA homologues. Two are closely related differing by the presence or absence of putative Zn ligands, for example Zn-

binding RbpA in *S. coelicolor*, and the putative Zn-free Rv2050 in *M. tuberculosis* (Section 6.3). The third class is distinct from RbpA and is represented by the paralogue, RbpB in *S. coelicolor* (Section 3.1.2). A multiple sequence alignment was generated for 22 RbpA homologues, selected from a wide range of Actinobacteria in order to reveal highly conserved amino acid residues (Fig. 3.1). The multiple sequence alignment showed that the RbpA homologues share a high level of similarity (Fig. 3.1).



RbpA_Sc	112	I	A	V	H	P	R	D	S	R	-----	K	S	A	-----																																						
RbpA_Sa	112	I	A	V	H	P	R	D	S	R	-----	K	S	A	-----																																						
RbpA_Sp	112	I	A	V	H	P	R	D	S	R	-----	K	S	A	-----																																						
RbpA_Ca	114	D	V	V	R	P	A	P	A	R	T	M	A	G	R	A	K	S	A	A	A	K	A	S	S	V	K	A	P	A	P	A	G	K	P	K	T	T	A	A	R	S	G	A	R	K	A	A	A	H	S	S	D
RbpA_Tc	111	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	K	S	A	-----																																						
RbpA_Sr	111	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	K	S	A	-----																																						
RbpA_Fs	113	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	R	S	A	-----																																						
RbpA_Se	114	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----																																						
RbpA_Am	114	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----																																						
RbpA_Sare	113	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----																																						
RbpA_Maur	113	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----																																						
RbpA_Ns	114	H	L	H	R	A	N	Q	K	K	----	R	K	A	T	A	----																																				
RbpA_Kf	114	H	L	H	R	A	R	N	T	S	----	G	K	R	T	A	----																																				
RbpA_Kr	110	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	I	K	R	S	A	----																																				
RbpA_Bd	114	E	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----																																						
RbpA_Bl	114	E	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----																																						
RbpA_Ah	115	R	H	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----																																						
RbpA_Cd	113	R	L	L	K	Q	Q	Q	E	E	A	A	----	A	Q	N	S	----																																			
RbpA_Cg	113	K	L	L	K	A	Q	Q	E	A	E	E	A	----	E	K	A	A	E	E	V	----																															
RbpA_Ma	111	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----																																						
RbpA_Mt	98	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----																																						
RbpB_Sc	110	-	I	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----																																						

RbpA1

RbpA2

RbpB

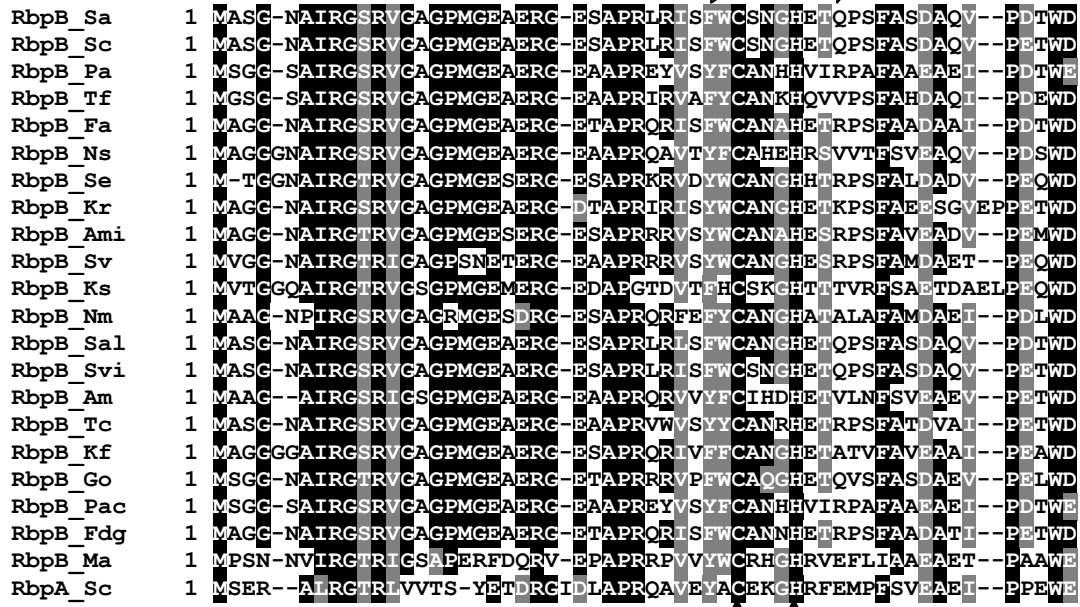
**Figure 3.1: Multiple sequence alignment of 22 homologues of RbpA in the Actinobacteria phylum.**

The Clustal W2 multiple sequence alignment tool and the BOXSHADE 3.21 server were used. Amino acid residues shown in white font with black highlight showed identical residues whilst the amino acid residues in white font with grey highlight showed similar residues. The homology was tested to the 0.8 fraction that must agree for shading. The black vertical arrows show the conserved residues that co-ordinate Zn. The PSIPRED server was used to predict the secondary structure of RbpA in *S. coelicolor* with the horizontal green arrows representing the  $\beta$ -strands, the purple cylinder: Helix 1 and the blue cylinder: Helix 2. On the side of the alignment, RbpA1: Zn-containing, RbpA2: Non-Zn containing, RbpB: paralogue of RbpA. RbpA\_Sc: *Streptomyces coelicolor* A3(2), RbpA\_Sa: *Streptomyces avermitilis* MA-4680, RbpA\_Sp: *Streptomyces pristinaespiralis* ATCC 25486, RbpA\_Ca: *Catenulispora acidiphila* DSM 44928, RbpA\_Tc: *Thermomonospora curvata* DSM 43183, RbpA\_Sr: *Streptosporangium roseum* DSM 43021, RbpA\_Fs: *Francisci3\_2270* *Frankia* sp. Ccl3, RbpA\_Se: *Saccharopolyspora erythraea* NRRL 2338, RbpA\_Am: *Actinosynnema mirum* DSM 43827, RbpA\_Sare: *Salinispora arenicola* CNS-205, RbpA\_Maur: *Micromonospora aurantiaca* ATCC 27029, RbpA\_Ns: *Nocardioideis* sp. JS614, RbpA\_Kf: *Kribbella flavida* DSM 17836, RbpA\_Kr: *Kineococcus radiotolerans*

SRS30216, RbpA\_Bd: *Bifidobacterium dentium* Bd1, RbpA\_Bl: *Bifidobacterium longum* subsp. *longum* JDM301, RbpA\_Ah: *Arcanobacterium haemolyticum* DSM 20595, RbpA\_Cd: *Corynebacterium diphtheriae* NCTC 13129, RbpA\_Cg: *Corynebacterium glutamicum* ATCC 13032, RbpA\_Ma: *Mycobacterium avium* subsp. *paratuberculosis* K-10, RbpA\_Mt: *Mycobacterium tuberculosis* H37Ra, RbpB\_Sc: *Streptomyces coelicolor* A3 (2).

### 3.1.2 The RbpA paralogue, RbpB

There is a paralogue of RbpA in *S. coelicolor*, known as RbpB (Fig. 3.1) (Paget *et al.*, 2001b). To determine whether RbpB is as widely conserved in the Actinobacteria as RbpA, BLASTP and TBLASTN searches were performed using the IMG database. This revealed 63 orthologues of RbpB (Table A1-Appendices). RbpB was shown to be 100% conserved amongst the *Streptomyces* genus and the *Frankia* genus. Other examples of organisms that have RbpB include *Micromonospora* sp., *Nocardioides* sp., *Propionibacterium acnes*, *Thermobifido fusca* (Fig. 3.2). In contrast, the *Mycobacterium* genus completely lacked RbpB. The putative Zn ligands, three cysteines and one histidine, are well conserved in RbpB homologues suggesting that RbpB might be a Zn metalloprotein.


$\beta$ -strands


Sequence alignment of RbpB proteins from various species, highlighting beta-strands with green arrows. The alignment shows conserved regions across the proteins, with the beta-strands indicated by green arrows pointing to specific residues.

Species	Sequence
RbpB_Sa	1 MASG-NAIRGSRVGAGPMGEAERG-ESAPRLRISFWCSNGHETQPSFASDAQV--PDTWD
RbpB_Sc	1 MASG-NAIRGSRVGAGPMGEAERG-ESAPRLRISFWCSNGHETQPSFASDAQV--PETWD
RbpB_Pa	1 MSGG-SAIRGSRVGAGPMGEAERG-EAAPREYVSYFCANHHVIRPAFAAEAEI--PDTWE
RbpB_Tf	1 MSGG-SAIRGSRVGAGPMGEAERG-EAAPRLRVAFYCANKHQVVPFSAHDAQI--PDEWD
RbpB_Fa	1 MAGG-NAIRGSRVGAGPMGEAERG-ETAPRORISFWCANAHETRPSFAADAAT--PDTWD
RbpB_Ns	1 MAGGNAIRGSRVGAGPMGEAERG-EAAPROAVTYFCAHEHRSVVTFSEAEQV--PDSWD
RbpB_Se	1 M-TGGNAIRGTRVGAGPMGESERG-ESAPRRVDYWCANGHETRPSFALDADV--PEQWD
RbpB_Kr	1 MAGG-NAIRGSRVGAGPMGEAERG-DTAPRLRISYWCANGHETKPSFAEESGVEPPETWD
RbpB_Ami	1 MAGG-NAIRGTRVGAGPMGESERG-ESAPRRRVSYWCANAHESRPSFAVEADV--PEMWD
RbpB_Sv	1 MVGG-NAIRGTRIGAGPSNETERG-EAAPRRRVSYWCANGHESRPSFAMDAET--PEQWD
RbpB_Ks	1 MVTGGQAIRGTRVGSGPMGESERG-EDAPGTDVTFHCSKGHTTIVRFSAETDAELPEQWD
RbpB_Nm	1 MAAG-NPIRGSRVGAGRMGESDRG-ESAPRRFEFYCANGHATALAFAMDAEI--PDLWD
RbpB_Sal	1 MASG-NAIRGSRVGAGPMGEAERG-ESAPRLRISFWCANGHETQPSFASDAQV--PDTWD
RbpB_Svi	1 MASG-NAIRGSRVGAGPMGEAERG-ESAPRLRISFWCSNGHETQPSFASDAQV--PETWD
RbpB_Am	1 MAAG--AIRGSRIGSGPMGEAERG-EAAPRRRVYVFCIHDEETVLNFSVEAEV--PETWD
RbpB_Tc	1 MASG-NAIRGSRVGAGPMGEAERG-EAAPRVVVSYYCANRHETRPSFATDVAI--PETWD
RbpB_Kf	1 MAGGGCAIRGSRVGAGPMGEAERG-ESAPRRIVFFCANGHETATVFVAEAI--PEAWD
RbpB_Go	1 MSGG-NAIRGTRVGAGPMGEAERG-ETAPRRRVPEWCAQGHETQVSFASDAEV--PELWD
RbpB_Pac	1 MSGG-SAIRGSRVGAGPMGEAERG-EAAPREYVSYFCANHHVIRPAFAAEAEI--PDTWE
RbpB_Fdg	1 MAGG-NAIRGSRVGAGPMGEAERG-ETAPRORISFWCANHETRPSEADATI--PETWD
RbpB_Ma	1 MPSN-NVIRGTRIGSAPERFDQRV-EPAPRRVVYWCRRHGRVEFLIAPAEI--PAawe
RbpA_Sc	1 MSER--ALRGTRLVVTs-YETDRGIDLAPRQAVEYACEKGRFEMPFSVDAEI--PPEWE

Helix 1

Helix 2



Sequence alignment of RbpB proteins from various species, highlighting Helix 1 and Helix 2 with purple and blue bars. The alignment shows conserved regions across the proteins, with Helix 1 and Helix 2 indicated by purple and blue bars respectively.

Species	Sequence
RbpB_Sa	56 CPRCGFPAGODQDNPPDPPTTEPYKTHLAYVRERRSDADGEAILAEALAKLRGEI-----
RbpB_Sc	56 CPRCGFPAGQDRDNPPDPPTTEPYKTHLAYVRERRSDADGEAILAEALAKLRGEI-----
RbpB_Pa	56 CPRCGLPANRDQANPPAPPKIVPYKTHLAYVKERRSEAEAAAILAEALENLRQRRRDGEV
RbpB_Tf	56 CPRCGFPAGKDPENPPAPPKTEPYKTHLAYVKERRSEADGEAILQEATAKLRAERAAMSS
RbpB_Fa	40 CPSCGFPAGRDNDNTPAPPPTTEPYKTHLAYVRERRNEKQGEAILAEALAKLRGAP-----
RbpB_Ns	57 CPKCGLPASLDSNPAPPKTEPYKTHLAYVKERRSDQEAADILDEALQLRSRRKSGDI
RbpB_Se	40 CPRCGLPAGQDEQNPPPEARNEPYKSHLAYVKERRSDSDGMAILLEALSKLKERKGR---
RbpB_Kr	58 CPRCGFPAGTDNANPPAPPKNEPYKTHLAYVKERRSDADGEAILNEALDSLRRAGLIR--
RbpB_Ami	40 CPRCGLPAGRDEQAPPAPPKNEPYKTHLAYVKERRSDADGEAILLEEALTKLRRQREEEG
RbpB_Sv	56 CPRCGLPAGRDRENPPAAPRTTEPYKSHLAYVKERRSDADGEAILAEALERLRRRREGAEG
RbpB_Ks	59 CTRCGLPAGTDNANPPDPPTTTPYKTHLAYVQERRDDAAGAILDEALGKLREQRGTA--
RbpB_Nm	56 CPRCGSPAGRDRAAPPAPATVEPYKSHLAYVKERRSDEGEAILAEALARLTDKRTPRS-
RbpB_Sal	56 CPRCGFPAGQDRDDPPAPPPTTEPYKTHLAYVRERRSDADGEAILAEALAKLRGEI-----
RbpB_Svi	56 CPRCGFPAGQDRDNPPDPPTTEPYKTHLAYVRERRSDADGEAILAEALAKLRGEI-----
RbpB_Am	55 CPRCGMPASPDSDNRPTAPKTEPYKTHLAYVKERRSDDEAQQILAEALDILLRSRRKSSEV
RbpB_Tc	56 CPRCGFPAGKDPDNPPAPPKTEPYKTHLAYVKERRSDADGQAILLEEALAKLRERRAAVKA
RbpB_Kf	57 CPRCGLPSTSDVNNPPPPPKTEPYKTHLAYVKERRSEQEAAQILLEALERLRARRAGGEI
RbpB_Go	56 CPRCGLPAGQDQAAPPAPPKTEPYKTHLAYVRERRSDADGDALLEEALAKLRARRGA---
RbpB_Pac	56 CPRCGLPANRDQANPPAPPKIVPYKTHLAYVKERRSEAEAAAILAEALENLRQRRRDGEV
RbpB_Fdg	40 CPNCGFPAGRDNDNAPAPPKTEPYKTHLAYVRERRNEKQGEAILAEALAKLRGTA-----
RbpB_Ma	56 CPRCGEPAGDPDGDPPGROAEPEYKTHLAYVQERRTPESGEAILAEALDALRRRRRGRA--
RbpA_Sc	55 CKVCGAQALLVDGDGPEEKAKAPARTHWDMMERRTREELLEVLEERLAVLRSGAMNIAV



RbpB_Sa	111	-----
RbpB_Sc	111	-----
RbpB_Pa	116	IY-----
RbpB_Tf	116	ALKSFTD--
RbpB_Fa	95	-----
RbpB_Ns	117	IF-----
RbpB_Se	97	-----
RbpB_Kr	116	-----
RbpB_Ami	98	-----
RbpB_Sv	115	-----
RbpB_Ks	117	-----
RbpB_Nm	115	-----
RbpB_Sal	111	-----
RbpB_Svi	111	-----
RbpB_Am	115	IF-----
RbpB_Tc	116	AMESVNNRN
RbpB_Kf	117	LF-----
RbpB_Go	113	-----
RbpB_Pac	116	IY-----
RbpB_Fdg	95	-----
RbpB_Ma	114	-----
RbpA_Sc	115	HPRDSRKSA

**Figure 3.2: Multiple sequence alignment of 22 homologues of RbpB in the Actinobacteria phylum.**

The Clustal W2 multiple sequence alignment tool and the BOXSHADE 3.21 server were used. Amino acid residues shown in white font with black highlight showed identical residues whilst the amino acid residues in white font with grey highlight showed similar residues. The homology was tested to the 0.7 fraction that must agree for shading. The black vertical arrows show the conserved residues that co-ordinate Zn. The PSIPRED server was used to predict the secondary structure of RbpB in *S. coelicolor* with the horizontal green arrows representing the  $\beta$ -strands, purple cyclinder: Helix 1 and blue cylinder: Helix 2. RbpB\_Sc: *Streptomyces coelicolor* A3 (2), RbpB\_Sa: *Streptomyces avermitilis*, RbpB\_Pa: *Propionibacterium acnes*, RbpB\_Tf: *Thermobifida fusca*, RbpB\_Fa: *Frankia alni* ACN14a, RbpB\_Ns: *Nocardioides* sp. JS614, RbpB\_Se: *Saccharopolyspora erythraea* NRRL 2338, RbpB\_Kr: *Kineococcus radiotolerans* SRS30216, RbpB\_Ami: *Actinosynnema mirum* DSM 43827, RbpB\_Sv: *Saccharomonospora viridis* DSM 43017, RbpB\_Ks: *Kytococcus sedentarius* DSM 20547, RbpB\_Nm: *Nakamurella multipartita* DSM 44233, RbpB\_Sal: *Streptomyces albus* J1074, RbpB\_Svi: *Streptomyces viridochromogenes* DSM 40736, RbpB\_Am: *Aeromicrobium marinum* DSM 15272, RbpB\_Tc: *Thermomonospora curvata* DSM 43183, RbpB\_Kf: *Kribbella flavida* DSM 17836,

RbpB\_Go: *Geodermatophilus obscurus* DSM 43160, RbpB\_Pac: *Propionibacterium acnes* J139, RbpB\_Fdg: *Frankia* symbiont of *Datisca glomerata* ctg00111, RbpB\_Ma: *Micromonospora aurantiaca* ATCC 27029, RbpA\_Sc: *Streptomyces coelicolor* A3 (2).

### 3.1.3 The secondary structural predictions of RbpA

Prior to this study (see Chapter 6), a structure of a member of the RbpA family was not available and there seemed to be no related proteins in the protein structure database. Therefore, a secondary structure prediction of RbpA was performed using the PSIPRED protein structure prediction server (Fig. 3.1). RbpA was predicted to have a significant amount of random coil at the N-terminus, suggesting that parts of the protein might be disordered. RbpA also appears to have four  $\beta$ -strands located N-terminally (Fig. 3.1) and two  $\alpha$ -helices located C-terminally (Fig. 3.1). The two helices are labelled Helix 1 (H1) and Helix 2 (H2) and seem to be connected by a short loop (Fig. 3.1). The helices have a high confidence level of 9 or 8.

#### 3.1.3.1 The N-terminal region

The sequence-based structural prediction programme Phyre (Protein Homology/analogy Recognition Engine) was used to identify possible structural homologues of RbpA. RbpA appeared to be related to members of the rubredoxin superfamily and has a predicted rubredoxin-like fold ( $\beta$ -fold). An example of a protein distantly similar in structure to RbpA is the rubredoxin (Protein Data Bank, d6rxna) from *Desulfovibrio desulfuricans* strain 27774. The d6rxna Chain A structure has 45 amino acids and is complexed with iron. When the d6rxna secondary structure was compared to that of RbpA, a rubredoxin-like fold was predicted between amino acid residues 30-60 in RbpA.

Another distantly related structure to RbpA was PDB c2v3b chain B from *Pseudomonas aeruginosa*. Its crystal structure was solved and it was shown to

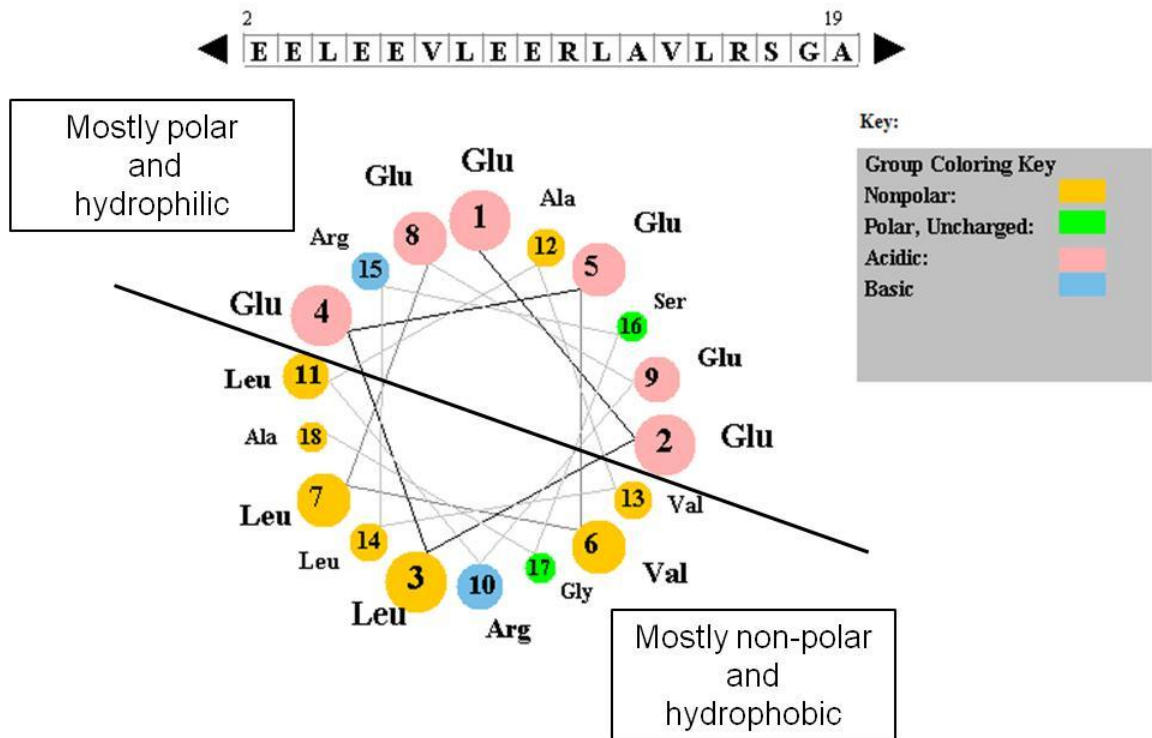
contain two chains: chain A, which is a Rubredoxin Reductase and Chain B, which is Rubredoxin 2 (Hagelueken *et al.*, 2007). Chain B has 55 residues with 10% of it being helical and 21% occurring as  $\beta$ -sheets. This domain has important functions in iron ion binding and electron carrier activity (Hagelueken *et al.*, 2007).

PDB d6rxna and c2v3b both bind to a metal complex (iron). This iron is coordinated by four cysteine residues (Hagelueken *et al.*, 2007). RbpA has also been shown to complex a metal (Zn) (P. Doughty, M. Paget, personal communication). Hence, it is likely that RbpA is similar to these rubredoxins structurally but not functionally.

### 3.1.3.2 The C-terminal region

The C-terminal region of RbpA was predicted to have two  $\alpha$ -helices separated by a short turn. H1 was composed of residues that were very highly conserved in most RbpA homologues (Fig. 3.1). H1 was also present in RbpB and was highly conserved amongst the RbpB homologues (Fig. 3.2). The ERR motif located just downstream of H1 was also very highly conserved amongst the RbpA and RbpB homologues (Fig. 3.1 & 3.2).

H2 in RbpA has high conservation in leucine (L), glutamate (E) and arginine (R) residues. Due to the high presence of hydrophobic residues (L & V) and hydrophilic residues (E), the helical wheel applet was used to determine whether this was an amphipathic helix. From the helical wheel representation, it was seen that one side of H2 is mainly acidic and hydrophilic (E), whilst the other side is mostly non-polar and hydrophobic (L & V) (Fig. 3.3). This indicates that H2 is an amphipathic helix (EELEEVL EERLAVL). In RbpB, H2 is not amphipathic as it has high presence of hydrophobic residues (alanine- A & L) with very few hydrophilic residues (E & aspartate- D) (helical wheel not shown).



**Figure 3.3: A helical wheel representation of Helix 2 showing the amino acid distribution in the helix.**

The line drawn in the middle separates the hydrophilic residues (Glu) from the hydrophobic residues (Leu & Val). The group coloring key is included at the top right hand side of the image. The size of the coloured circles signify the position of these residues within the helix, that is large to small circles represent residues from start to further down in the helix.

### **3.2 Biological role of RbpA in *S. coelicolor***

#### **3.2.1 Essential for normal growth**

Much of the work in this thesis involves the use of *S. coelicolor* strains S129 and J1981 (Table 3 - Section 2.1.3.2), which have  $\Delta rbpA::aac(3)IV$  and wild-type (w/t) alleles, respectively, together with a His-tagged  $\beta'$  subunit. The strains were streaked to single colonies on MS agar plates and left to grow at 30°C for 4 to 5 days. J1981 (w/t) showed normal growth with appropriately sized colonies,

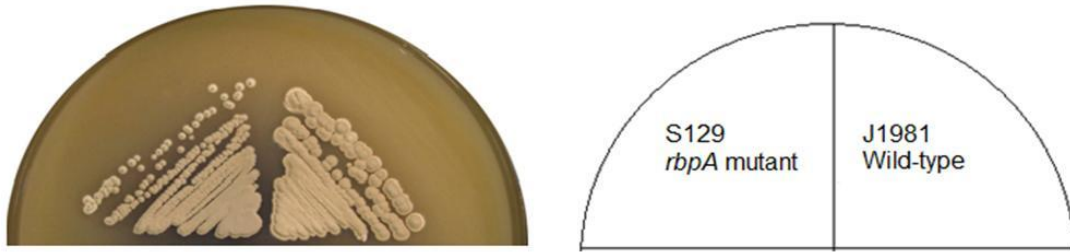
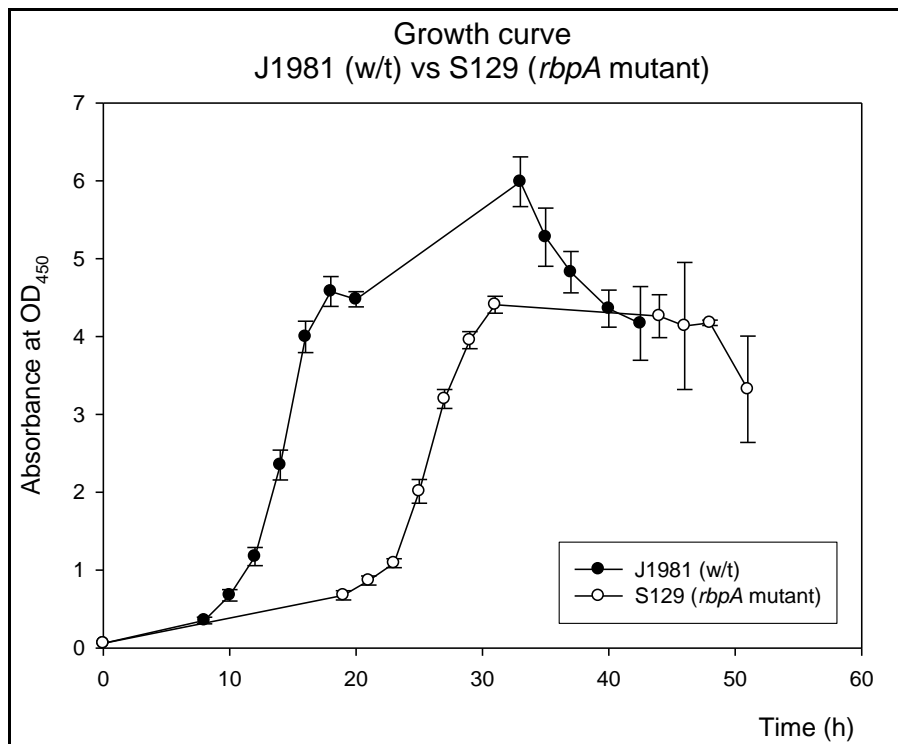
efficient sporulation (grey) and appropriate production of the antibiotic actinorhodin (Fig. 3.4A). S129 ( $\Delta rbpA::aac(3)/V$ ) grew slowly producing smaller sized colonies, efficient sporulation (grey), and the over-production of actinorhodin (Fig. 3.4A). Hence, RbpA was shown to be required for the normal growth of *S. coelicolor* as already shown by Newell, K., *et al* 2006.

Previous growth curves were performed using a supplemented minimal medium (Newell *et al.*, 2006). To confirm if the growth rate phenotype was retained in a rich medium, S129 and J1981 were grown in YEME medium (Fig. 3.4B). The spores were pre-germinated for 4 h with a starting OD<sub>450</sub> of 0.06 (Section 2.2.5.3). These spores were then inoculated in YEME (10% (w/v) sucrose) liquid media and left to grow to early exponential phase before absorbance readings were taken at 2 h intervals until stationary phase (Fig. 3.4B).

J1981 had a short lag phase, entering early exponential phase (OD<sub>450</sub> of 0.8) in approximately 8 h and reached early stationary phase (OD<sub>450</sub> of 5) around 22 h (Fig. 3.4B). S129, on the other hand, had a longer lag phase entering early exponential phase (OD<sub>450</sub> of 0.7) in approximately 19 h and reached early stationary (OD<sub>450</sub> of 4.2) in approximately 31 h. Thus, S129 had a 10 h longer lag phase compared to J1981.

The doubling time of J1981 during exponential phase is 2 h and that of S129 varied from 2 to 3 h. However, at particular points during exponential phase the doubling time of S129 was only slightly slower than that of the w/t. The doubling time of S129 observed in YEME is faster than that observed with NMMP by Newell, K., *et al* (2006). J1981 exhibited a classical transition phase, shifting from primary to secondary metabolism around 18 h. This transition occurs when the nutrient level is being depleted and therefore, the cells stop growing and start switching on secondary metabolism, e.g. production of antibiotics. This was then followed by the major transition state from late exponential phase to early stationary phase around 22 h (Fig. 3.4B). S129, also exhibited a classical

transition phase around 27 h, which was only seen when the dry weights of the cultures were taken, shown in Fig. A1 (Appendices), and entered stationary phase at approximately 31 h (Fig. 3.4B).

**A****B**

**Figure 3.4: RbpA is essential to maintain the normal growth rate of *S. coelicolor*.**

**(A)** *S. coelicolor* J1981 (w/t) and S129 ( $\Delta rbpA::aac(3)/V$ ) strains were streaked into single colonies on MS agar and left to grow at 30°C for 4-5 days. **(B)** Graphical representation of the growth curve of J1981 and S129 using SigmaPlot. Samples were taken at early exponential phase and then at 2 h intervals until stationary phase and the absorbance at OD<sub>450</sub> were measured. The error bars shown represent the standard deviation of triplicate samples.

### 3.2.2 Gene expression in *S. coelicolor* $\Delta rbpA$ strain

*S. coelicolor*  $\Delta rbpA$  strains have a slower growth rate, form small colonies and over-produce the antibiotic actinorhodin (Fig. 3.4). S101, an  $\Delta rbpA$  strain, together with its parental strain, J1915, were used to obtain microarray data on the differential expression of genes in  $\Delta rbpA$  relative to its parent (Table 3- Section 2.1.3.2). For each strain, four independent NMMP + 0.5% glycerol cultures were inoculated with a starting OD<sub>450</sub> of 0.05 and the RNA was isolated in early exponential phase with absorbance (OD<sub>450</sub>) readings of 0.5 to 0.6 (Section 2.2.4). The RNA purification protocol included a final QIAGEN RNeasy clean-up step to obtain highly purified RNA. The RNA was quantified with a Bioanalyzer Prokaryote Total RNA Nano chip (Agilent) and only RNA having a RNA Integrity Number of 6.5 or above was labelled. Hybridisations were performed and analysed by the Surrey Microarray Group (G. Bucca, E. Laing and C.P. Smith, University of Surrey, UK) using custom-designed 4 x 44,000 Agilent Technology slides. In each of the four hybridisations, Cy3- or Cy5-labelled cDNA from S101 was mixed with Cy3- or Cy5-labelled cDNA from J1915 (In collaboration with M. Paget., University of Sussex).

The data was filtered based on quality (as assessed by the Agilent Feature Extraction software), such that probes that did not have good quality data were filtered out. Genes were then filtered out of the data analysis if they did not have a value in each of the four arrays, in order to retain the dye balance in the experiment.

Genes that were differentially expressed in  $\Delta rbpA$  relative to w/t were identified by performing a Rank products analysis (Breitling *et al.*, 2004) (E. Laing, University of Surrey). This approach looks at conservation of rank in the data set rather than the level. The genes with a pfp value of less than 0.15 were chosen as they had more than 95% probability that the difference in expression had not occurred by chance. A table was generated of the genes down-regulated and up-regulated in S101 compared to J1915 (Table A2 & A3- Appendices).



### 3.2.2.1 Down-regulated genes

In S101 ( $\Delta rbpA$ ), it was found that 59 genes were down-regulated relative to w/t. Of these genes, many had unknown functions and were assigned as hypothetical proteins (Table A2- Appendix). Some of the genes with assigned functions include genes involved in cold shock response, amino acid synthesis, amino-acylated tRNA synthesis and ribosomal protein synthesis, which are all involved in protein synthesis (Table 9). Although these data might indicate that the genes are regulated by *rbpA*, it is also possible that the genes are subject to growth rate control, since  $\Delta rbpA$  mutants have a decreased growth rate.

**Table 9: Genes that are downregulated in *S. coelicolor*  $\Delta rbpA$  strain, S101, relative to wild-type strain, J1915.**

Gene Name	Inv. log <sub>2</sub>	Description
SCO0527	0.47	Cold shock protein: involved in cell survival at low temperatures
*SCO1578	0.44	Acetylglutamate kinase: has a role in L-arginine synthesis
*SCO1579	0.50	Bifunctional ornithine acetyltransferase/N-acetylglutamate synthase protein: has a role in L-arginine synthesis
*SCO2628	0.56	Amino acid permease: amino acid transmembrane transporter
SCO3731	0.50	Cold shock protein involved in cell survival at low temperatures
*SCO3795	0.47	Aspartyl-tRNA synthetase: catalyses attachment of aspartate to tRNA
*SCO3961	0.42	Seryl-tRNA synthetase: catalyses attachment of serine to tRNA
*SCO4078	0.49	Phosphoribosyl formylglycinamide synthase I: involved in purine metabolism
*SCO4079	0.62	Phosphoribosylformylglycinamide synthase subunit II: involved in purine metabolism
SCO4089	0.59	Valine dehydrogenase: involved in L-valine degradation
SCO4652	0.74	50S ribosomal protein L10: formation of ribosome
*SCO4660	0.58	30S ribosomal protein S7: formation of ribosome
*SCO4716	0.7	30S ribosomal protein S8: formation of ribosome
*SCO4717	0.76	50S ribosomal protein L6: formation of ribosome
*SCO4718	0.72	50S ribosomal protein L18: formation of ribosome
*SCO4719	0.69	30S ribosomal protein S5: formation of ribosome
*SCO4720	0.74	50S ribosomal protein L30: formation of ribosome
*SCO4721	0.73	50S ribosomal protein L15: formation of ribosome
SCO5209	0.49	tetR-family transcriptional regulator: controls TetR, a repressor of antibiotic biosynthesis, efflux pumps and osmotic stress
*SCO5680	0.57	Cytidine deaminase: Pyrimidine metabolism
*SCO7036	0.53	Argininosuccinate synthase: involved in L-arginine biosynthesis

\*: represent genes expressed in an operon.

The table lists the gene names of 21 genes that had a pfp value < 0.15, therefore, there is >95% probability that the difference in expression has not occurred by chance.

Gene names and functions were obtained from UniProtKB.

#### 3.2.2.2 Up-regulated genes

It was found that approximately 200 genes were up-regulated in S101 compared to J1915 (Table A3– Appendices). Many of these genes have unknown functions and thus, were allocated as hypothetical proteins. The genes that were up-regulated in S101 relative to J1915 with predicted or known function included those involved in glycerol metabolism, *Streptomyces* differentiation and sporulation, antibiotic actinorhodin production, antibiotic RED production, gas vesicle synthesis and the production of stress proteins (Table 10). These results are consistent with the fact that  $\Delta rbpA$  over-produce actinorhodin (blue) (Fig. 3.4A) and also RED antibiotic (this was observed when grown in YEME liquid media).

**Table 10: Genes that are upregulated in *S. coelicolor*  $\Delta rbpA$  strain, S101, relative to wild-type strain, J1915.**

Gene Name	Inv. $\log_2$	Description
SCO0409	2.3	spore-associated protein precursor
SCO0600	1.64	RNA polymerase sigma factor: sig8
*SCO0650	1.85	gas vesicle synthesis-like protein
*SCO1659	2.11	glycerol uptake facilitator protein: glycerol metabolism and transport
*SCO1660	1.71	glycerol kinase: Key enzyme in the regulation of glycerol uptake and metabolism
*SCO1661	1.63	glycerol-3-phosphate dehydrogenase: involved in glycerol-3-phosphate metabolism
SCO2113	2.24	bacterioferritin: iron storage protein
*SCO2828	1.64	probable amino acid ABC transporter protein, solute-binding component
SCO3034	2.53	sporulation regulatory protein: WhiB
SCO3090	2.51	ABC transporter integral membrane protein
*SCO3110	1.64	ABC transport system integral membrane protein
*SCO3111	2.93	ABC transport system ATP-binding protein
SCO3202	2.27	RNA polymerase sigma factor: $\sigma^{\text{HrdD}}$
SCO3323	1.95	RNA polymerase sigma protein: $\sigma^{\text{BidN}}$ involved in formation of aerial hyphae
SCO3397	1.67	integral membrane lysyl-tRNA synthetase: catalyses attachment of lysine to tRNA
*SCO4148	1.63	ABC transport system ATP-binding protein
SCO4277	1.74	tellurium resistance protein: responds to bacterial stress
*SCO4562	1.82	NADH-quinone oxidoreductase $\alpha$ subunit: transfers electrons from NADH to quinines in respiratory chain through FMN and Fe-S centers
*SCO4585	1.77	ABC transporter ATP-binding protein
SCO4635	1.71	50S ribosomal protein L33: formation of ribosome
*SCO5080	2.04	hydrolase: in actinorhodin cluster
*SCO5084	2.18	hypothetical protein: in actinorhodin cluster
*SCO5085	4.14	actinorhodin cluster activator protein: <i>actII-ORF4</i>
SCO5086	1.77	ketoacyl reductase: in actinorhodin cluster
*SCO5088	2.78	actinorhodin polyketide $\beta$ -ketoacyl synthase $\beta$ subunit: in actinorhodin cluster
*SCO5091	1.85	cyclase: in actinorhodin cluster
*SCO5130	1.92	ABC transporter integral membrane protein

Gene Name	Inv. Log <sub>2</sub>	Description
*SCO5877	2.11	transcriptional regulator RedD: in RED cluster
*SCO5885	2.83	hypothetical protein: in RED cluster
*SCO6500	1.72	gas vesicle synthesis-like protein: might aid in spore buoyancy
*SCO6501	1.76	gas vesicle synthesis protein: gas vesicle organisation
*SCO6502	3.04	gas vesicle synthesis protein
*SCO6815	1.76	Bi-domain ABC transporter permease protein
*SCO6816	2.98	ABC transporter binding lipoprotein
SCO7325	1.86	Stage II sporulation protein: belongs to anti-sigma factor antagonist

\*: represent genes expressed in an operon.

The table lists the gene names of 35 genes that had a pfp value <0.15, therefore, there is > 95% probability that the difference in expression has not occurred by chance.

Gene names and function obtained from UniProtKB.

### 3.2.3 The level of $\sigma^{\text{HrdB}}$ during growth of *S. coelicolor*

As RbpA was shown to act by activating transcription of  $\sigma^{\text{HrdB}}$ -dependent genes (P. Doughty and M. Paget, personal communication), it was decided to determine the level of  $\sigma^{\text{HrdB}}$  in both J1981 (w/t) and S129 ( $\Delta rbpA::aac(3)/V$ ) strains. J1981 and S129 spores were pre-germinated for 3 h and 4 h, respectively, and inoculated with a starting OD<sub>450</sub> of 0.06 in YEME containing 10% (w/v) sucrose. J1981 and S129 were incubated at 30°C with vigorous shaking for 8 h and 19 h, respectively, before entering exponential phase. At specific time points, samples were collected from each culture in duplicate, and one sample was used to determine the total protein concentration using BCA Reagent (Section 2.4.2) whilst the other subject to SDS polyacrylamide gel analysis (Section 2.4.1).

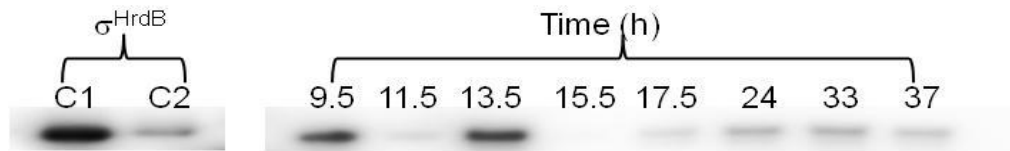
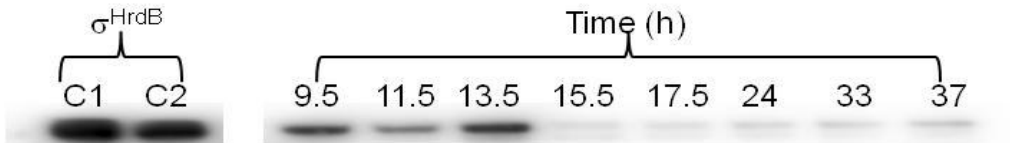
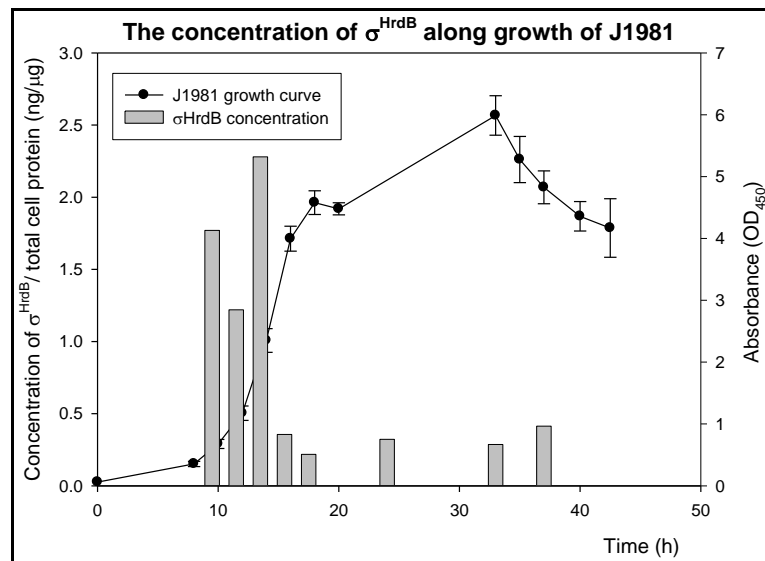
#### 3.2.3.1 J1981

Duplicate samples of 200 to 500  $\mu\text{l}$  of J1981 were taken from early exponential phase to mid-stationary phase at time points of 9.5 (OD<sub>450</sub> 0.66), 11.5 (OD<sub>450</sub>

2.2), 13.5 (OD<sub>450</sub> 3.6), 15.5 (OD<sub>450</sub> 4.3), 17.5 (OD<sub>450</sub> 5.3), 24 (OD<sub>450</sub> 5.5), 33 (OD<sub>450</sub> 6.6), and 37 h (OD<sub>450</sub> 5.4). The samples were prepared for western blotting as explained in Section 2.4.1. Samples of 8.7 µg total protein were separated on SDS polyacrylamide gel together with 10 and 20 ng of pure  $\sigma^{\text{HrdB}}$ .

The proteins on the gel were then transferred onto nitrocellulose membrane and probed with the primary anti- $\sigma^{\text{HrdB}}$  antibody SK3523 (Section 2.4.3). The antibody was used at a dilution of 1:6000 and left overnight at 4°C with gentle rocking. Anti-Rabbit HRP (Sigma) was used as the secondary antibody and diluted into blocking solution to 1:1000 (Section 2.4.4). The bands were detected using ECL detection Kit (GE Healthcare) and membrane developed using DNR Bio-Imaging Systems F-chemiBIS 3.2 M machine and visualised by the Gel capture software (Section 2.4.5).

A graph of J1981 growth curve together with the concentration of  $\sigma^{\text{HrdB}}$  at specific times was plotted using SigmaPlot (Fig. 3.5C). The bands were quantified and normalised to the 20 ng  $\sigma^{\text{HrdB}}$  control using ImageJ software. It was evident that the concentration of  $\sigma^{\text{HrdB}}$  was high during the exponential phase, that is from 9.5 to 13.5 h (OD<sub>450</sub> 0.66 to 3.6), then decreased significantly towards the late exponential phase, 15.5 h (OD<sub>450</sub> 4.3), and remained relatively constant during the stationary phase, 17.5 to 37 h (OD<sub>450</sub> 5.3 to 5.4) (Fig. 3.5A & B).

**A****B****C**

**Figure 3.5: The concentration of  $\sigma^{\text{HrdB}}$  during growth of J1981.**

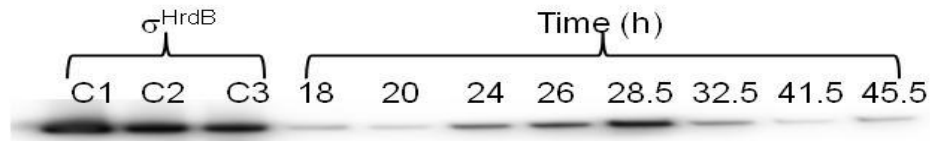
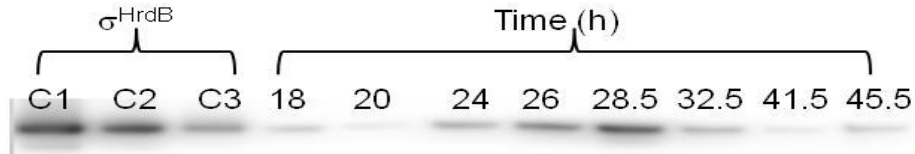
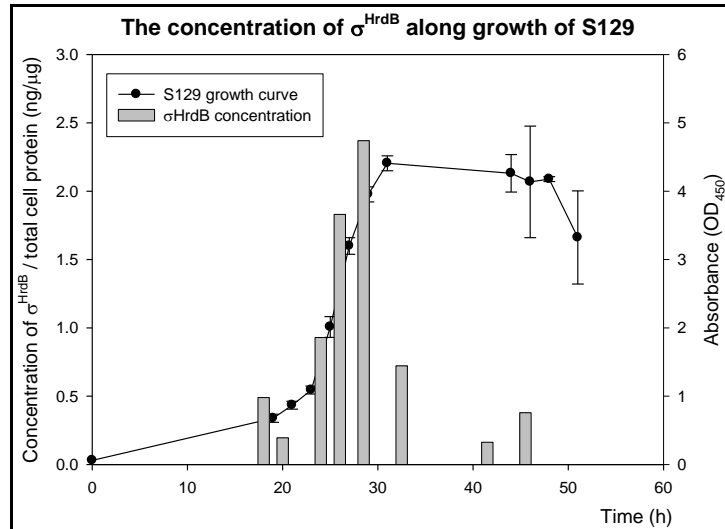
**(A)** 4-12% Bis-Tris polyacrylamide gel showing the level of  $\sigma^{\text{HrdB}}$  of the 8 samples collected along J1981 growth at: 9.5, 11.5, 13.5, 15.5, 17.5, 24, 33 and 37 h. C1 (20 ng) and C2 (10 ng) are  $\sigma^{\text{HrdB}}$  controls. The gel was separated at 120 V for 80 min and proteins were transferred onto nitrocellulose membrane at 100 V for 90 min. **(B)** Same as **A** to serve as duplicate. **(C)** A graphical representation of the concentration of  $\sigma^{\text{HrdB}}$  relative to the total cell protein (ng/ $\mu$ g) present at specific time points along the growth curve of J1981 plotted using SigmaPlot. The second set of samples from **B** were plotted to give a general representation. These were normalised to the 20 ng  $\sigma^{\text{HrdB}}$  control (C1).

### 3.2.3.2 S129

S129 samples of 200 to 500  $\mu\text{l}$  were collected in duplicate along a growth curve from early exponential phase to mid-stationary phase at specific time points of 18 ( $\text{OD}_{450}$  0.61), 20 ( $\text{OD}_{450}$  1.35), 24 ( $\text{OD}_{450}$  2.04), 26 ( $\text{OD}_{450}$  2.5), 28.5 ( $\text{OD}_{450}$  3.2), 32.5 ( $\text{OD}_{450}$  3.3), 41.5 ( $\text{OD}_{450}$  3.34), and 45.5 h ( $\text{OD}_{450}$  3.6). The samples were prepared for western blotting as explained in Section 2.4.1.

A graph of S129 growth curve together with the concentration of  $\sigma^{\text{HrdB}}$  at specific times was plotted using SigmaPlot (Fig. 3.6C). The bands were quantified and normalised to the 20 ng  $\sigma^{\text{HrdB}}$  control using ImageJ software. It was apparent that the concentration of  $\sigma^{\text{HrdB}}$  was low during the early exponential phase, that is from 18 to 20 h ( $\text{OD}_{450}$  0.61 to 1.35), then increased significantly towards the late exponential phase, 24 to 28 h ( $\text{OD}_{450}$  2.04 to 3.2), and finally decreased before remaining relatively constant during the stationary phase, 32.5 to 45.5 h ( $\text{OD}_{450}$  3.3 to 3.6) (Fig. 3.6A & B).



**A****B****C**

**Figure 3.6: The concentration of  $\sigma^{\text{HrdB}}$  during growth of S129.**

**(A)** 4-12% Bis-Tris polyacrylamide gel showing the level of  $\sigma^{\text{HrdB}}$  during a S129 growth curve at 18, 20, 24, 26, 28.5, 32.5, 41.5, and 45.5 h. C1 (40 ng), C2 (20 ng), and C3 (10 ng) are  $\sigma^{\text{HrdB}}$  controls. The gel was separated at 120 V for 80 min and proteins were transferred onto nitrocellulose membrane at 100 V for 90 min. **(B)** Same as **A** to serve as duplicate. **(C)** A graphical representation of the concentration of  $\sigma^{\text{HrdB}}$  relative to the total cell protein (ng/ $\mu$ l) present at specific time points along the growth curve of S129 plotted using SigmaPlot. The second set of samples from **B** were plotted to give a general representation. These were normalised to the 20 ng  $\sigma^{\text{HrdB}}$  control (C2).

### 3.2.4 RbpA in holoenzyme formation

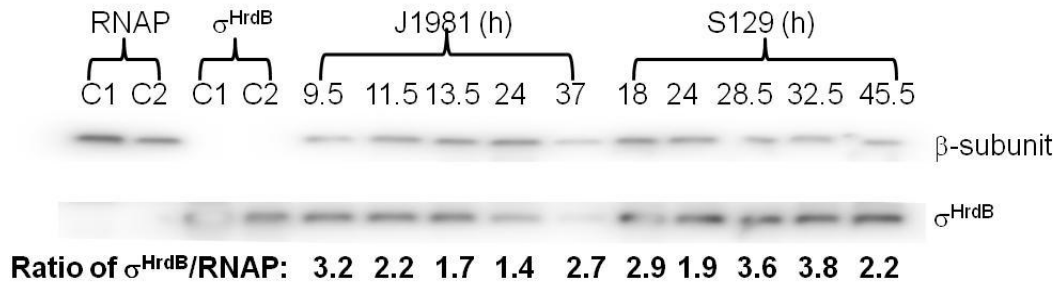
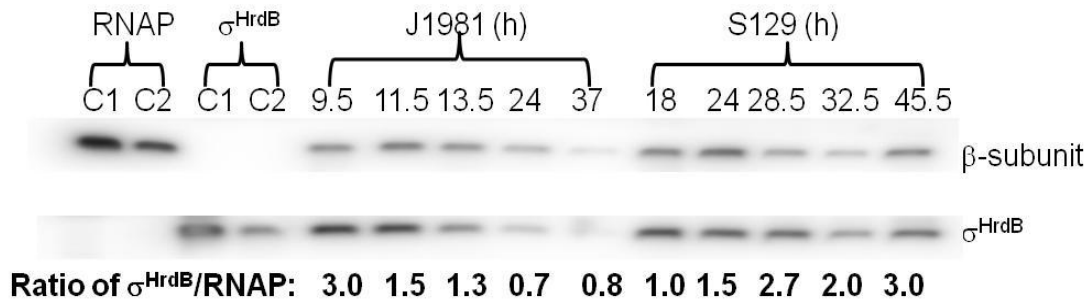
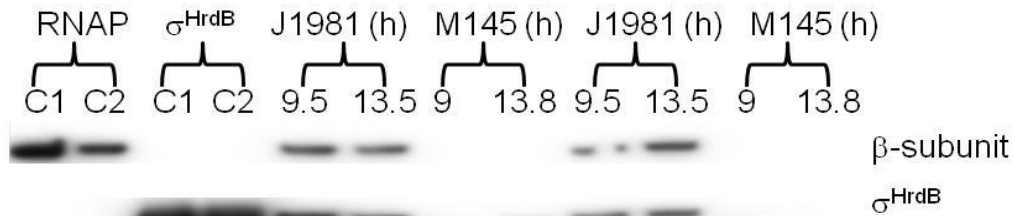
RbpA appears to be involved in an early stage of transcription initiation (P. Doughty and M. Paget, personal communication), therefore it was decided to test the role of RbpA in recruiting  $\sigma^{\text{HrdB}}$  to bind to the core RNAP, that is the formation of the holoenzyme. The effect of RbpA on holoenzyme formation was tested *in vivo* using J1981 (w/t) and S129 ( $\Delta rbpA::aac(3)IV$ ) *S. coelicolor* strains. J1981 and S129 were pre-germinated with a starting OD<sub>450</sub> of 0.06 for 3 to 4 h, respectively, and then inoculated in YEME containing 10% (w/v) sucrose. These were left to grow at 30°C with vigorous shaking and samples of 2 to 8 ml were collected at five specific time points; early exponential, mid-exponential, late-exponential, early stationary and mid-stationary phase. The J1981 samples were taken at 9.5 (OD<sub>450</sub> 0.7), 11.5 (OD<sub>450</sub> 2), 13.5 (OD<sub>450</sub> 3.6), 24 (OD<sub>450</sub> 5.5), and 37 h (OD<sub>450</sub> 5.4), whilst the S129 samples were taken at 18 (OD<sub>450</sub> 0.6), 24 (OD<sub>450</sub> 2.0), 28.5 (OD<sub>450</sub> 3.2), 32.5 (OD<sub>450</sub> 3.3), and 45.5 h (OD<sub>450</sub> 3.6). The protein concentration was determined by collecting 200 to 500 µl samples at the exact time points, followed by performing BCA assay (Section 2.3.5.2).

The samples were immediately centrifuged at 16,100 x g for 30 s at 4°C and the pellet was resuspended in 1 ml ice-cold 0.9% (w/v) NaCl and again centrifuged. The pellets were placed on dry ice to immediately freeze. To test whether RbpA affects  $\sigma^{\text{HrdB}}$  recruitment to core RNAP, the appropriate concentration of the samples were passed through Ni-NTA spin column (Qiagen) and protocol followed as per manufacturer's instruction (Section 2.3.2.3). J1981 and S129 have His-tagged *rpoC*, therefore the  $\beta'$ -subunit of RNAP attaches to the Ni-NTA column bringing any bound  $\sigma^{\text{HrdB}}$  with it. The flow-through was kept to ensure the final protein concentration loaded on the gel was equal (data not included). The eluate samples were normalised and separated by SDS polyacrylamide gel together with *S. coelicolor* core RNAP and  $\sigma^{\text{HrdB}}$  as controls.

The proteins were then transferred onto nitrocellulose membrane. The top part of the membrane was incubated with mouse monoclonal anti-RNAP  $\beta$  subunit

(Abcam Ltd.) primary antibody diluted with blocking solution to 1:5000 with gentle rocking overnight at 4°C. It was then incubated with secondary anti-Mouse HRP (Sigma) diluted to 1:10000 in blocking solution for 1 h with gentle rocking. The bottom half of the membrane was incubated with anti- $\sigma^{\text{HrdB}}$  SK3523 diluted with blocking solution to 1:6000 overnight at 4°C. It was then incubated with secondary anti-Rabbit HRP (Sigma) at 1:1000 in blocking solution for 1 h (Section 2.4.4). The bands on the membrane were then detected using ECL detection kit (GE Healthcare), developed by DNr Bio-Imaging Systems F-chemiBIS 3.2 M machine and visualised by the Gel capture software (Section 2.4.5). Four cultures were tested in duplicates to ensure accurate results.

The level of  $\sigma^{\text{HrdB}}$  in relation to  $\beta$ -subunit of RNAP was not higher in J1981 compared to S129 (Fig. 3.7A & B). M145 was used as the negative control for J1981 as it does not have His-tagged *rpoC*. Two time points were selected 9 (OD<sub>450</sub> 0.78) and 13.8 h (OD<sub>450</sub> 3.9) and were duplicated. No RNAP  $\beta$ -subunit or  $\sigma^{\text{HrdB}}$  were detected on the membrane of M145 samples as was expected (Fig. 3.7C), suggesting that all  $\sigma^{\text{HrdB}}$  detected was bound to RNAP. The  $\sigma^{\text{HrdB}}$  and RNAP  $\beta$ -subunit bands were quantified using ImageJ and normalised to the 10 ng  $\sigma^{\text{HrdB}}$  (C2) and 10 ng RNAP (C2), respectively. The ratio of  $\sigma^{\text{HrdB}}$  to RNAP was then calculated at each time point (Fig. 3.7A & B) and it was found that this ratio was not higher in J1981 compared to S129. Thus, RbpA appears to have no effect in recruiting  $\sigma^{\text{HrdB}}$  to RNAP, that is in holoenzyme formation.

**A****B****C**

**Figure 3.7: The level of  $\sigma^{\text{HrdB}}$  relative to RNAP in J1981 and S129.**

**(A)** 4-12% Bis-Tris polyacrylamide gel showing the 5 samples collected along J1981 and S129 growth curves. The J1981 samples were taken at 9.5, 11.5, 13.5, 24, and 37 h whilst the S129 samples were taken at 18, 24, 28.5, 32.5, and 45.5 h. RNAP controls: C1 (30 ng) and C2 (10 ng) and  $\sigma^{\text{HrdB}}$  controls: C1 (20 ng) and C2 (10 ng). The gel was separated at 120 V for 80 min and proteins were transferred onto nitrocellulose membrane at 100 V for 90 min. C2 of both RNAP and  $\sigma^{\text{HrdB}}$  were used for normalising RNAP and  $\sigma^{\text{HrdB}}$  samples, respectively, using ImageJ. **(B)** Same as **A** to serve as duplicate. **(C)** M145 served as the control of J1981 with samples taken at two time points: 9 ( $\text{OD}_{450}$  0.78) and 13.8 ( $\text{OD}_{450}$  3.9).

### **3.3 Discussion**

#### **3.3.1 Actinobacteria lacking RbpA homologues**

RbpA was shown to be present in almost all the Actinobacteria; however, it is absent from a few Actinobacteria (Section 3.1.1 and Table A1 - Appendices). These organisms were then checked whether their genome was completely sequenced using NCBI database. Out of the 19 strains, 18 were fully sequenced, examples include: *Acidimicrobium ferrooxidans*, *Atopobium parvulum*, *Collinsella aerofaciens*, *Conexibacter woesei*, *Cryptobacterium curtum*, *Eggerthella lenta*, *Olsenella uli*, *Rubrobacter xylanophilus*, *Slackia heliotrinireducens* RHS 1, and *Tropheryma whippiei*.

Most of these organisms belong to the family of *Coriobacteriaceae*, and are found in the human oral cavity, e.g.: *Atopobium parvulum* (Copeland *et al.*, 2009), *Cryptobacterium curtum* (Mavrommatis *et al.*, 2009), and *Olsenella uli* (Goker *et al.*, 2010) or in the intestinal microflora e.g.: *Eggerthella lenta* (Saunders *et al.*, 2009) and *Slackia heliotrinireducens* RHS 1 (Pukall *et al.*, 2009). These organisms are either pathogenic or reside within the host and therefore, may rely on the host for survival. The GC content of these organisms varied from 72.7% for *Conexibacter woesei* to 46.3% for *Tropheryma whippiei* suggesting that the GC content may not be a factor in the presence or absence of RbpA.

An interesting organism, *Tropheryma whippiei* Twist, is an intracellular human pathogen that has a reduced genome of only 927,303bp and encodes 808 predicted protein-coding genes. It lacks many of the genes involved in amino acid metabolism, and possibly orthologues of the thioredoxin pathway (Raoult *et al.*, 2003). Thus, this organism appeared to have adapted different ways of surviving within the host and therefore, has lost many genes including the *rbpA* gene. It is not known why RbpA is absent from these organisms and how these

organisms have evolved to transcribe genes without the need for RbpA (Section 4.3.3).

### 3.3.2 RbpB in *S. coelicolor*

*S. coelicolor*  $\Delta rbpA$  strains have a very evident phenotypic and morphological defect (Fig. 3.4). However, previous work showed that *S. coelicolor*  $\Delta rbpB$  strains, on the other hand, had no phenotypic effect on growth or morphological defect in colony size (Newell *et al.*, 2006) suggesting that RbpB is not essential. Initial attempts to obtain a knock-out of both *rbpA* and *rbpB* have not yet been successful. This might be due to the fact that RbpB might be taking over some of the roles of RbpA and consequently when both are mutated then the bacterium cannot survive as either one of these genes is essential for survival (Section 4.4.3). Further work to obtain an *S. coelicolor* strain with both *rbpA/rbpB* knock-out is in progress (R. Lewis and M. Paget, work in progress).

### 3.3.3 Gene expression in *S. coelicolor* $\Delta rbpA$ mutant

From the microarray data, it was identified that genes involved in amino acid synthesis, amino-acylated tRNA synthesis, and ribosomal proteins were all down-regulated in *S. coelicolor*  $\Delta rbpA$  strain relative to w/t (Table 9). Many of these genes were present in operons. To determine whether they may be direct targets of RbpA or are down-regulated due to the slow growth phenotype of *S. coelicolor*  $\Delta rbpA$  strains, the genes in the different operons were all checked for their expression level.

It was found that the ribosomal genes were present in three separate operons and all three operons together with the genes within the operons were all down-regulated. Also, genes involved in amino-acyl tRNA synthesis e.g. aspartyl- and seryl-tRNA were present in operons and each gene in the operon was down-regulated. Interestingly, the whole operon for arginine synthesis, consisting of five genes, was down-regulated significantly. In addition, genes involved in purine metabolism were found to occur in an operon and their expression was

down-regulated. Does this suggest that *S. coelicolor*  $\Delta rbpA$  is an auxotroph that has lost its ability to synthesis arginine and purine which is required for growth? This is not thought to be the case as these mutants grow slowly even in a rich media such as YEME. Many of the genes down-regulated in the *rbpA* mutant are house-keeping genes and are therefore likely to be transcribed by RNAP containing  $\sigma^{\text{HrdB}}$ ; it follows that the promoters might additionally require RbpA for full activity. This may explain why  $\Delta rbpA$  strains have a reduced growth compared to w/t. Alternatively, the observed reduction in activity might be caused by an indirect effect on account of their possible growth rate control.

The antibiotic actinorhodin is also over-produced in *S. coelicolor*  $\Delta rbpA$  strains and this was further verified by the up-regulation of genes present in the *act* cluster in  $\Delta rbpA$  relative to the w/t (Table 10). One of these genes is the *actII-ORF4* and it is the main transcriptional regulator that controls production of actinorhodin. S129 grown in YEME liquid media produced RED antibiotic as the cells approached stationary phase and this observation was backed up by the microarray data, which showed that genes present in the *red* cluster were also up-regulated in the  $\Delta rbpA$  strain relative to w/t (Table 10).

### 3.3.4 $\sigma^{\text{HrdB}}$ level during growth of J1981 and S129

The expression of the  $\sigma^{\text{HrdB}}$  gene was previously shown to be relatively constant throughout growth and stationary phase in *S. coelicolor* (Kang *et al.*, 1997). This fits in with the model that  $\sigma^{\text{HrdB}}$  is required to express rRNA genes during exponential phase as well as RED and actinorhodin biosynthetic genes during stationary phase (Fujii *et al.*, 1996, Kang *et al.*, 1997). However, the level of RNAP- $\sigma^{\text{HrdB}}$  holoenzyme appeared to be abundant during exponential phase before dropping during stationary phase (Kang *et al.*, 1997). This latter observation was also observed in Fig. 3.7 where the level of RNAP- $\sigma^{\text{HrdB}}$  was highest during exponential phase in J1981 (w/t) before decreasing during stationary phase (Fig. 3.7). Particularly striking however, was the changes in the total concentration of  $\sigma^{\text{HrdB}}$ , which increased during growth to a maximum

towards the end of exponential phase before dropping again during stationary phase (Fig. 3.5). Why is the level of  $\sigma^{\text{HrdB}}$  decreased in stationary phase when its expression is constant throughout growth?

In *Bacillus subtilis*, the expression of the principal sigma factor ( $\sigma^A$ ) is constant throughout growth and development but the level of RNAP- $\sigma^A$  decreases during stationary phase as observed in *S. coelicolor* (Haldenwang *et al.*, 1981). In *E. coli*, this was also observed for  $\sigma^{70}$ , which is expressed at a relatively constant level throughout growth. During stationary phase, the level of  $\sigma^S$  rises and leads to decrease in relative level of RNAP- $\sigma^{70}$  (Jishage & Ishihama, 1995). Alternative sigma factors compete with  $\sigma^{70}$  for core RNAP, however, the affinity of  $\sigma^{70}$  for core RNAP is much higher than that for any of the other alternative sigma factors. For this reason, these alternative sigma factors might require additional regulators to be able to successfully compete for core RNAP.

Global regulators such as ppGpp, DksA and 6S RNA might play a role in sigma factor competition for core RNAP. ppGpp, DksA and 6S RNA interact with RNAP- $\sigma^{70}$  complex, decreasing  $\sigma^{70}$ -dependent transcription and thereby, allowing other alternative sigma factors to compete with  $\sigma^{70}$  for binding to core RNAP (Section 1.2.5.1, 1.2.5.2 & 1.2.6.1). Rsd, an anti-sigma factor of  $\sigma^{70}$ , has also been shown to interact with free  $\sigma^{70}$ , core RNAP and RNAP- $\sigma^{70}$  complex causing the dissociation of  $\sigma^{70}$  and allowing alternative sigma competition (Section 1.2.7.1) (Ilag *et al.*, 2004). Crl, a stationary phase protein that interacts with  $\sigma^S$ , has been shown to increase the affinity of  $\sigma^S$  for core RNAP and specifically aids  $\sigma^S$  to compete with  $\sigma^{70}$  for the core RNAP during stationary phase (Typas *et al.*, 2007).

Therefore, the decrease in the level of RNAP- $\sigma^{\text{HrdB}}$  holoenzyme at stationary phase may be due to alternative sigma competition, where factors such as ppGpp and DksA might hinder the  $\sigma^{\text{HrdB}}$ -dependent transcription of rRNA genes



causing the dissociation of  $\sigma^{\text{HrdB}}$  to allow alternative sigma factors to interact with RNAP and direct secondary metabolism. Free  $\sigma^{\text{HrdB}}$  may be unstable in the cells which would lead to lower levels of  $\sigma^{\text{HrdB}}$  being detected (Fig. 3.5).

Unlike J1981 (w/t), S129 ( $\Delta rbpA::aac(3)/V$ ) had low levels of  $\sigma^{\text{HrdB}}$  in early exponential phase (Fig. 3.6). However, the level of  $\sigma^{\text{HrdB}}$  increased in mid-exponential phase to late exponential phase before decreasing and remaining relatively constant throughout stationary phase as observed in J1981 (Fig. 3.5 & 3.6). The reduced  $\sigma^{\text{HrdB}}$  level in early exponential phase might be due to the long lag phase and slow start of the exponential growth owing to the absence of RbpA. In addition, if  $\Delta rbpA$  strain grows because RbpB is produced, then the expression of *rbpB* might limit the growth of the mutant. In support, RbpB suppresses the  $\Delta rbpA$  phenotype when overexpressed under the control of a strong promoter *ermEp*. The natural expression of *rbpB* has not been studied in detail.

### 3.3.5 RbpA in holoenzyme formation

RbpA and Crl are unusual positive regulators of transcription (Section 1.2.9.1 & 1.2.9.2). Crl, as mentioned above, is a positive regulator of  $\sigma^S$  that increases the affinity of  $\sigma^S$  for core RNAP. To determine whether RbpA was also involved in holoenzyme formation, *in vivo* analysis of holoenzyme  $\sigma^{\text{HrdB}}$  content was determined in *S. coelicolor* J1981 (w/t) and S129 ( $\Delta rbpA::aac(3)/V$ ) strains (Section 3.2.4). The results (Fig. 3.7) indicate that the ratio of  $\sigma^{\text{HrdB}}$  to RNAP (as judged by the concentration of  $\beta$ ) in J1981 and S129 was comparable throughout growth. Hence, RbpA does not appear to be involved in holoenzyme formation; recruiting free  $\sigma^{\text{HrdB}}$  and facilitating its interaction with the core RNAP. However, it remains possible that homeostatic mechanisms present *in vivo* mask an effect, and so this needs to be verified *in vitro*.

#### **4- Mapping RbpA interactions with $\sigma^{\text{HrdB}}$**

#### **4.0 Overview**

This chapter focuses on investigating possible interactions between RbpA (*S. coelicolor*) or Rv2050 (*M. tuberculosis*) with their respective sigma factors. Protein-protein interactions were monitored using the Bacterial Adenylate Cyclase-based Two-Hybrid System (BACTH) and by *in vitro* pull down assays. Previously, it was observed that RbpA interacted with the principal sigma factor,  $\sigma^{\text{HrdB}}$  (P. Doughty and M. Paget, personal communication). However, the specificity of RbpA for different sigma factors was unknown, so this was investigated by BACTH. RbpA was tested for interaction with sigma factors that are closely related to  $\sigma^{\text{HrdB}}$  as well as selected Group III and Group IV sigma factors. Finally, as RbpA was shown to interact with  $\sigma^{\text{HrdB}}$ , it was important to map the exact region on  $\sigma^{\text{HrdB}}$  that RbpA binds to.  $\sigma^{\text{HrdB}}$  was divided into five fragments based on its structural domains.

#### **4.1 BACTH analysis of sigma factors with RbpA**

BACTH exploits the principle that the *Bordetella pertussis*'s adenylate cyclase catalytic site is made up of two domains, T25 and T18. When these two domains are expressed as separate polypeptides, the enzyme is inactive because the domains have no means to associate. However, the two domains can be brought together when each is fused to one partner of a protein-protein interaction pair. This leads to the activation of adenylate cyclase, which increases cyclic AMP production. cAMP will interact with the catabolite activator protein (CAP) forming the cAMP-CAP complex. This complex is recruited to cAMP-CAP-dependent promoters, thereby regulating gene expression including the *lac* operon of *E.coli*. Activation of the *lac* operon will induce the expression of the *lacZ* gene leading to the production of  $\beta$ -galactosidase, which can then be assayed. In order to detect protein-protein interactions on a media plate, the chromogenic substrate X-gal was included and it is hydrolysed to a blue dye by  $\beta$ -galactosidase resulting in blue colonies. The detailed methodology for BACTH is described in Section 2.6.

#### 4.1.1 RbpA interaction with $\sigma^{\text{HrdB}}$

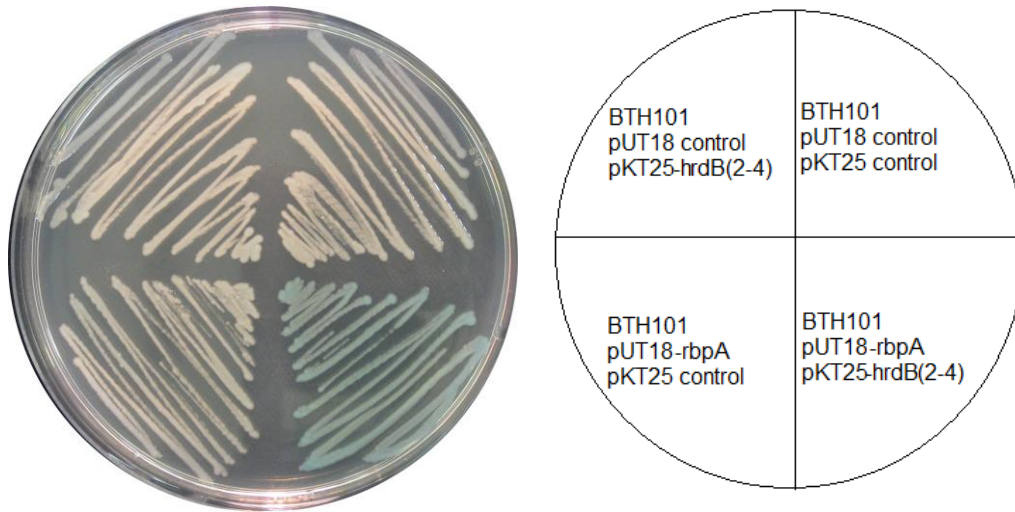
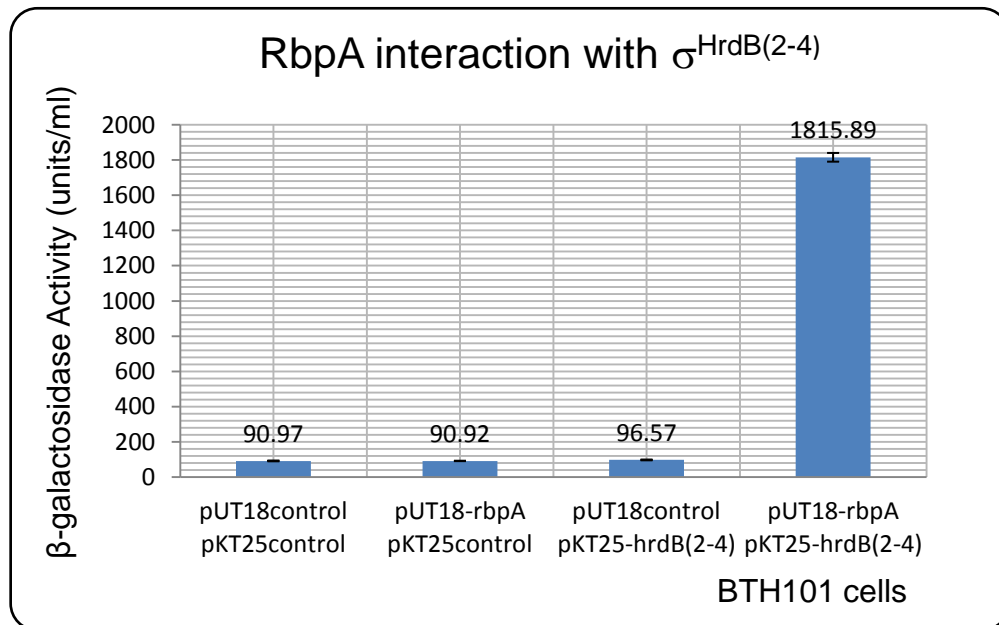
$\sigma^{\text{HrdB}}$  is the principal sigma factor of *S. coelicolor* that controls housekeeping genes and is essential for survival of the organism (Buttner *et al.*, 1990). RbpA was previously shown to activate transcription of  $\sigma^{\text{HrdB}}$ -dependent promoters using *in vitro* transcription assays (P. Doughty and M. Paget, personal communication). Furthermore, RbpA was shown to bind directly to  $\sigma^{\text{HrdB}}$  *in vitro* (P. Doughty and M. Paget, personal communication). In order to confirm these data and to develop a system for further characterising RbpA-sigma interactions, the BACTH system was applied.

It was decided to fuse RbpA to the T18 domain of adenylate cyclase in pUT18, generating a RbpA-T18 fusion, and  $\sigma^{\text{HrdB}}$  and other sigma factors to the T25 domain generating for example a T25- $\sigma^{\text{HrdB}}$  fusion. Constructs containing the entire *hrdB* gene appeared to be toxic in *E. coli*, and so a fragment lacking region 1.1 of  $\sigma^{\text{HrdB}}$  but containing domains 2 to 4 ( $\sigma^{\text{HrdB}(2-4)}$ ) was used, which was determined from the structural domain organisation shown in Fig. 4.4.

The  $\sigma^{\text{HrdB}(2-4)}$  fragment was amplified using *T25\_hrdB(2-4)* forward and reverse primers (Table 5- Section 2.1.5.1), introducing XbaI and EcoRI sites, respectively, allowing the gene to be cloned in-frame with the T25 domain when cloned into pKT25. *rbpA* was amplified with *T18\_rbpA* forward and reverse primers (Table 5- Section 2.1.5.1), introducing BamHI and EcoRI sites, respectively, and keeping the gene in frame with the T18 domain in pUT18. Since the T18 subunit is located downstream of *rbpA*, the stop codon of *rbpA* was removed. Following amplification, the PCR products were cloned into pBlueScript II SK+, sequenced, then subcloned into either pKT25 and pUT18.

The recombinant plasmids pKT25-*hrdB*<sup>(2-4)</sup> and pUT18-*rbpA* were used to co-transform BTH101 competent cells by electroporation. Controls were also prepared containing vector only, pKT25-*hrdB*<sup>(2-4)</sup> alone, or pUT18-*rbpA* alone. Transformants were selected on LA plates with additional X-gal and IPTG for an

initial assessment (Fig. 4.1A). BTH101 containing both pUT18-*rbpA* and pKT25-*hrdB*<sup>(2-4)</sup> produced blue colonies, whereas all negative controls produced white colonies. To quantify and confirm these interactions  $\beta$ -galactosidase assays were performed (Section 2.6) (Fig. 4.1B). BTH101 containing both pUT18-*rbpA* and pKT25-*hrdB*<sup>(2-4)</sup> exhibited 20-fold higher  $\beta$ -galactosidase activity compared to the controls, indicating that RbpA interacts with  $\sigma^{\text{HrdB}(2-4)}$  in this assay (Fig. 4.1).

**A****B**

**Figure 4.1: BACTH assay of RbpA interaction with principal sigma factor,  $\sigma^{\text{HrdB}(2-4)}$ , of *S. coelicolor*.**

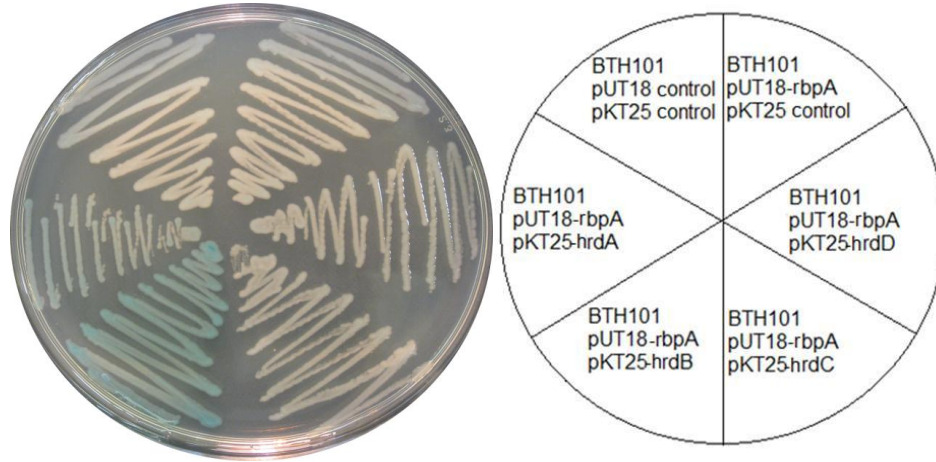
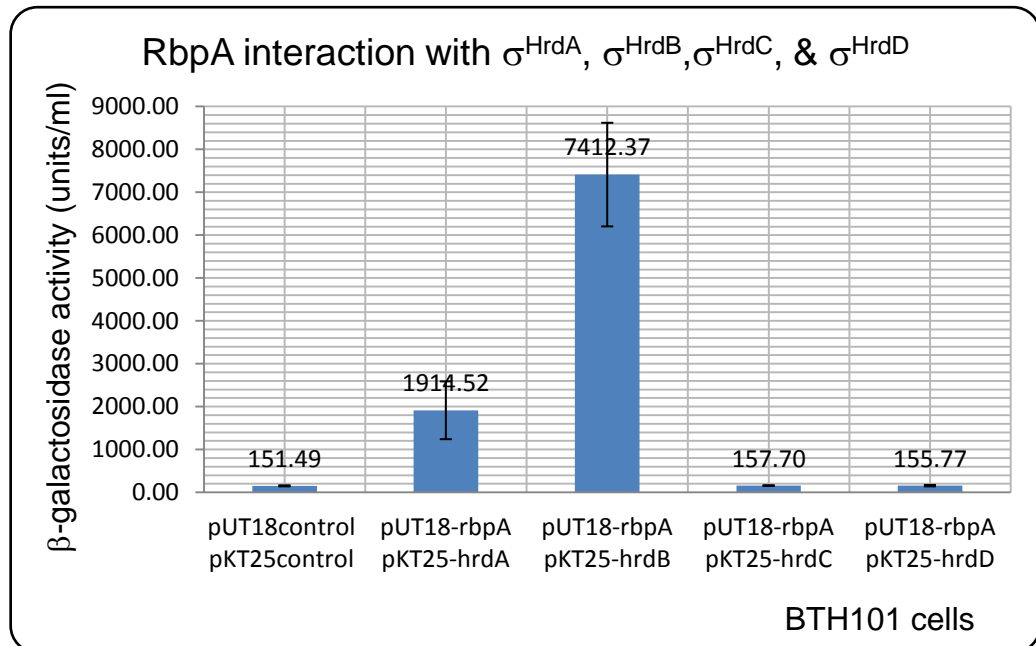
**(A)** Blue/White screening on LA plates left to grow at 30°C for 16-24 h. **(B)**  $\beta$ -galactosidase assay to determine RbpA interaction with  $\sigma^{\text{HrdB}(2-4)}$ .  $\beta$ -galactosidase activity was measured for BTH101 containing vector only, pUT18-*rbpA* only, pKT25-*hrdB*<sup>(2-4)</sup> only and pUT18-*rbpA* pKT25-*hrdB*<sup>(2-4)</sup>. This assay was repeated in triplicate with error bars showing standard deviation.

#### 4.1.2 RbpA interaction with Group II

The interaction of RbpA with the principal sigma factor,  $\sigma^{\text{HrdB}}$ , raises the question of whether RbpA is a sigma-specific transcription factor, thereby influencing certain holoenzymes, or whether it binds to all sigma factors. It was therefore, important to test for possible interactions between RbpA and different sigma factors. Initially, we tested the three sigma factors that are most similar to  $\sigma^{\text{HrdB}}$ :  $\sigma^{\text{HrdA}}$ ,  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$ . Each of these is non-essential and shows extensive amino acid sequence similarity between domains 2 to 4, and so the domains 2-4 of each were tested for interaction with RbpA.

The conserved domains 2-4 regions of *hrdA*, *hrdC* and *hrdD* were amplified and cloned into pKT25 as described for *hrdB* (see above) using forward and reverse primers *T25\_hrdA(2-4)*, *T25\_hrdC(2-4)* and *T25\_hrdD(2-4)*, respectively (Table 5- Section 2.1.5.1), introducing flanking XbaI and EcoRI sites and keeping the genes in frame when fused to T25. pKT25-*hrdA*, pKT25-*hrdC* and pKT25-*hrdD* were electroporated together with pUT18-*rbpA* into BTH101 competent cells. BTH101 control was also prepared containing vector only.

These transformants were plated on LA plates with X-gal and IPTG for preliminary assessment (Fig. 4.2A). BTH101 containing pUT18-*rbpA* and pKT25-*hrdA* or pKT25-*hrdB* produced blue colonies. However, BTH101 containing pUT18-*rbpA* and pKT25-*hrdC* or pKT25-*hrdD* together with negative controls produced white colonies. To quantify and confirm these interactions  $\beta$ -galactosidase assays were performed (Section 2.6) (Fig. 4.2B). BTH101 containing pUT18-*rbpA* and pKT25-*hrdA* or pKT25-*hrdB* had 12-fold and 50-fold increase in activity of  $\beta$ -galactosidase, respectively, compared to the control. These results indicate that RbpA interacts with  $\sigma^{\text{HrdA}(2-4)}$  and  $\sigma^{\text{HrdB}(2-4)}$  but does not interact with  $\sigma^{\text{HrdC}(2-4)}$  and  $\sigma^{\text{HrdD}(2-4)}$  in this assay (Fig.4.2).

**A****B**

**Figure 4.2: BACTH assay of RbpA interaction with domain 2-4 of Group I and II sigma factors,  $\sigma^{\text{HrdA}}$ ,  $\sigma^{\text{HrdB}}$ ,  $\sigma^{\text{HrdC}}$ ,  $\sigma^{\text{HrdD}}$ , of *S. coelicolor*.**

**(A)** Blue/White screening on LA plates grown at 30°C for 16-24 h. **(B)**  $\beta$ -galactosidase assay of RbpA interaction with  $\sigma^{\text{HrdA}}$ ,  $\sigma^{\text{HrdB}}$ ,  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$ .  $\beta$ -galactosidase assay for BTH101 cells containing vector only, pUT18-rbpA and pKT25-hrdA, pKT25-hrdB, pKT25-hrdC and pKT25-hrdD were measured. These assays were repeated using three separate colonies in triplicate with error bars showing standard deviation.



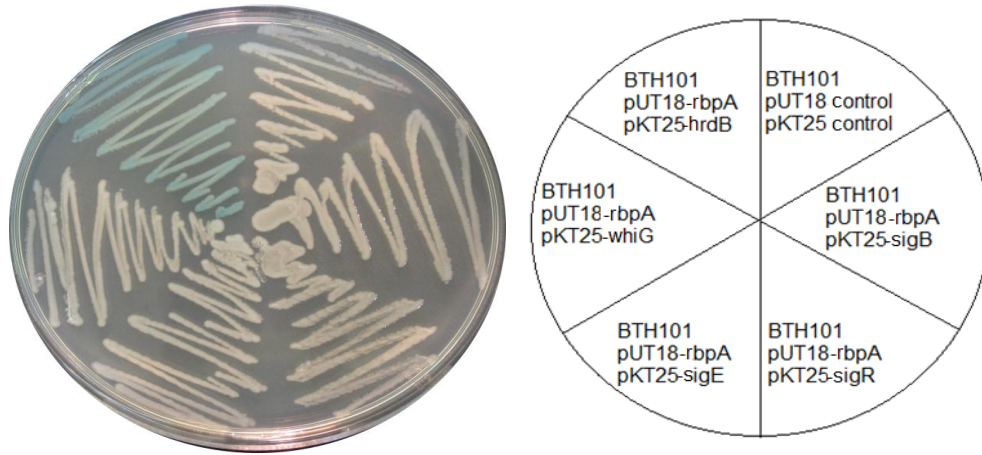
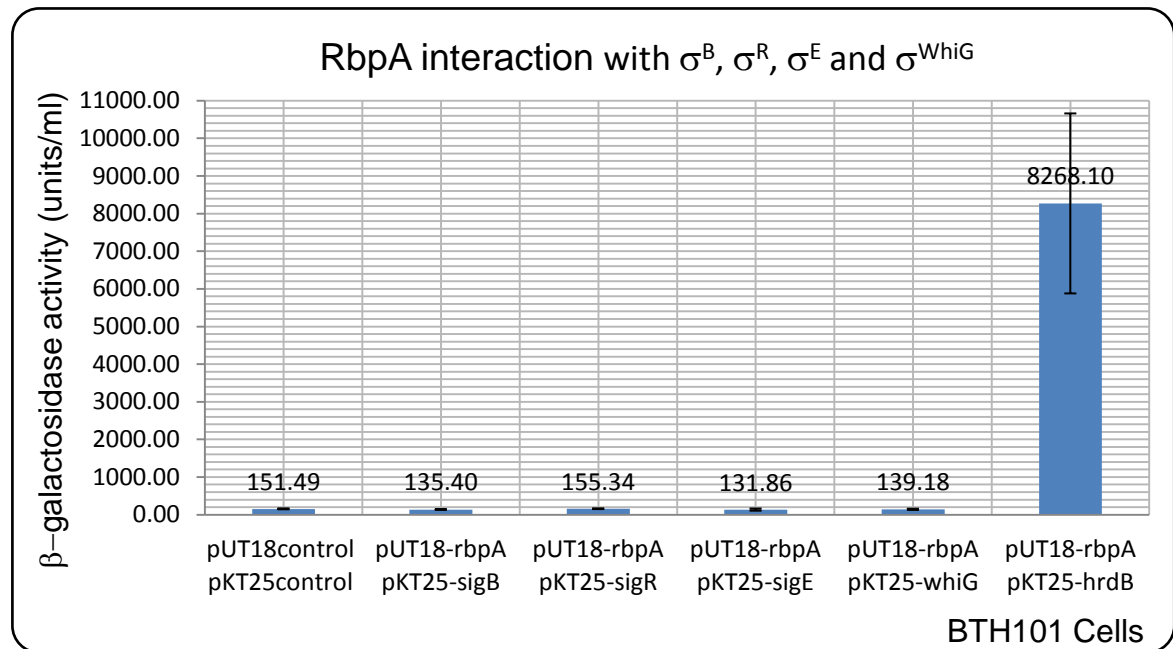
### 4.1.3 RbpA interaction with Group III and IV

In addition to the Group I (principal) and Group II (principal-related) sigma factors, there are two remaining Groups (Groups III and IV) that include alternative sigma factors. Group III includes sigma factors closely related to  $\sigma^B$  of *B. subtilis*, as well as  $\sigma^{WhiG}$ , which is phylogenetically distinct from other *S. coelicolor* sigma factors, resembling the flagella-related sigma factors such as FliA (Helmann, 1991). Group IV contains approximately 49 ECF sigma factors of *S. coelicolor* (Paget & Helmann, 2003). The sigma factors selected to test for possible interaction with RbpA were the Group III sigma factors,  $\sigma^B$  and  $\sigma^{WhiG}$ , and the Group IV sigma factors  $\sigma^R$  and  $\sigma^E$ .  $\sigma^B$  controls general stress response and is a major regulator in oxidative and osmotic stress (Viollier *et al.*, 2003).  $\sigma^{WhiG}$  is a sporulation sigma factor that triggers the onset of sporulation in aerial hyphae (Tan *et al.*, 1998).  $\sigma^R$  is involved in an oxidative stress response (Paget *et al.*, 1998), whilst  $\sigma^E$  is essential for maintaining the integrity of the cell envelope (Paget *et al.*, 1999).

The four different sigma factor genes *sigB*, *sigR*, *sigE* and *whiG* were amplified using *T25\_sigBSc*, *T25\_sigR*, *T25\_sigE* and *T25\_whiG* forward and reverse primers (Table 5- Section 2.1.5.1), respectively, introducing XbaI and EcoRI sites and keeping the genes in frame when fused to T25. The inserts were cloned into pBlueScript II SK+, sequenced, then sub-cloned into pKT25. Recombinants pKT25-*sigB*, pKT25-*sigR*, pKT25-*sigE* and pKT25-*whiG* were electroporated together with pUT18-*rbpA* into BTH101 competent cells. Positive and negative control strains were also prepared (see above). These were streaked on LA plates with additional selection with X-gal and IPTG.

BTH101 containing pUT18-*rbpA* and pKT25-*sigB*, pKT25-*sigR*, pKT25-*sigE* and pKT25-*whiG* produced white colonies as the negative control. However, the positive control BTH101 containing pUT18-*rbpA* and pKT25-*hrdB* produced blue colonies (Fig. 4.3A). To confirm this result the  $\beta$ -galactosidase assays were performed (Section 2.6). From the assays, there was no observed increase in  $\beta$ -

galactosidase activity for BTH101 pUT18-*rbpA* and pKT25-*sigB*, pKT25-*sigR*, pKT25-*sigE* and pKT25-*whiG* (Fig. 4.3B). These assays indicate that RbpA does not interact with  $\sigma^B$ ,  $\sigma^R$ ,  $\sigma^E$  and  $\sigma^{WhiG}$ .

**A****B**

**Figure 4.3: BACTH assay of RbpA interaction with Group III and Group IV sigma factors:  $\sigma^B$ ,  $\sigma^R$ ,  $\sigma^E$  and  $\sigma^{WhiG}$ , of *S. coelicolor*.**

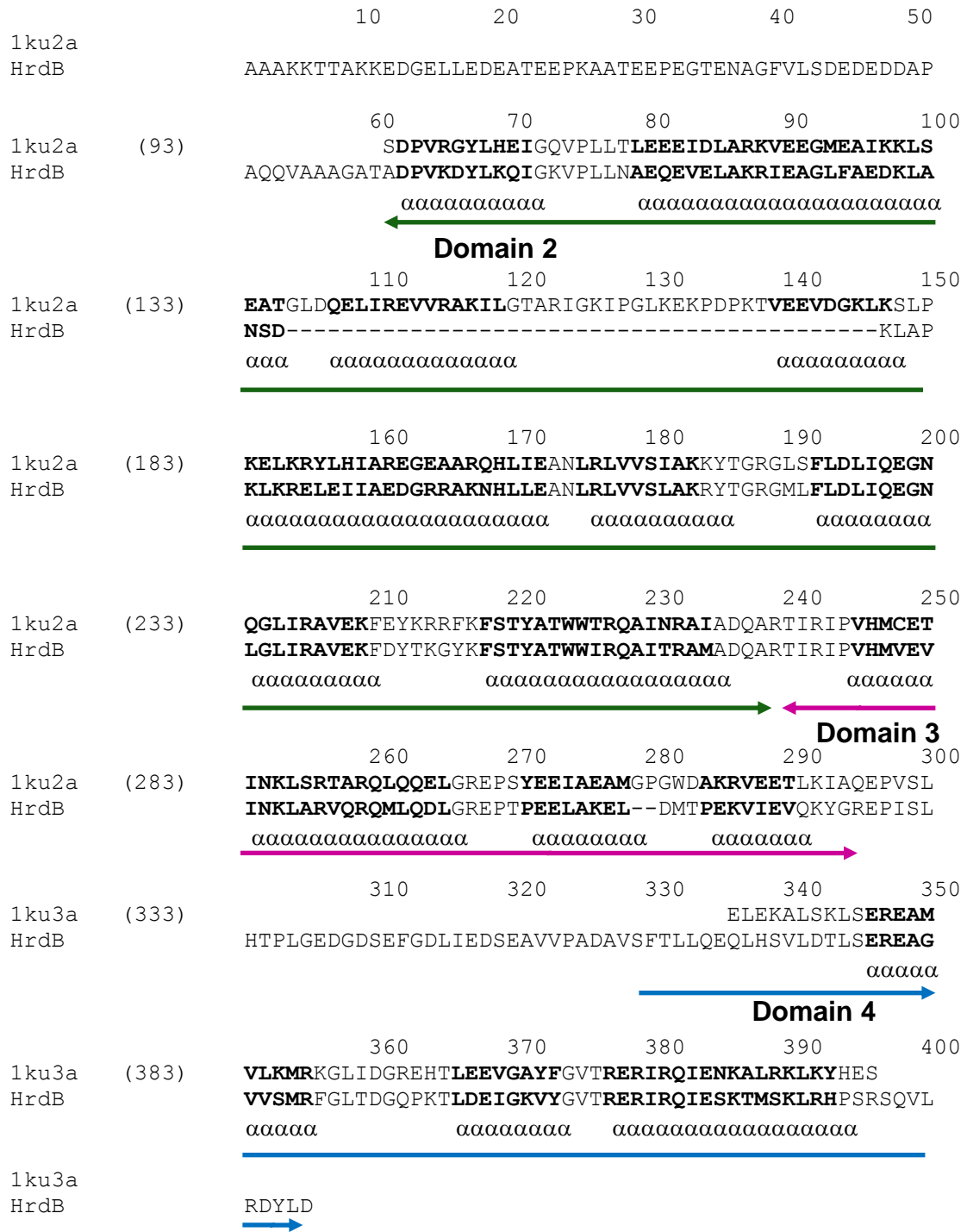
**(A)** Blue/White screening on LA plates grown at 30°C for 16-24 h. **(B)**  $\beta$ -galactosidase assay of RbpA interaction with  $\sigma^B$ ,  $\sigma^R$ ,  $\sigma^E$  and  $\sigma^{WhiG}$ . The activity of  $\beta$ -galactosidase for BTH101 containing vector only, pUT18-*rbpA* and pKT25-*sigB*, pKT25-*sigR*, pKT25-*sigE*, pKT25-*whiG* and pKT25-*hrdB* were measured. These assays were repeated using three separate colonies in triplicate with error bars showing standard deviation.

## **4.2 Mapping the binding site of RbpA on $\sigma^{\text{HrdB}}$**

After showing that RbpA interacts with  $\sigma^{\text{HrdB}}$ , it was important to determine the region to which RbpA binds. Since sigma factors are composed of functionally distinct domains, this might shed some light on the role that RbpA plays in transcription initiation. Furthermore, identifying a small region of the sigma factor that binds to RbpA would aid in the identification of suitable sigma fragments for RbpA-sigma structural studies (Chapter 6).

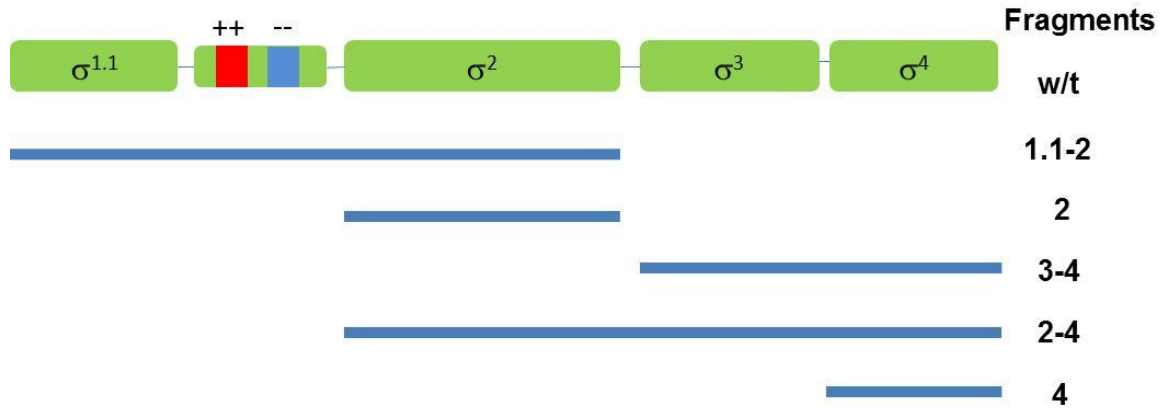
As described in Chapter 1,  $\sigma^{\text{HrdB}}$  is composed of four domains that are interconnected by flexible linkers (Young *et al.*, 2002) (Figure 4.5). These four domains have conserved regions 1.1, 1.2-2.4, 3.0-3.1 and 4.1-4.2. Region 1.1 is actually a relatively non-conserved region that has an auto-inhibitory effect on binding to promoter DNA. Domain 2 contains region 2.3 and 2.4 that recognise and bind to the -10 element of the promoter DNA and is closely linked to region 1.2 which interacts with the “discriminator region” and is important in open complex stability (Haugen *et al.*, 2008). Domain 3 recognises the “extended -10 element”, enhancing the stability of the open complex formation. Domain 4 interacts with -35 element of the promoter (Murakami & Darst, 2003).

The structure of  $\sigma^{\text{A}}$  of *T. aquaticus* was aligned against the sequence of  $\sigma^{\text{HrdB}}$  using FUGUE to identify the distinct structural domains of  $\sigma^{\text{HrdB}}$  (Fig. 4.4). The  $\sigma^{\text{HrdB}}$  fragments were designed and constructed to cover the whole region of the sigma factor. Vassylyev *et al.*, aligned the structure of  $\sigma^{70}$  of *T. thermophilus*, *E. coli* and *B. subtilis* together and then separated the domains based on their structural organisation (Vassylyev *et al.*, 2002). Using this structural organisation the fragments were constructed and are shown below in Figure 4.5 and Table 11.



**Figure 4.4: Alignment of *T. aquaticus*  $\sigma^A$  and  $\sigma^{\text{HrdB}}$ .**

lku2a is the PDB entry of  $\sigma^A$  domain 2-3 and lku3a is the PDB entry of  $\sigma^A$  domain 4. There is a region within 1.2 of  $\sigma^A$  that is not conserved in  $\sigma^{\text{HrdB}}$  and the flexible linker between domains 3-4. This shows the different structural domains of  $\sigma^{\text{HrdB}}$  together with the amino acid compositions.



**Figure 4.5: Constructed structural domains of  $\sigma^{\text{HrdB}}$ .**

The fragments were designed to cover the whole region of the wild-type  $\sigma^{\text{HrdB}}$  (Refer to Table 11).

**Table 11:  $\sigma^{\text{HrdB}}$  fragments for BACTH analysis and *in vitro* pull-down assays.**

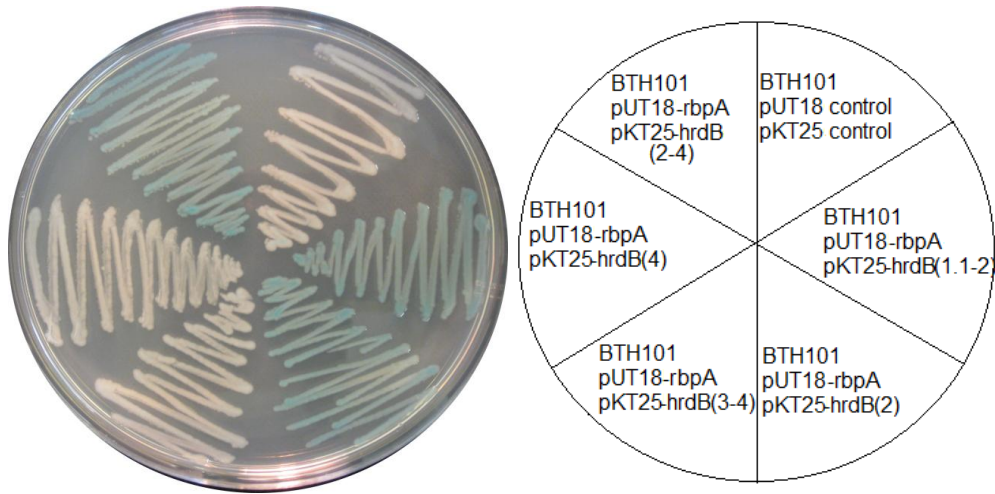
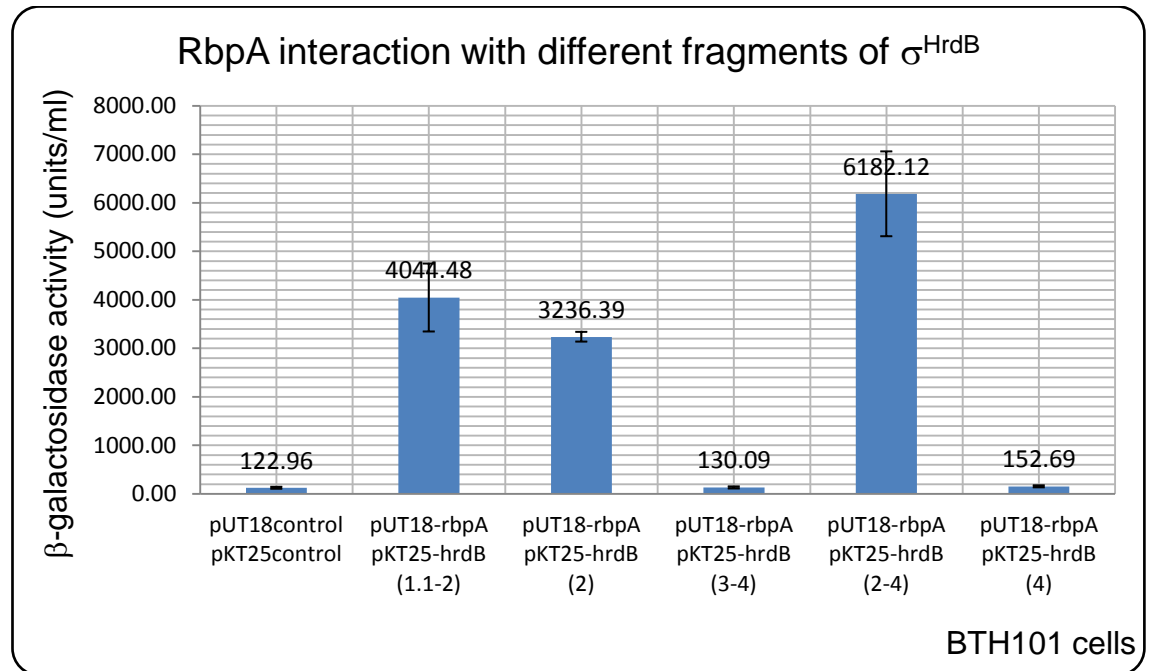
$\sigma^{\text{HrdB}}$ Fragments	Nucleotide length (bp)	Amino Acids	N-terminal sequence	C-terminal sequence
Domain 1.1 and 2 (1.1-2)	~1000	364	fMSASTS	MADQAR
Domain 2 (2)	~430	137	TADPVK	MADQAR
Domain 3 and 4 (3-4)	~512	164	TIRIPV	LRDYLD
Domain 2, 3 and 4 (2-4)	~900	301	TADPVK	LRDYLD
Domain 4 (4)	~250	77	SFTLLQ	LRDYLD

#### 4.2.1 *In vivo* analysis

The five constructed  $\sigma^{\text{HrdB}}$  fragments were used to identify the region to which RbpA binds. pKT25-*hrdB*<sup>(2-4)</sup> was previously obtained for BACTH analysis. The remaining four fragments containing domain 1.1-2, 2, 3-4, and 4, were amplified using *T25\_hrdB(1.1-2)*, *T25\_hrdB(2)*, *T25\_hrdB(3-4)* and *T25\_hrdB(4)* forward

and reverse primers, respectively (Table 5- Section 2.1.5.1). The fragments were flanked by XbaI and EcoRI sites and the fragments were cloned into pKT25. The recombinant pKT25-*hrdB*<sup>(1.1-2)</sup>, pKT25-*hrdB*<sup>(2)</sup>, pKT25-*hrdB*<sup>(3-4)</sup>, pKT25-*hrdB*<sup>(2-4)</sup> and pKT25-*hrdB*<sup>(4)</sup> were electroporated with pUT18-*rbpA* into BTH101 competent cells. Control strains containing vector only were also prepared. These were all plated on LA with additional selection of X-gal and IPTG.

Initial transformation experiments indicated that BTH101 containing pUT18-*rbpA* and pKT25 derivatives carrying domains 3-4 and 4 produced white colonies. However, BTH101 carrying pUT18-*rbpA* and pKT25 derivatives carrying domains 1.1-2, 2 and 2-4 produced blue colonies (Fig. 4.6A).  $\beta$ -galactosidase activities confirmed these data: fragments containing domain 1.1-2, 2 and 2-4 induced 33-, 26- and 50-fold increases in  $\beta$ -galactosidase activity compared to vector only controls (Fig. 4.6B). These results show that RbpA binds to  $\sigma^{\text{HrdB}(1.1-2)}$ ,  $\sigma^{\text{HrdB}(2)}$  and  $\sigma^{\text{HrdB}(2-4)}$  but does not bind to  $\sigma^{\text{HrdB}(3-4)}$  and  $\sigma^{\text{HrdB}(4)}$ .

**A****B**

**Figure 4.6: BACTH assay of RbpA interaction with the constructed fragments of  $\sigma^{\text{HrdB}}$ .**

**(A)** Blue/White Screening on LA plates grown at 30°C for 16-24 h. **(B)**  $\beta$ -galactosidase assay of RbpA interaction with the constructed fragments of  $\sigma^{\text{HrdB}}$ . The  $\beta$ -galactosidase activity was measured for BTH101 vector only and BTH101 pUT18-rbpA and pKT25-hrdB<sup>(1.1-2)</sup>, pKT25-hrdB<sup>(2)</sup>, pKT25-hrdB<sup>(3-4)</sup>, pKT25-hrdB<sup>(2-4)</sup> and pKT25-hrdB<sup>(4)</sup>. Three individual colonies were tested and the assays were done in triplicate for each colony with error bars showing the standard deviation.



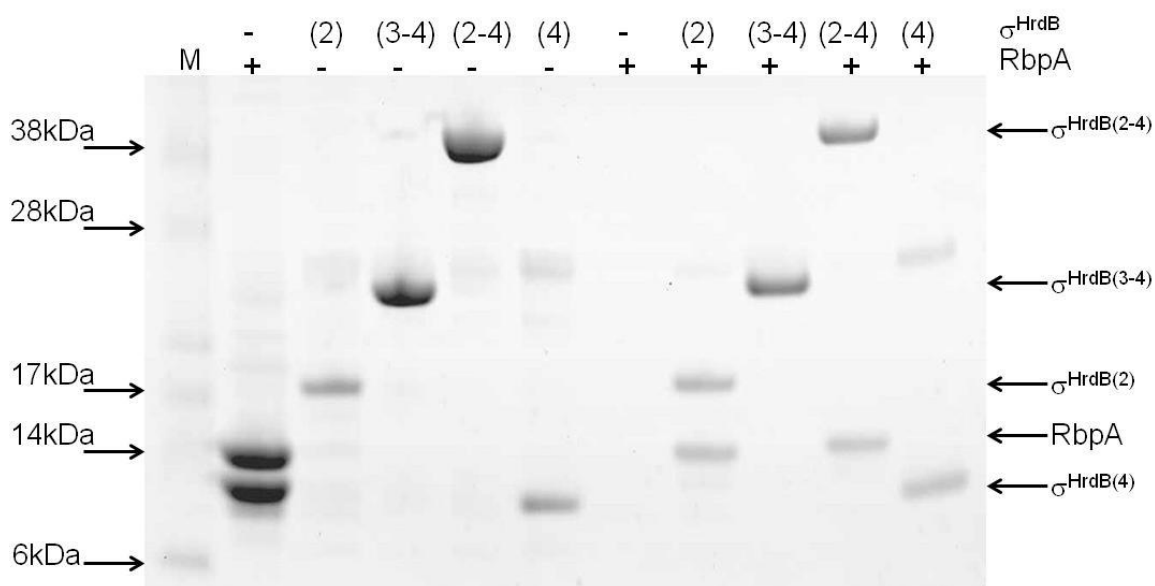
#### 4.2.2 *In vitro* analysis

Using the BACTH system, RbpA was shown to bind to  $\sigma^{\text{HrdB}(1.1-2)}$ ,  $\sigma^{\text{HrdB}(2)}$  and  $\sigma^{\text{HrdB}(2-4)}$  but does not seem to interact with  $\sigma^{\text{HrdB}(3-4)}$  and  $\sigma^{\text{HrdB}(4)}$  (Fig. 4.6). To further prove this, *in vitro* pull-down assays were designed. In this experiment, the fragments would be His-tagged, bound by Ni-affinity magnetic beads, and tested for their ability to sequester RbpA from solution. Each fragment was cloned as an NdeI-BglII fragment from pBlueScript II Sk (+)::*hrdB*<sup>(1.1-2)</sup>, *hrdB*<sup>(2)</sup>, *hrdB*<sup>(3-4)</sup>, *hrdB*<sup>(2-4)</sup>, *hrdB*<sup>(4)</sup> (Table 1- Section 2.1.2) into the expression vector pET15b, which had been digested with NdeI and BamHI. In the resulting expression plasmids, the  $\sigma^{\text{HrdB}}$  fragments would be produced with N-terminal His-tags. Each recombinant pET15b::*hrdB* fragment plasmid was used to transform into *E. coli* BL21 (pLysS) and tested for expression (Section 2.3.1.1). Unfortunately,  $\sigma^{\text{HrdB}(1.1-2)}$  was insoluble and therefore, work on it was discontinued. The remaining four fragments were soluble and purified using Ni-NTA spin columns (Qiagen) (Section 2.3.2.3) (Fig. 4.7, lanes 3-6).

In order to produce non-tagged native RbpA, *rbpA* was subcloned as an NdeI-BamHI fragment from pBlueScript::*rbpA* (Table 1- Section 2.1.2) into the expression vector pET20b. pET20b::*rbpA* was used to transform into BL21 (pLysS) and RbpA was purified by Q.F.F anion-exchange chromatography, Mono Q anion-exchange chromatography, followed by gel filtration (Fig. 4.7, lane 2) (Section 2.3.2.5 & 2.3.2.6). Note that RbpA often runs on SDS polyacrylamide gels as two bands, which is thought to be due to partial degradation (see below).

To perform the *in vitro* pull-down assays, the purified His-tagged  $\sigma^{\text{HrdB}}$  fragments (containing domains 2, 3-4, 2-4 and 4) were mixed with purified native RbpA on ice for 15 min and then mixed with magnetic Ni-affinity Dynabeads (Section 2.3.2.4). As a control, RbpA alone was mixed with the magnetic Dynabeads. The Dynabeads were washed four times with buffer, then any bound protein eluted using 100  $\mu$ l elution buffer. As shown in Figure 4.7 lane 7, RbpA by itself did not bind to the Ni-affinity Dynabeads. In addition, there was no binding when mixed with  $\sigma^{\text{HrdB}(3-4)}$  and  $\sigma^{\text{HrdB}(4)}$  (Fig. 4.7; lane 9 & 11

respectively). However, when mixed with  $\sigma^{\text{HrdB}(2)}$  and  $\sigma^{\text{HrdB}(2-4)}$  RbpA co-eluted from the beads, confirming that RbpA binds specifically to  $\sigma^{\text{HrdB}}$  fragments containing conserved region 1.2-2.4 (Fig. 4.7; lane 8 & 10 respectively). As mentioned above, RbpA was separated by SDS polyacrylamide gel as two bands, which might be possibly due to degradation. Indeed, in support of this the lower band tended to increase in relative intensity in later steps of the purification. Significantly, it is worth noting that only the top band interacted with  $\sigma^{\text{HrdB}}$  (Chapter 5 and 6).



**Figure 4.7: The interaction of RbpA with  $\sigma^{\text{HrdB}(2)}$  and  $\sigma^{\text{HrdB}(2-4)}$ .**

The pull down-assays were assessed by separating the proteins on SDS polyacrylamide gel. The Marker used was SeeBlue<sup>®</sup>Plus2 (Lane 1). Lanes 2 to 6 are showing RbpA (~14kDa),  $\sigma^{\text{HrdB}(2)}$  (~17kDa),  $\sigma^{\text{HrdB}(3-4)}$  (~22.5kDa),  $\sigma^{\text{HrdB}(2-4)}$  (~38) and  $\sigma^{\text{HrdB}(4)}$  (~12kDa) after purification and their respective concentrations. Lanes 7 to 11 are the eluates from the dynabeads column. Lane 7: RbpA only, Lane 8: RbpA and  $\sigma^{\text{HrdB}(2)}$ , Lane 9: RbpA and  $\sigma^{\text{HrdB}(3-4)}$ , Lane 10: RbpA and  $\sigma^{\text{HrdB}(2-4)}$  and Lane 11: RbpA and  $\sigma^{\text{HrdB}(4)}$ . Proteins were separated on a 4-12% Bis-Tris polyacrylamide gel at 120 V for 80 min, and visualised by Coomassie staining.

### **4.3 Rv2050 interaction with *M. tuberculosis* sigma factors**

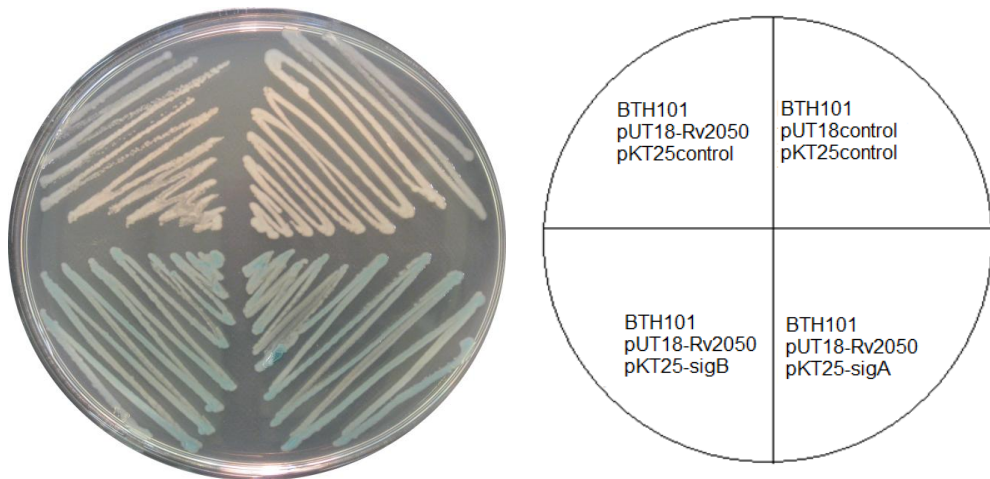
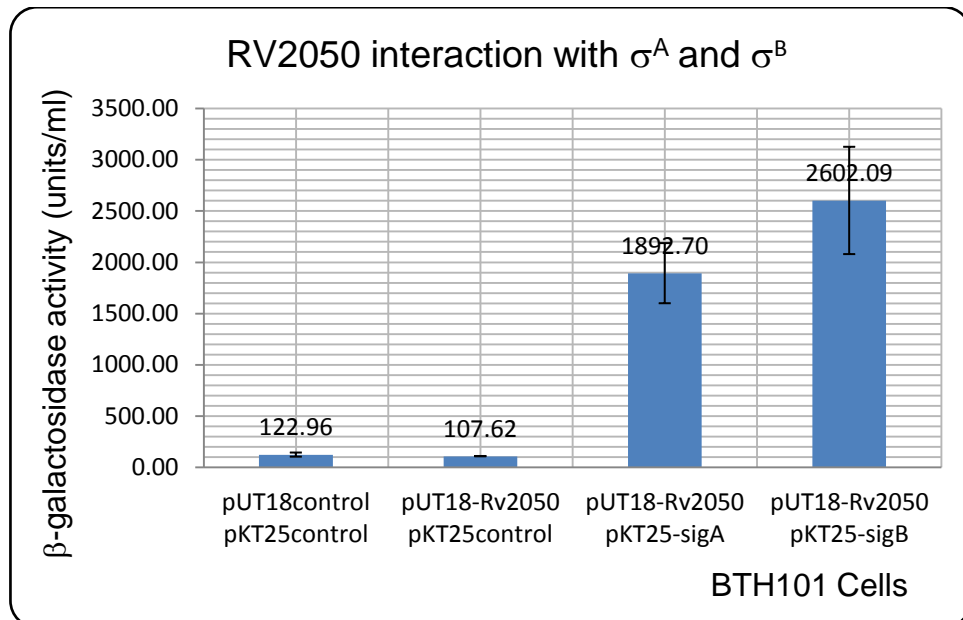
RbpA homologues are found in virtually all Actinobacteria including the major human pathogen, *M. tuberculosis*. This is of particular interest because RNAP is the target of the frontline tuberculosis antibiotic, rifampicin. Therefore, the next step was to identify whether the RbpA homologue in *M. tuberculosis*, Rv2050, behaves in a similar manner to RbpA in *S. coelicolor*. Recently, it was found that Rv2050 is an essential gene in *M. tuberculosis* (Forti *et al.*, 2011) raising the possibility that Rv2050 might be a useful drug target.

$\sigma^{\text{HrdB}}$ , the principal sigma factor in *S. coelicolor*, has two homologues in *M. tuberculosis*; the first homologue is Rv2703 ( $\sigma^{\text{A}}$ ).  $\sigma^{\text{A}}$  is the principal sigma factor in *M. tuberculosis* and is involved in expression of house-keeping genes and also modulates expression of some virulence genes (Rodrigue *et al.*, 2006). It is 81% identical to  $\sigma^{\text{HrdB}}$  with 528 amino acids and a molecular mass of 57768.1 Da. The other homologue of  $\sigma^{\text{HrdB}}$  is Rv2710 ( $\sigma^{\text{B}}$ ) which is a Group II/principal-like sigma factor. It controls genes involved in stress response, example heat-shock response, and stationary phase survival, and is itself induced by stress (Rodrigue *et al.*, 2006). It is 61% identical to  $\sigma^{\text{HrdB}}$  with 323 amino acids and a molecular mass of 36343.2 Da.

#### **4.3.1 $\sigma^{\text{A}}$ and $\sigma^{\text{B}}$**

Rv2050 was amplified with T18\_Rv2050 forward and reverse primers (Table 5- Section 2.1.5.1) introducing BamHI and EcoRI sites and keeping the gene in frame when fused to pUT18. *sigA* and *sigB* were amplified using T25\_sigAMt and T25\_sigBMt forward and reverse primers (Table 5- Section 2.1.5.1), respectively, introducing an XbaI and EcoRI site and keeping the genes in frame when fused to pKT25. The inserts were cloned into pBlueScript II SK+, sequenced and sub-cloned into pUT18 or pKT25. Recombinant pUT18\_Rv2050 together with pKT25\_sigAMt or pKT25\_sigBMt were electroporated into BTH101 competent cells. Control strains were also prepared. These were streaked on LA plates with additional selection with X-gal and IPTG.

Negative control BTH101 containing vector only and pUT18-*Rv2050* only produced white colonies. However, BTH101 containing pUT18-*Rv2050* and pKT25-*sigAMt* and pKT25-*sigBMt* produced blue colonies (Fig. 4.8A). To confirm this result, the  $\beta$ -galactosidase assays were performed (Section 2.6).  $\beta$ -galactosidase activity of BTH101 pUT18-*Rv2050* and pKT25-*sigAMt*, pKT25-*sigBMt* had 18- and 24- fold higher  $\beta$ -galactosidase activity compared to the negative controls, respectively (Fig. 4.8B). These results suggest that Rv2050 interacts with  $\sigma^A$  and  $\sigma^B$  in *M. tuberculosis*.

**A****B**

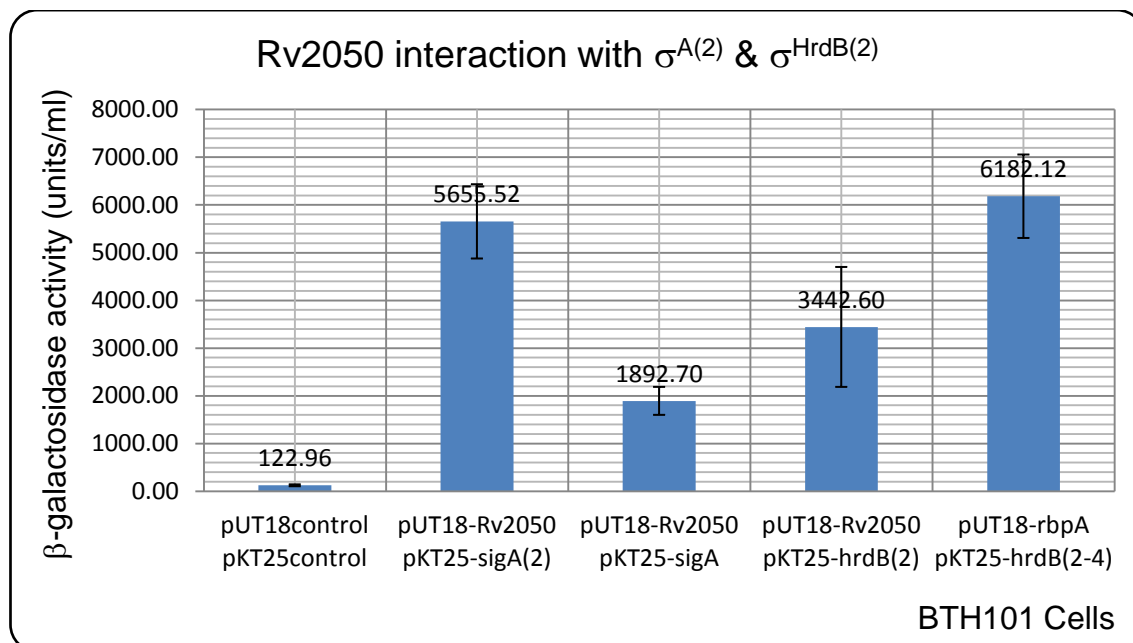
**Figure 4.8: BACTH assay of Rv2050 interaction with principal sigma factor,  $\sigma^A$ , and principal-like sigma factor,  $\sigma^B$ , of *M. tuberculosis*.**

**(A)** Blue/White Screening on LA plates grown at 30°C for 16-24 h. **(B)**  $\beta$ -galactosidase assay of Rv2050 interaction with  $\sigma^A$  and  $\sigma^B$ . The  $\beta$ -galactosidase activity was measured for BTH101 containing vector only, pUT18-Rv2050 only, pUT18-Rv2050 and pKT25-sigAMt and pKT25-sigBMt. Three individual colonies were tested and the assays were done in triplicate for each colony with error bars showing the standard deviation.

### 4.3.2 Rv2050 interaction with sigma domain 2

#### 4.3.2.1 *In vivo* analysis

RbpA was shown to bind to  $\sigma^{\text{HrdB}(2)}$  (Fig. 4.6) and Rv2050 was shown to interact with  $\sigma^{\text{A}}$  (Fig. 4.8). Therefore, it was decided to test if Rv2050 also interacts with domain 2 *i.e* region 1.2-2.4 of its principal sigma factor,  $\sigma^{\text{A}}$ , and  $\sigma^{\text{HrdB}(2)}$  using BACTH assay. The equivalent region of  $\sigma^{\text{HrdB}(2)}$  was determined in  $\sigma^{\text{A}}$  using Clustal W alignment tool (data not shown). *sigA*<sup>(2)</sup> fragment was amplified using T25\_*sigA*(2) forward and reverse primers (Table 5- Section 2.1.5.1), which was flanked by XbaI and EcoRI sites and cloned into pKT25. Recombinant pKT25-*sigA*<sup>(2)</sup> and pKT25-*hrdB*<sup>(2)</sup>(see above) were co-transformed together with pUT18-Rv2050 into competent BTH101 cells by electroporation. Positive and negative control strains were also prepared (see above). To quantify and confirm these interactions, the  $\beta$ -galactosidase assay was performed (Section 2.6). The assay showed that domain 2 of  $\sigma^{\text{A}}$  and  $\sigma^{\text{HrdB}}$  induced 46- and 28-fold increase in  $\beta$ -galactosidase activity compared to the vector only control, respectively. This result suggests that Rv2050 interacts with both  $\sigma^{\text{A}(2)}$  and  $\sigma^{\text{HrdB}(2)}$  (Fig. 4.9).



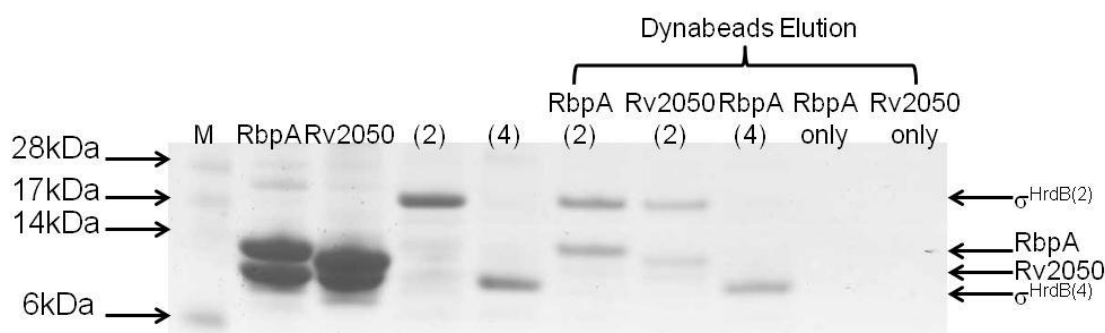
**Figure 4.9: The  $\beta$ -galactosidase activity of the interaction of Rv2050 with  $\sigma^{A(2)}$  and  $\sigma^{HrdB(2)}$ .**

The  $\beta$ -galactosidase assay was performed on BTH101 containing vector only, pUT18-Rv2050 and pKT25-sigA<sup>(2)</sup>, pKT25-sigA, pKT25-hrdB<sup>(2)</sup> together with the positive control pUT18-rbpA and pKT25-hrdB<sup>(2-4)</sup>. Three individual colonies were tested and the assays were repeated in triplicate for each colony with error bars showing the standard deviation.

#### 4.3.2.2 In vitro analysis

To further verify the results obtained from BACTH that Rv2050 interacts with  $\sigma^{HrdB(2)}$ , the *in vitro* pull down assay was used.  $\sigma^{HrdB(2)}$  fragment was previously His-tagged and purified (Section 4.2.2). In order to produce non His-tagged native Rv2050, Rv2050 was amplified using Rv2050 reverse and forward primers (Table 7- Section 2.1.5.3), cloned into pBlueScript II SK+ and subcloned as an NdeI-BamHI fragment into the expression vector pET20b. pET20b::Rv2050 was used to transform into BL21 (pLysS) and Rv2050 was purified by QFF anion-exchange chromatography, Mono Q anion-exchange chromatography, followed by gel filtration (Section 2.3.2.5 & 2.3.2.6) (Fig. 4.10, lane 3).

Rv2050 and His-tagged  $\sigma^{\text{HrdB}(2)}$  were mixed together with magnetic Ni-affinity Dynabeads (Section 2.3.2.4). The beads were washed four times with wash buffer and eluted in 100  $\mu\text{l}$  elution buffer (see above). As shown in Fig. 4.10, RbpA and Rv2050 by themselves did not bind to the Ni-affinity dynabeads (Fig. 4.10, lane 9 & 10). Also, RbpA did not bind to  $\sigma^{\text{HrdB}(4)}$  (Fig. 4.10, lane 8). However, when Rv2050 and RbpA were mixed with  $\sigma^{\text{HrdB}(2)}$  both co-eluted from the beads (Fig. 4.10, lane 6 & 7). These results confirm that Rv2050 binds to domain 2 of  $\sigma^{\text{HrdB}}$ .



**Figure 4.10: The interaction of Rv2050 with  $\sigma^{\text{HrdB}(2)}$ .**

The pull down-assays were assessed by separating the proteins on SDS polyacrylamide gel. The Marker used was SeeBlue® Plus2 (Lane 1). Lanes 2 to 5 are showing RbpA (~14kDa), Rv2050 (~12.5kDa),  $\sigma^{\text{HrdB}(2)}$  (17kDa) and  $\sigma^{\text{HrdB}(4)}$  (~12kDa), respectively, after purification. Lanes 6 to 10 are the eluates from the dynabeads column. Lane 6: RbpA and  $\sigma^{\text{HrdB}(2)}$ , Lane 7: Rv2050 and  $\sigma^{\text{HrdB}(2)}$ , Lane 8: RbpA and  $\sigma^{\text{HrdB}(4)}$ , Lane 9: RbpA only and Lane 10: Rv2050 only. The gel used was 4-12% Bis-Tris polyacrylamide gel which was run at 120 V for 80 min. The proteins were visualised using Coomassie Blue staining.

## **4.4 Discussion**

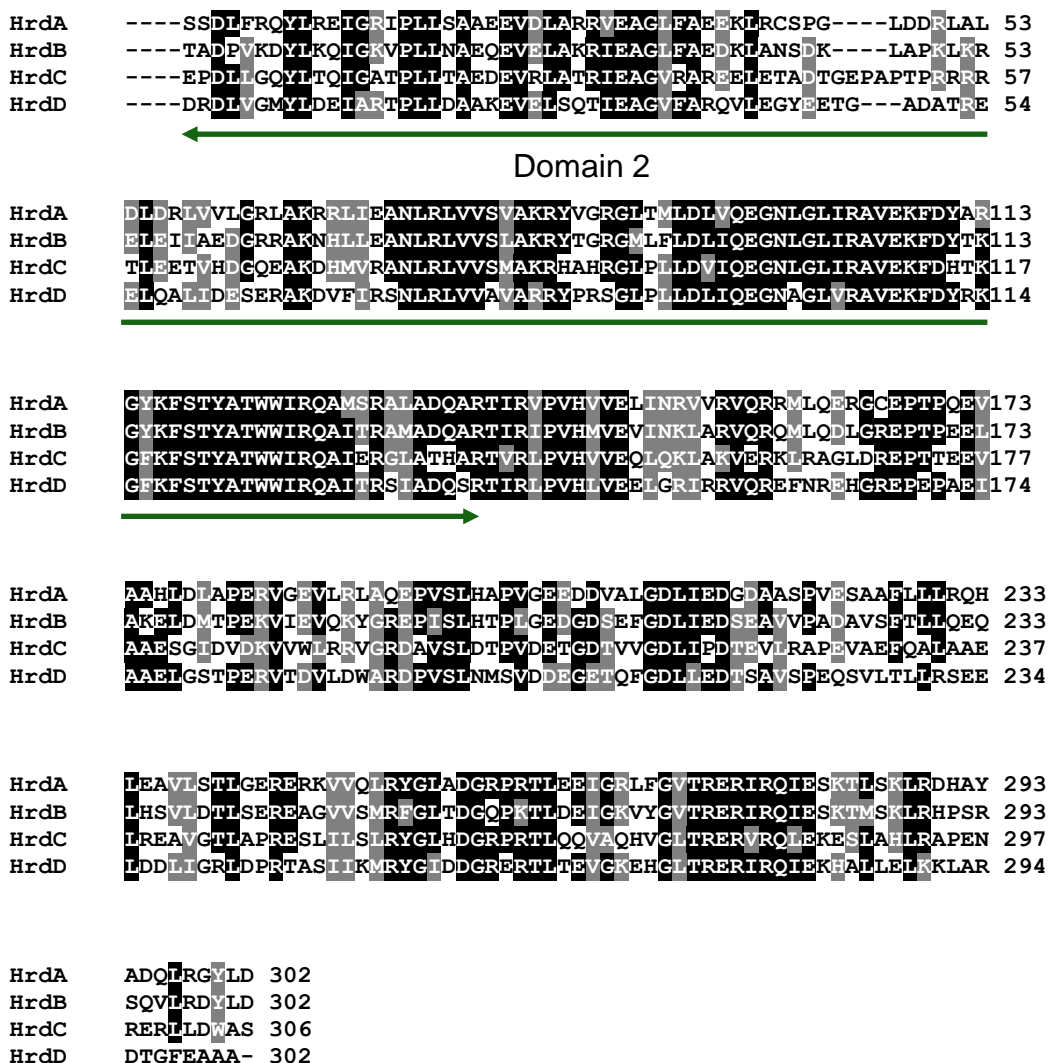
### **4.4.1 RbpA binds to $\sigma^{\text{HrdA}}$ and $\sigma^{\text{HrdB}}$**

Using BACTH, RbpA was shown to interact with the principal sigma factor,  $\sigma^{\text{HrdB}}$  and to some extent with the principal-like sigma factor,  $\sigma^{\text{HrdA}}$ . There was no interaction of RbpA with the other two principal-like sigma factors,  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$ . The reason for RbpA interaction with  $\sigma^{\text{HrdA}}$  and  $\sigma^{\text{HrdB}}$  but not with  $\sigma^{\text{HrdC}}$



and  $\sigma^{\text{HrdD}}$  is currently unknown. However, it is worth noting that the level of amino acid conservation between  $\sigma^{\text{HrdB}}$  and  $\sigma^{\text{HrdA}}$  along conserved region 1.2 to 2.4 is slightly higher than the equivalent region of  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$  (Fig. 4.11). This suggests that key amino acids that are involved in sigma-RbpA interaction in  $\sigma^{\text{HrdB}}$  are absent in  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$ .

However, the result obtained from the BACTH assay of the interaction of RbpA with  $\sigma^{\text{HrdD}}$  was contradictory to previous work performed using surface plasmon resonance (BIAcore 2000), where  $\sigma^{\text{HrdD}}$  was shown to bind to RbpA (P. Doughty, and M. Paget, personal communication). However, the  $\sigma^{\text{HrdD}}$ -RbpA interaction appeared to be at least 4-fold weaker than the  $\sigma^{\text{HrdB}}$ -RbpA interaction. This weaker interaction might not be detected by the BACTH assays. Also, another disadvantage of the BACTH system is that the proteins fused to the pKT25 and pUT18 vectors might not be expressed or might be unstable when produced, and therefore degraded. As a result, further work needs to be done to determine whether the proteins,  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$  were indeed expressed and not degraded *in vivo*. To verify these interactions, *in vitro* pull-down assays can also be performed.



**Figure 4.11: The multiple sequence alignment of domains 2-4 of  $\sigma^{\text{HrdA}}$ ,  $\sigma^{\text{HrdB}}$ ,  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$ .**

The Clustal W2 multiple sequence alignment tool and the BOXSHADE 3.21 server were used. Amino acids in white font with black highlight show identical residues whilst amino acids in white font with grey highlight show similar residues. The homology was tested to the 0.7 fraction that must agree for shading. The most conserved domain 2 is indicated with arrows.

#### 4.4.2 The principal sigma factor, $\sigma^{\text{HrdB}}$

RbpA has been shown to interact with  $\sigma^{\text{HrdB}}$  and  $\sigma^{\text{HrdA}}$  using BACTH assays (Fig. 4.2). However, which of these sigma factors does RbpA bind to *in vivo*? A BLAST search was done using the amino acid sequence of  $\sigma^{\text{HrdB}}$  against the Actinobacteria in the Integrated Microbial Genome (IMG) database. It was shown that some of these organisms only contained a principal sigma factor and no principal-like sigma factor, *e.g.* *Actinomyces coleocanis*, *Actinomyces oris*, and *Bifidobacterium bifidum*. Bacteria contain at least one sigma factor that plays a major role in expression of housekeeping genes and is vital for the viability of the organism. RbpA is nevertheless, present in these organisms that lack principal-like sigma factors, which further suggests that  $\sigma^{\text{HrdB}}$  is the main target of RbpA *in vivo* (Table A1- Appendices).

As RbpA interacts with  $\sigma^{\text{HrdB}}$ , it was decided to determine the level of conservation of  $\sigma^{\text{HrdB}}$  in the Actinobacteria. Alignment of  $\sigma^{\text{HrdB}}$  homologues from the different genera obtained from the BLAST search above, showed that  $\sigma^{\text{HrdB}}$  was well conserved amongst the Actinobacteria (Fig.A2- Appendices; Sig\_Mt to Sig\_Fa). Therefore, both RbpA and  $\sigma^{\text{HrdB}}$  were highly conserved amongst the Actinobacteria, although RbpA was absent in some species (Table A1- Appendices).

It was decided to determine whether there is any clear difference in  $\sigma^{\text{HrdB}}$  sequences in organisms containing RbpA with those that do not in the Actinobacteria. Hence, the amino acid sequences of  $\sigma^{\text{HrdB}}$  homologues from the Actinobacteria lacking RbpA, together with  $\sigma^{\text{HrdB}}$  homologues from Actinobacteria containing RbpA, were aligned (Fig.A2- Appendices). From this multiple sequence alignment, there is no apparent difference in the level of amino acid conservation of  $\sigma^{\text{HrdB}}$  in either the Actinobacteria containing RbpA (Fig.A2- Appendices, bottom 8 sequences) or those lacking RbpA (Fig. A2- Appendices, top 7 sequences). Therefore, the organisms lacking RbpA may have evolved a different way to overcome the need for RbpA whilst keeping the principal sigma factor sequences relatively similar (Section 3.3.1). As a result, it is unclear why RbpA is utilised in some Actinobacteria but not in others.

#### 4.4.3 RbpB interaction with sigma factors

RbpA was shown to have a paralogue, RbpB, in *S. coelicolor* (Section 3.1.2), which was shown to partially compensate for the loss of RbpA in *S. coelicolor* strains when overexpressed (K. Newell and M. Paget, personal communication). Therefore, it was considered that the role of RbpB might be to activate one or more principal-like sigma factors, or possibly other sigma factors in *S. coelicolor*  $\Delta rbpA$  strains, therefore ensuring viability. However, recent BACTH work has shown that RbpB interacted with  $\sigma^{\text{HrdA}}$  and  $\sigma^{\text{HrdB}}$  but not with  $\sigma^{\text{HrdC}}$ ,  $\sigma^{\text{HrdD}}$ ,  $\sigma^{\text{B}}$ ,  $\sigma^{\text{E}}$ ,  $\sigma^{\text{R}}$  and  $\sigma^{\text{WhiG}}$  (In collaboration with R. Lewis and M. Paget, unpublished observations). This might explain why *S. coelicolor*  $\Delta rbpA$  strains are viable; RbpB might partially take over the role of RbpA through binding to  $\sigma^{\text{HrdB}}$ , leading to the activation of  $\sigma^{\text{HrdB}}$ -dependent promoters, hence, maintaining the survival of the organism. In addition, since both RbpA and RbpB interact with  $\sigma^{\text{HrdB}}$ , they might recognise different promoter sequences and possibly modulate the activity of  $\sigma^{\text{HrdB}}$ .

#### 4.4.4 RbpA interacts at domain 2 of $\sigma^{\text{HrdB}}$

RbpA was shown to interact with all the fragments of  $\sigma^{\text{HrdB}}$  that contained domain 2: 1.1-2, 2 and 2-4 (Fig. 4.6). Even though BACTH is not a quantitative test, the  $\beta$ -galactosidase activity was 2-fold higher for the interaction of RbpA with  $\sigma^{\text{HrdB}}$  containing domain 2-4 compared to the interaction of RbpA with  $\sigma^{\text{HrdB}}$  containing domain 2. This suggests that domain 2-4 may provide a better binding surface for RbpA. Another possibility for the decrease in  $\beta$ -galactosidase activity in  $\sigma^{\text{HrdB}(2)}$  compared to  $\sigma^{\text{HrdB}(2-4)}$  might be due to the difference in the folding of the fragments and their conformation, which will affect RbpA binding.

Sigma domain 2 is composed of sub-regions 1.2, 2.1, 2.2, 2.3 and 2.4 which are highly conserved in the  $\sigma^{70}$  family (Murakami & Darst, 2003). Each of these regions play an important role in the core RNAP binding (region 1.2-2.2), promoter melting (region 2.3), the recognition and binding of the -10 element (region 2.4), which leads to open complex formation (Section 1.2.4.1). *E. coli*  $\sigma^{70(2)}$  was shown to make multiple contacts with the core RNAP via the  $\alpha$ -helical

coiled-coil  $\beta'$  domain (Vassylyev *et al.*, 2002). The RNAP  $\beta'$  coiled-coil subunit interacts with  $\sigma^{70}$  region 2.2, inducing the A<sub>-11</sub> base in the non-template strand of the -10 element to flip out, followed by a 90° kink in the non-template DNA strand, which causes the T<sub>-7</sub> base of the non-template strand of the -10 element to also flip. This leads to complete promoter melting, the -10 element sequence recognition, and thereby, the formation of the open complex (Young *et al.*, 2001, Feklistov & Darst, 2011).

As RbpA interacts with region 1.2-2.4 of  $\sigma^{\text{HrdB}}$ , it may aid in the melting of the DNA, allowing region 2.4 to bind to the -10 element thus, leading to the open complex formation. It may also increase the stability of the open complex. A role in open complex formation or stability is supported by the KMnO<sub>4</sub> assay, which showed that in the presence of RbpA, the level of promoter DNA melting increased (P. Doughty and M. Paget, personal communication). The exact role of RbpA in transcription initiation, however, has not yet been defined.

#### 4.4.5 $\sigma^{\text{HrdB}}$ lacks a non-conserved region between region 1.2 to 2.1

The ability of RbpA to interact with the *T. thermophilus* principal sigma factor,  $\sigma^{\text{A}(2)}$ , was tested using BACTH assays. Interestingly, RbpA was shown not to interact with  $\sigma^{\text{A}(2)}$  (In collaboration with R. Lewis and M. Paget, unpublished observation). The main difference between  $\sigma^{\text{A}(2)}$  and  $\sigma^{\text{HrdB}(2)}$  is that  $\sigma^{\text{HrdB}(2)}$  lacks a non-conserved amino acid region that is present between region 1.2 to 2.1 in  $\sigma^{\text{A}(2)}$  (Fig. 4.12). This NCR between region 1.2 to 2.1 is also present in *E. coli*, *Salmonella typhimurium*, *Pseudomonas aeruginosa* and other principal sigma factors (Malhotra *et al.*, 1996). However, this region is absent from the principal sigma factors of *S. coelicolor* (Fig. 4.12- dark grey highlight), *B. subtilis*, *Staphylococcus aureus* and other principal sigma factors (Malhotra *et al.*, 1996).

$\sigma^{70}$  NCR was shown to have no effect on transcription *in vitro* (Kumar *et al.*, 1995) on some promoters. However, the deletion of  $\sigma^{70}$  NCR was shown to be important for cell viability (Leibman & Hochschild, 2007).  $\sigma^{70}$  NCR was shown to interact with  $\beta'$  subunit of RNAP close to its N-terminus and this interaction was shown to inhibit  $\sigma^{70}$ -dependent pausing (Leibman & Hochschild, 2007). The  $\sigma^{70}$ -

dependent pausing occurs when  $\sigma^{70(2)}$  interacts with  $\beta'$  coiled-coil subunit of RNAP, recognises and binds to a -10-like element on the non-template strand during elongation (Ring *et al.*, 1996).  $\sigma^{70}$  NCR- $\beta'$  interaction was also shown to assist in promoter escape in contrast to the role of  $\sigma^{70(2)}$ - $\beta'$  coiled-coil interaction, which appears to obstruct promoter escape (Leibman & Hochschild, 2007). Therefore, one can speculate that a possible function of RbpA is to bind to  $\sigma^{\text{HrdB}(2)}$  and possibly  $\beta'$  subunit of RNAP and thereby aid in promoter escape.

SigAtt	MKSKRKNAQAQEAQETEVLVQEEAEELPEFPGEPEPDPLEDDPDLEDDLLDLPEEGEG	60
HrdB(2)	-----	
SigAtt	LDLEEEEDLPIPKISTSDPVRQYLHEIGQVPLLTL EEVELARKVEEGMEAIKKLSEIT	120
HrdB(2)	-----TADPVKDYLKQIGKVPLLNAEQEVELAKRIEAGLFAEDKL-----	40
	*:***:***:***:***. *:*****::* *: * .**	
SigAtt	GLDPDLIREVVRKILGSARVRHIPGLKETLDPKTVEEIDQKLKSLPKEHKRYLHIAREG	180
HrdB(2)	-----G	41
	*	
SigAtt	EAARQHLEIANLRLVVSIAKKYTGRGLSFLDLIQEGNQGLIRAVEKF EYKRRFKFSTYAT	240
HrdB(2)	RRAKNHLLEANLRLVVSIAKRYTGRGMLFLDLIQEGNLGLIRAVEKF DYTGYKFSTYAT	101
	. *:***:*****:***:***: ***** *****:*. : *****	
SigAtt	WWIRQAINRAIADQARTIRIPVHVMETINKLSRTARQLQQLGREPTYEEIAEAMGPGWD	300
HrdB(2)	WWIRQAITRAMADQAR-----	117
	*****.***:*****	
SigAtt	AKRVEETLKIAQEPVSLETPIGDEKDSFYGDFIPDEHLPSPVDAATQSLSEELEKALSK	360
HrdB(2)	-----	
SigAtt	LSEREAMVLKLRKGLIDGREHTLEEVGAF FGVTRERIRQIENKALRKLYHESRTRKLRD	420
HrdB(2)	-----	
SigAtt	FLD	423
HrdB(2)	---	

**Figure 4.12: The pairwise sequence alignment of  $\sigma^{\text{HrdB}(2)}$  and *T. thermophilus* principal sigma factor,  $\sigma^A$ .**

The light grey shaded region indicates the start and end amino acid sequences of region 1.2 to 2.4. The dark grey shaded region indicates the amino acids absent from  $\sigma^{\text{HrdB}(2)}$ . This alignment was performed using Clustal W 2.1. The asterisk (\*)= completely conserved residues, double dots (:)= highly conserved residues, and single dot (.)= slightly conserved residues.

#### 4.4.6 Crl, a possible analogue of RbpA

RbpA is the first described positive regulator of the principal sigma factor that interacts with region 1.2 to 2.4. Most proteins that interact with domain 2 of the sigma factor are the anti-sigma factors, for example ChrR binds to domain 2 and 4 of  $\sigma^E$  in *Rhodobacter sphaeroides* preventing  $\sigma^E$  from forming stable complex with core RNAP (Campbell *et al.*, 2007). Another protein that has recently been shown to interact with sigma and might function in an analogous way to RbpA is Crl (Section 1.2.9.1). Crl, in *Salmonella enteric*, was shown to bind to domain 2 of  $\sigma^S$ , a stationary-phase sigma factor (Monteil *et al.*, 2010b), and leads to the activation of  $\sigma^S$ -dependent genes (England *et al.*, 2008). It is thought that Crl might be involved in aiding holoenzyme formation (Gaal *et al.*, 2006), or formation of the open complex (Monteil *et al.*, 2010b). There is some evidence that Crl also binds to  $\sigma^{70}$  and activates  $\sigma^{70}$ -dependent transcription (Gaal *et al.*, 2006).

Crl is not present in the Actinobacteria. A BLAST search showed that there are 143 genomes that contain Crl, all in *Gammaproteobacteria* that contain an *rpoS* gene (Monteil *et al.*, 2010a). Using PSIPRED secondary structure prediction, Crl was shown to have an N-terminal helix, followed by four  $\beta$ -strands and a C-terminal amphipathic helix (Fig. 4.13). RbpA was shown in Section 3.1.1 to have four  $\beta$ -strands and a C-terminal amphiphatic helix. Due to the similarity in structure and some conserved residues in the amphiphatic helix (H2) (Fig. A3-Appendix), it is possible that the interaction of these proteins with sigma region 1.2 to 2.4 might involve some residues from the H2 (Chapter 5). Hence, Crl and RbpA might have analogous function.

[illegible]

**Figure 4.13: Secondary structure prediction of Crl.**

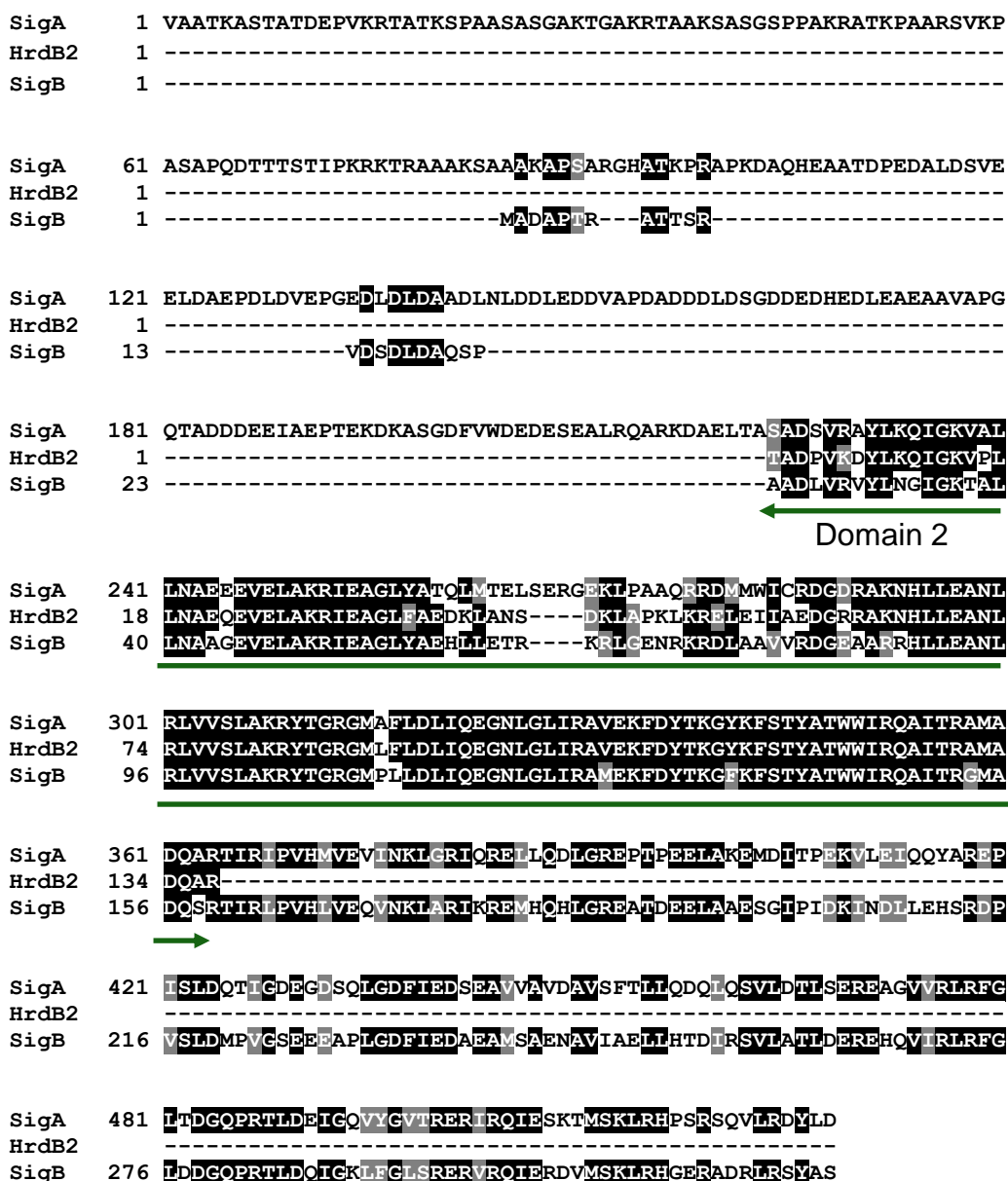
The grey shaded region show the amphiphatic helix of Crl. Using the PSIPRED HFORMAT (V2.6) by David Jones, Conf: confidence level, Pred: secondary prediction and AA: amino acids. C represents random coils, E:  $\beta$ -strands, and H: helices.

#### 4.4.7 The importance of Rv2050 in *M. tuberculosis*

Rv2050, the homologue of RbpA in *M. tuberculosis*, has been shown to be an essential gene in *M. tuberculosis*. Recent work involving knocking out Rv2050 by transposon mutagenesis have shown that *M. tuberculosis* can no longer grow (Forti *et al.*, 2011). Consistent with this, Rv2050 was shown to interact with its principal essential sigma factor,  $\sigma^A$ , and its principal-like sigma factor,  $\sigma^B$  (Fig. 4.8).  $\sigma^A$  is involved in controlling housekeeping genes and also some virulence genes, whilst  $\sigma^B$  is involved in general stress response and in the regulation of stationary-phase genes (Section 1.2.3.2). Rv2050 was also shown to interact with  $\sigma^{\text{HrdB}(2)}$  (Fig. 4.10) and this data is consistent with earlier studies where *Rv2050* introduced into *S. coelicolor*  $\Delta rbpA$  strain would partially complement the  $\Delta rbpA$  strain (Newell *et al.*, 2006). A multiple sequence alignment of  $\sigma^A$ ,  $\sigma^B$  and  $\sigma^{\text{HrdB}(2)}$  showed a high level of amino acid conservation between region 1.2 to 2.4 of these sigma factors (Fig. 4.14). This explains the strong interaction observed for Rv2050 interaction with both sigma factors,  $\sigma^A$  and  $\sigma^B$ , in *M. tuberculosis* (Fig. 4.8).



Rv2050 was shown to have no paralogues in *M. tuberculosis* unlike the case in *S. coelicolor* (Table A1- Appendices). In *S. coelicolor*, RbpB was shown to interact with  $\sigma^{\text{HrdB}}$  and  $\sigma^{\text{HrdA}}$ , and may take over some of the roles of RbpA in a *S. coelicolor*  $\Delta rbpA$  strain (Section 4.4.3). This might explain why Rv2050 is essential in *M. tuberculosis*, while RbpA is not essential in *S. coelicolor*.



**Figure 4.14: The multiple sequence alignment of  $\sigma^{\text{HrdB}(2)}$  against  $\sigma^{\text{A}}$  and  $\sigma^{\text{B}}$  of *M. tuberculosis*.**

The Clustal W2 multiple sequence alignment tool and the BOXSHADE 3.21 server were used. Amino acids shown in white font with black highlight show identical residues whilst the amino acids in white font with grey highlight show similar residues. The homology was tested to the 0.7 fraction that must agree for shading. The most conserved domain 2 is indicated with arrows.

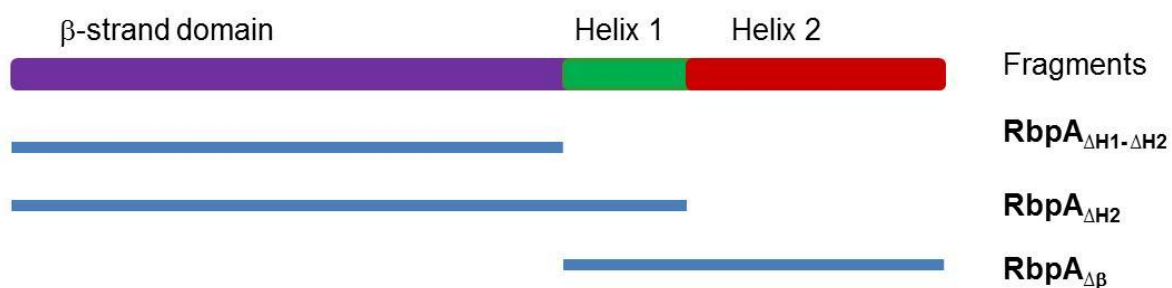
## **5- Mutagenesis of *rbpA***

## 5.0 Overview

In Chapter 4, RbpA was shown to interact with domain 2 of  $\sigma^{\text{HrdB}}$ . This chapter focuses on determining the region of RbpA that interacts with  $\sigma^{\text{HrdB}}$  using a combination of BACTH and site-directed mutagenesis. Constructed RbpA mutants were tested for interaction with  $\sigma^{\text{HrdB}}$  *in vivo* using BACTH and *in vitro* pull-down assays. Finally, *in vitro* transcription assays were performed to test whether selected mutations that appeared to affect RbpA function *in vivo*, influenced transcription initiation under defined conditions.

### 5.1 Identifying the $\sigma^{\text{HrdB}}$ binding region of RbpA

RbpA was found to interact with  $\sigma^{\text{HrdB}(2)}$  both *in vivo* and *in vitro* (Section 4.2). However, it was not known which region of RbpA was involved in this interaction. RbpA is predicted to have a  $\beta$ -strand region, a highly conserved Helix 1 and an amphiphatic helix, Helix 2 (Section 3.1.1). Therefore, three different truncated fragments of RbpA were designed and constructed as shown in Fig. 5.1 and Table 12: RbpA lacking Helix 1 and Helix 2 (RbpA $_{\Delta\text{H1-}\Delta\text{H2}}$ ); RbpA lacking Helix 2 (RbpA $_{\Delta\text{H2}}$ ); and RbpA containing only the C-terminal region, that is Helix 1 and Helix 2 (RbpA $_{\Delta\beta}$ ) (Fig. 5.1 and Table 12).



**Figure 5.1: Truncated fragments of RbpA.**

RbpA fragments were constructed containing the  $\beta$ -strand region only (RbpA $_{\Delta\text{H1-}\Delta\text{H2}}$ ), the  $\beta$ -strand and Helix 1 (RbpA $_{\Delta\text{H2}}$ ), and only Helix 1 and the Helix 2 region (RbpA $_{\Delta\beta}$ ).

**Table 12: The three different fragments designed for RbpA.**

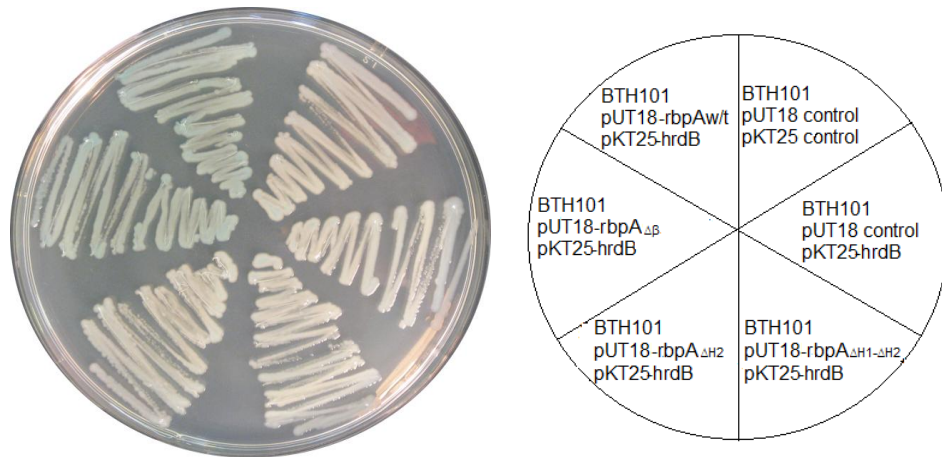
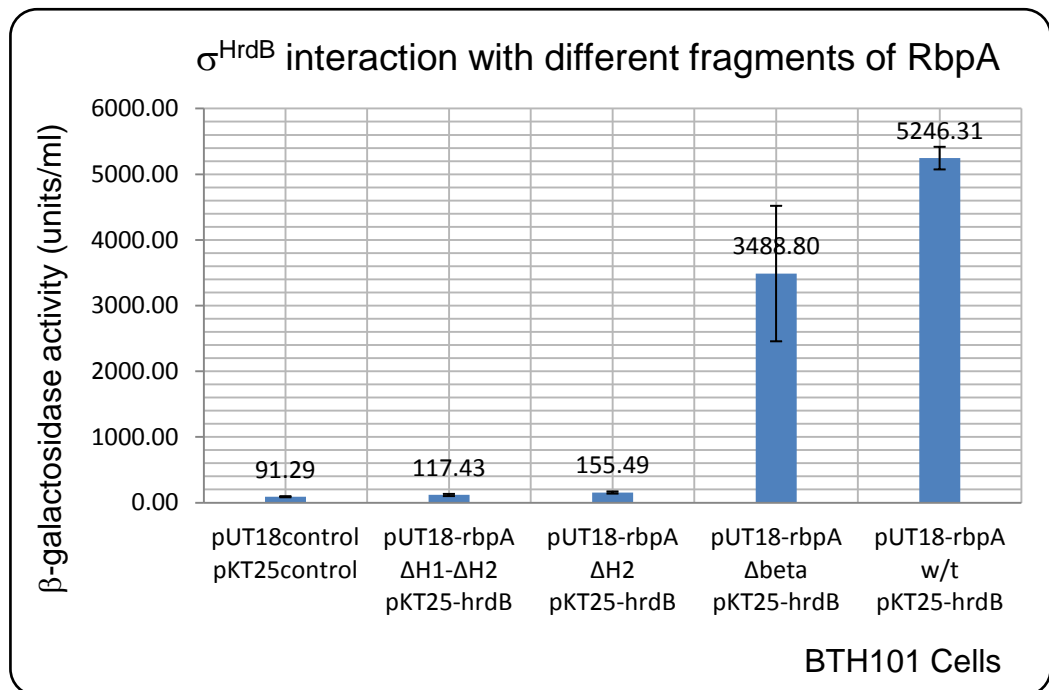
<b>RbpA Fragment</b>	<b>Regions of RbpA</b>	<b>Nucleotide length (bp)</b>	<b>Amino acids</b>	<b>N-terminal sequence</b>	<b>C-terminal sequence</b>
<b>RbpA<sub>ΔH1-ΔH2</sub></b>	β-strand region	~217	72	MSERA	GDGPE
<b>RbpA<sub>ΔH2</sub></b>	β-strand region and Helix 1	~261	90	MSERA	LMERR
<b>RbpA<sub>Δβ</sub></b>	Helix 1 and Helix 2	~158	52	EKKAK	SRKSA

### 5.1.1 $\sigma^{\text{HrdB}}$ interaction with RbpA fragments

The three truncated derivatives of RbpA, that is RbpA<sub>ΔH1-ΔH2</sub>, RbpA<sub>ΔH2</sub>, and RbpA<sub>Δβ</sub> were amplified using *T18\_rbpA<sub>ΔH1-ΔH2</sub>*, *T18\_rbpA<sub>ΔH2</sub>* and *T18\_rbpA<sub>Δβ</sub>* forward and reverse primers, respectively (Table 5- Section 2.1.5.1), and subcloned into pUT18 (Section 4.1.1). The recombinant plasmids pUT18-*rbpA<sub>ΔH1-ΔH2</sub>*, pUT18-*rbpA<sub>ΔH2</sub>* and pUT18-*rbpA<sub>Δβ</sub>* were co-transformed together with pKT25-*hrdB<sup>(2-4)</sup>* (Section 4.1.1) into *E. coli* BTH101 competent cells by electroporation. Controls were also prepared containing vector only as a negative control and containing pUT18-*rbpA* (Section 4.1.1) and pKT25-*hrdB<sup>(2-4)</sup>* as a positive control. Transformants were selected on LA with additional selection of X-gal and IPTG for initial assessment (Fig. 5.2A).

BTH101 containing pUT18-*rbpA<sub>Δβ</sub>* and pKT25-*hrdB<sup>(2-4)</sup>* together with the positive control produced blue colonies, whereas BTH101 containing pKT25-*hrdB<sup>(2-4)</sup>* and pUT18-*rbpA<sub>ΔH1-ΔH2</sub>*, pUT18-*rbpA<sub>ΔH2</sub>* and the negative control produced white colonies (Fig. 5.2A). To quantify and confirm these interactions β-galactosidase assays were performed (Section 2.6). BTH101 containing pUT18-*rbpA<sub>Δβ</sub>* and pKT25-*hrdB<sup>(2-4)</sup>* had 38-fold increase in β-galactosidase activity compared to vector only control (Fig. 5.2B). These results indicate that the C-terminal region of RbpA, including Helix 1 and Helix 2, interacts with  $\sigma^{\text{HrdB(2-4)}}$ . The assays also

suggest that the N-terminal region of RbpA, including the  $\beta$ -strand region, is not involved in the interaction with  $\sigma^{\text{HrdB}}$  (Fig. 5.2).

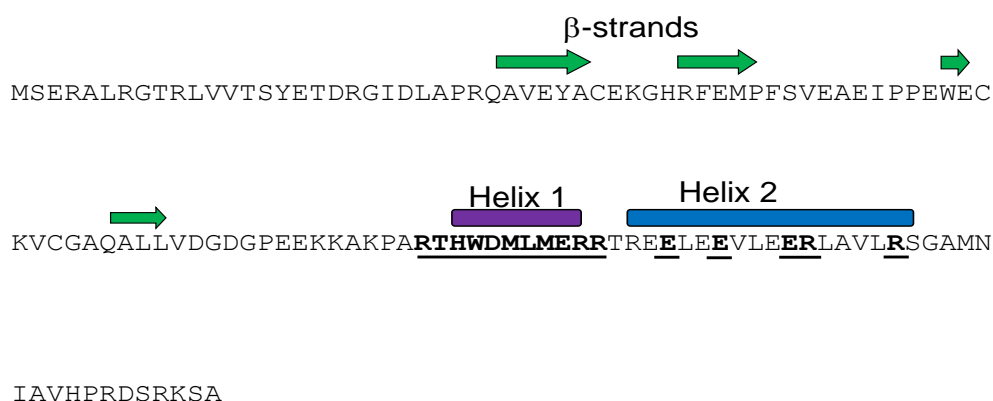
**A****B**

**Figure 5.2: BACTH assay of  $\sigma^{\text{HrdB}}$  interaction with truncated RbpA derivatives.**

**(A)** Blue/White screening on LA plates grown at 30°C for 16-24 h. **(B)**  $\beta$ -galactosidase assay of  $\sigma^{\text{HrdB}}$  interaction with the constructed fragments of RbpA: RbpA $_{\Delta H1-\Delta H2}$ , RbpA $_{\Delta H2}$ , and RbpA $_{\Delta\beta}$ . The  $\beta$ -galactosidase activity was measured for BTH101 containing pKT25-*hrdB* and pUT18-*rbpA* $_{\Delta H1-\Delta H2}$ , pUT18-*rbpA* $_{\Delta H2}$ , pUT18-*rbpA* $_{\Delta\beta}$ , and pUT18-*rbpA*<sub>w/t</sub> together with vector only control. Three individual colonies were tested and the assays were done in triplicate for each colony with error bars showing the standard deviation.

## 5.2 Site-directed mutagenesis of C-terminal region of RbpA

The C-terminal region of RbpA was shown to interact with  $\sigma^{\text{HrdB}}$ , using BACTH assays (Section 5.1.1). In order to identify amino acids that are crucial for this interaction, conserved residues within the C-terminal region of RbpA were selected (Fig. 5.3) and substituted by alanine using site-directed mutagenesis. Initially, the mutants were introduced into a *S. coelicolor*  $\Delta rbpA$  strain to test for complementation (Section 5.2). Subsequently, selected RbpA mutant proteins were checked for interaction with  $\sigma^{\text{HrdB}}$  (Section 5.3 & 5.4) and whether they could stimulate transcription initiation (Section 5.6).



**Figure 5.3: The amino acid sequence and secondary structure prediction of RbpA.**

The horizontal arrows represent the  $\beta$ -strands (green arrows), Helix 1 (purple cylinder) and the Helix 2 (blue cylinder). Residues underlined in Helix 1, ERR motif and Helix 2 were selected for site-directed mutagenesis.

### 5.2.1 Constructing *rbpA* mutants

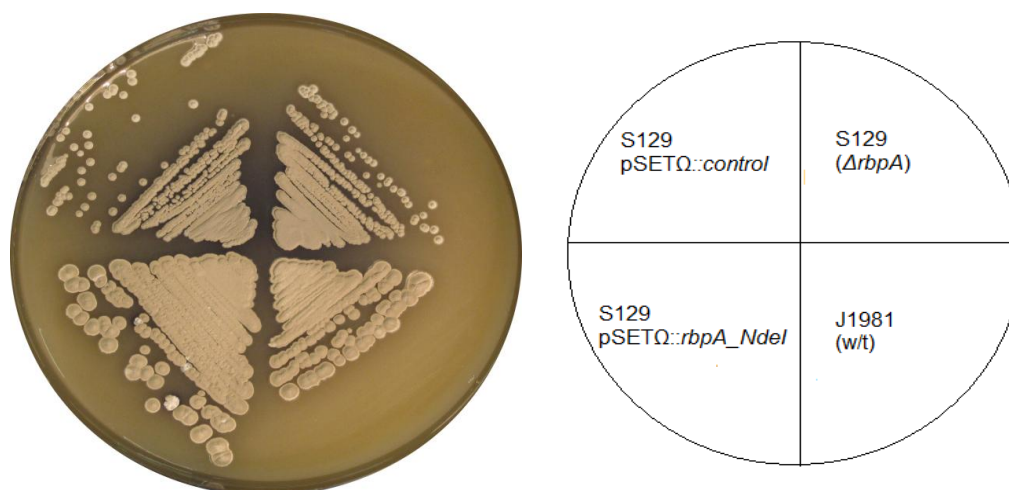
In order to allow the analysis of the constructed *rbpA* mutants *in vivo* and *in vitro*, an NdeI site was engineered at the start codon of the *rbpA* gene in the native promoter-ORF context. The *rbpA* gene together with its promoter region already present in pMT3000 vector (Table 1- Section 2.1.2) was amplified simultaneously introducing an NdeI site, using the *rbpA\_NdeI* forward and



reverse primers (Table 6- Section 2.1.5.2). The amplified pMT3000::*rbpA\_NdeI* was sequenced and the *rbpA\_NdeI* fragment subcloned into pSET $\Omega$  (Table 1- Section 2.1.2 & Fig.A4- Appendices). Prior to the use of this construct to analyse *rbpA* mutants, it was important to check whether this engineered gene, *rbpA\_NdeI*, works as the wild-type *rbpA* and that the NdeI site had not disrupted gene function.

pSET $\Omega$ ::*rbpA\_NdeI* was used to transform *E. coli* ETZ competent cells and the transformants inoculated in LB containing appropriate antibiotic selection. The inoculants were then conjugated into *S. coelicolor* strains: J1981 (w/t) and S129 ( $\Delta$ *rbpA*) (Section 2.2.5.1). Vector only control strains of J1981 and S129 were also generated. The ex-conjugants were left to grow at 30°C for 4-5 days and then streaked out into single colonies. A single colony was selected from each ex-conjugant and streaked into one MS plate to check for any phenotypic effect on *S. coelicolor* (Fig. 5.4).

*S. coelicolor* J1981 (w/t), grew normally producing normal sized colonies and the appropriate production of the antibiotic actinorhodin (Fig. 5.4). S129,  $\Delta$ *rbpA* strain, together with S129 vector only control formed small sized colonies and had an increase in actinorhodin production (Fig. 5.4). However, when pSET $\Omega$ ::*rbpA\_NdeI* was introduced into S129, normal sized colonies were formed and the appropriate production of actinorhodin (Fig. 5.4). Therefore, *rbpA\_NdeI* complemented the  $\Delta$ *rbpA* strain, which indicates that the newly engineered *rbpA\_NdeI* is functional.



**Figure 5.4: Complementation of *S. coelicolor* S129 strain by introducing the engineered *rbpA\_Ndel* vector.**

*S. coelicolor* strains: J1981 (w/t), S129 (Δ*rbpA*), S129 (pSETΩ::control), and S129 (pSETΩ::*rbpA\_Ndel*) were streaked into single colonies on MS agar and incubated at 30°C for 4-5 days.

### 5.2.2 Helix 1

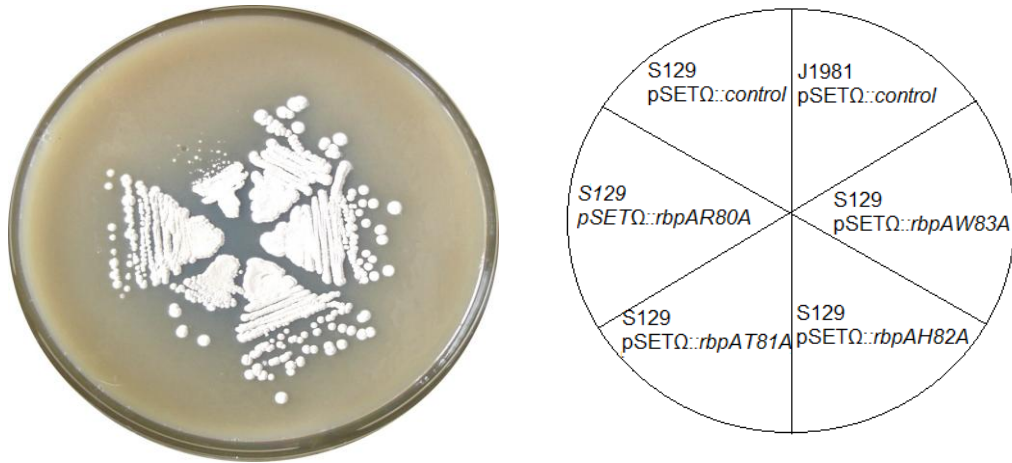
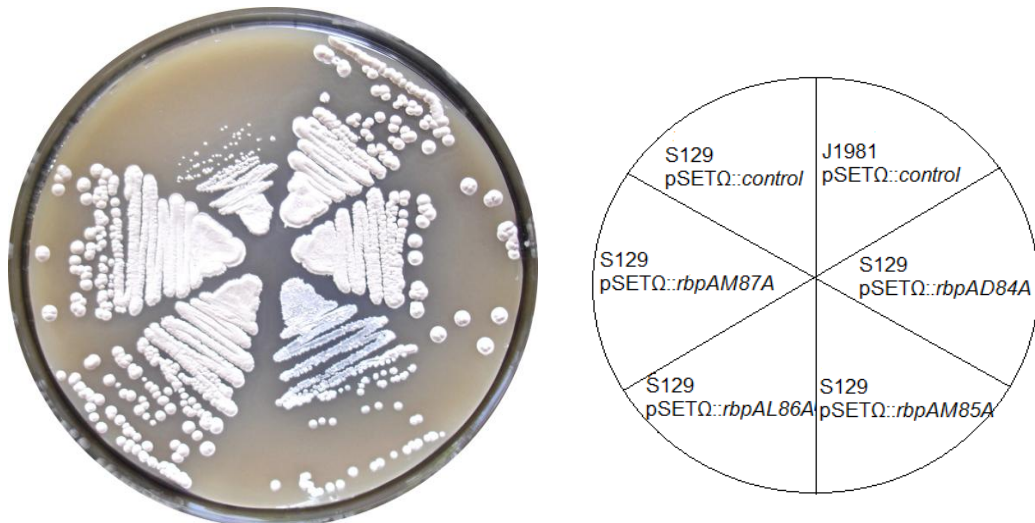
Site-directed mutagenesis was performed on all eight residues present in the predicted Helix 1 of RbpA; R80, T81, H82, W83, D84, M85, L86, and M87 (Fig. 5.2). These residues were selected because they are highly conserved amongst RbpA homologues (Fig. 3.1). The residues were mutated individually to alanine (A), which would remove any bulky or charged R group from these amino acids that may be involved in structural and/or functional interactions with other proteins.

Using pMT3000::*rbpA\_Ndel*, the forward and reverse primers of *rbpA*<sub>R80A</sub>, *rbpA*<sub>T81A</sub>, *rbpA*<sub>H82A</sub>, *rbpA*<sub>W83A</sub>, *rbpA*<sub>D84A</sub>, *rbpA*<sub>M85A</sub>, *rbpA*<sub>L86A</sub>, and *rbpA*<sub>M87A</sub> (Table 6- Section 2.1.5.2) were used to introduce the alanine substitution by inverse PCR mutagenesis (Section 2.2.1.2). The pMT3000::*rbpA* plasmids containing R80A, T81A, H82A, W83A, D84A, M85A, L86A and M87A substitutions were sequenced and each mutant *rbpA* fragment was extracted with BglII and subcloned into pSETΩ. pSETΩ containing *rbpA*<sub>R80A</sub>, *rbpA*<sub>T81A</sub>, *rbpA*<sub>H82A</sub>, *rbpA*<sub>W83A</sub>, *rbpA*<sub>D84A</sub>, *rbpA*<sub>M85A</sub>, *rbpA*<sub>L86A</sub>, and *rbpA*<sub>M87A</sub> were individually transformed into ETZ competent cells and were then conjugated into *S.*

*coelicolor* J1981 (w/t) and S129 ( $\Delta rbpA$ ) strains (Section 2.2.5.1). Vector only controls were also prepared. The ex-conjugants were left to grow at 30°C for 4-5 days, then streaked out into single colonies and spores harvested.

#### 5.2.2.1 Effect of mutations in RbpA helix 1 on the phenotype of *S. coelicolor*

A single colony was selected from each mutant ex-conjugant and streaked onto MS agar plates to check for phenotype, morphology and antibiotic production (Fig. 5.5). J1981 vector only control strain produced appropriate sized colonies, whilst S129 vector only control strain produced small sized colonies and an increase in the production of actinorhodin (Fig. 5.5). S129 containing pSET $\Omega$ ::*rbpA*<sub>R80A</sub> and pSET $\Omega$ ::*rbpA*<sub>M85A</sub> allowed partial complementation as medium sized colonies were formed (Fig. 5.5). The remaining six *rbpA* mutants restored appropriate sized colonies and therefore, appeared to fully complement the  $\Delta rbpA$  strains (Fig. 5.5).

**A****B**

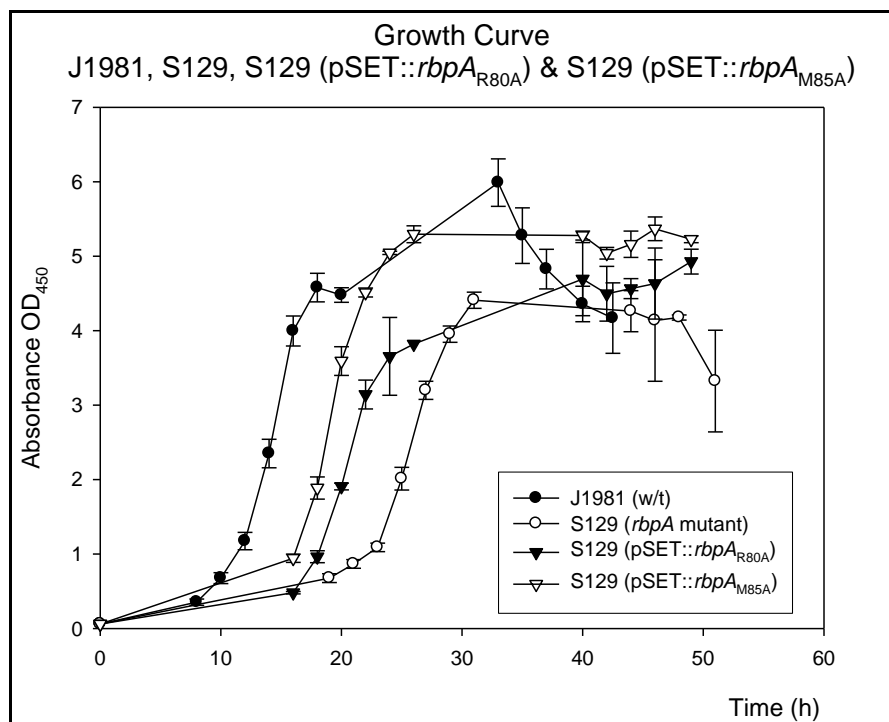
**Figure 5.5: Complementation of *S. coelicolor* S129 strains by introducing mutated residues of Helix 1 of RbpA.**

The strains were plated on MS agar with 100 µg/µl of spectinomycin and incubated at 30°C for 4-5 days. **(A)** Showing S129 (pSETΩ::rbpA<sub>R80A</sub>), S129 (pSETΩ::rbpA<sub>T81A</sub>), S129 (pSETΩ::rbpA<sub>H82A</sub>), S129 (pSETΩ::rbpA<sub>W83A</sub>) together with S129 and J1981 vector only controls. **(B)** Showing S129 (pSETΩ::rbpA<sub>D84A</sub>), S129 (pSETΩ::rbpA<sub>M85A</sub>), S129 (pSETΩ::rbpA<sub>L86A</sub>), S129 (pSETΩ::rbpA<sub>M87A</sub>) together with S129 and J1981 vector only controls.

#### 5.2.2.2 Effect of mutations in RbpA helix 1 on the growth curve of *S. coelicolor*

S129 (pSETΩ::*rbpA*) containing R80A or M85A substitution was shown to partially complement the  $\Delta$ *rbpA* phenotype (Fig. 5.5). Therefore, to confirm these results and determine whether the mutants affect the *S. coelicolor* growth, a growth curve was generated for S129 (pSETΩ::*rbpA*<sub>R80A</sub>), S129 (pSETΩ::*rbpA*<sub>M85A</sub>), S129 ( $\Delta$ *rbpA*) and J1981 (w/t) strains. The spores were pre-germinated with a starting OD<sub>450</sub> of 0.06 for 3 h and then inoculated in YEME (10% (w/v) sucrose) liquid media (Section 2.2.5.3) with 50 µg/µl spectinomycin at 30°C with vigorous shaking. At 2 h intervals, the OD<sub>450</sub> readings were taken (Section 2.2.5.3).

S129 containing pSETΩ::*rbpA*<sub>R80A</sub> had a slightly shorter lag phase than that of  $\Delta$ *rbpA* strain but entered stationary phase at an earlier lower OD reading (~4) relative to J1981 (w/t). S129 containing pSETΩ::*rbpA*<sub>M85A</sub> had a much shorter lag phase compared to S129 and grew to a similar OD reading (~5) to that of J1981 (~6). Therefore, M85A has a very little effect on *S. coelicolor* growth whilst R80A has a more significant effect on the growth of *S. coelicolor* (Fig. 5.6).



**Figure 5.6: A graphical representation of the growth curve of J1981 (w/t), S129 ( $\Delta$ *rbpA*), S129 (pSETΩ::*rbpA*<sub>R80A</sub>) and S129 (pSETΩ::*rbpA*<sub>M85A</sub>).**

The graph was generated using SigmaPlot. Absorbance reading at OD<sub>450</sub> was recorded at 2 h intervals for each of the *S. coelicolor* strains.

### 5.2.3 The ERR motif

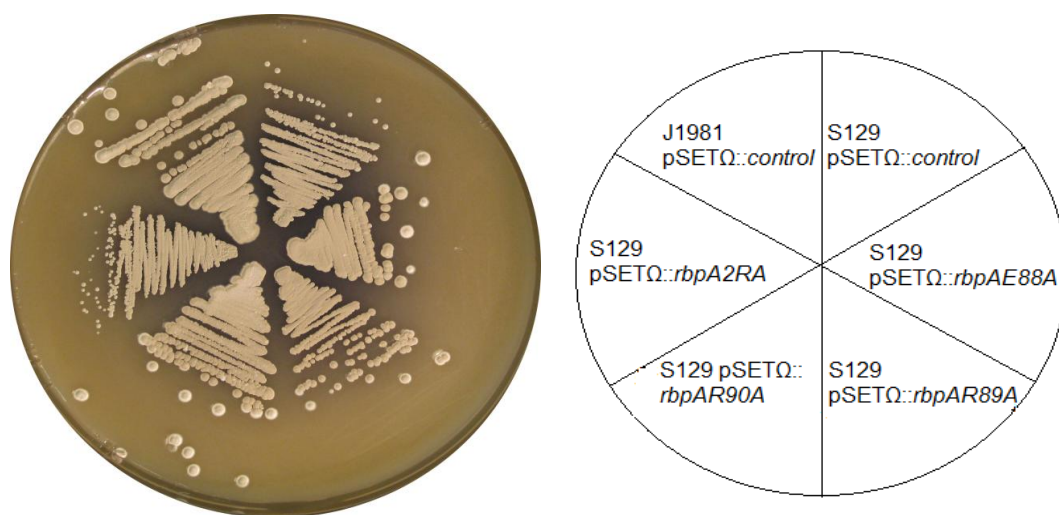
The ERR motif lies between the two helices of RbpA: Helix 1 and Helix 2. It is highly conserved amongst the RbpA and RbpB homologues (Fig. 3.1 & 3.2). Hence, E88, R89 and R90 were mutated individually to alanine (E88A, R89A and R90A), then the two arginines, R89 and R90, were mutated collectively to two alanines (2RA), and the mutants were tested for complementation of *S. coelicolor* S129 ( $\Delta$ *rbpA*).

*rbpA*<sub>E88A</sub>, *rbpA*<sub>R89A</sub>, *rbpA*<sub>R90A</sub> and *rbpA*<sub>2RA</sub> forward and reverse primers (Table 6-Section 2.1.5.2) were used to introduce their respective mutations into pMT3000::*rbpA*\_NdeI by inverse PCR (Section 2.2.1.2). The vectors were sequenced and the respective BglII fragments were subcloned into pSETΩ (see above). pSETΩ::*rbpA* containing E88A, R89A, R90A, and 2RA substitutions

were conjugated into *S. coelicolor* J1981 (w/t) and S129 ( $\Delta rbpA$ ) as described (Section 5.2.2).

#### 5.2.3.1 Effect of mutations in RbpA ERR motif on the phenotype of *S. coelicolor*

The expression of RbpA<sub>E88A</sub> or RbpA<sub>R90A</sub> in S129 complemented the  $\Delta rbpA$  phenotype, producing appropriate sized colonies (Fig. 5.7). However, expression of RbpA<sub>R89A</sub> in S129 only partially complemented the deletion strain (Fig. 5.7). Interestingly, the *rbpA*<sub>2RA</sub> allele, with both arginine residues mutated, did not complement S129, such that the strain formed small sized colonies and over-produced actinorhodin (Fig. 5.7).



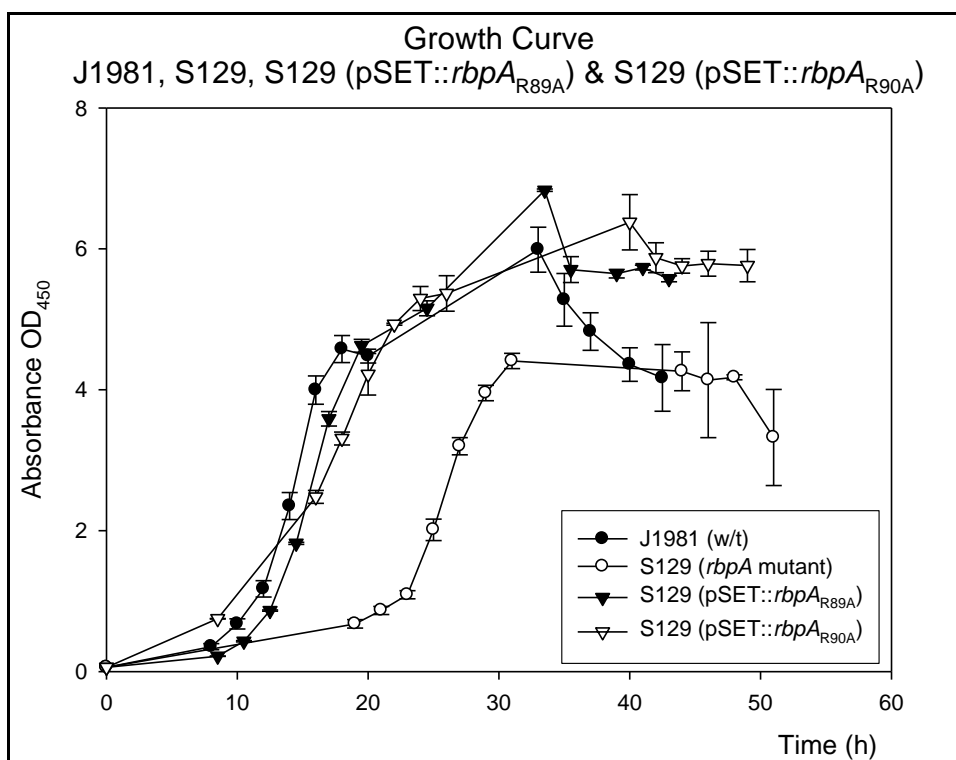
**Figure 5.7: Complementation of *S. coelicolor* S129 strains by introducing mutated residues of the ERR motif of RbpA.**

The spores of S129 (pSETΩ::rbpA<sub>E88A</sub>), S129 (pSETΩ::rbpA<sub>R89A</sub>), S129 (pSETΩ::rbpA<sub>R90A</sub>), and S129 (pSETΩ::rbpA<sub>2RA</sub>), together with S129 and J1981 vector only control were plated on MS agar with 100 µg/µl spectinomycin and incubated at 30°C for 4-5 days.

### 5.2.3.2 Effect of mutations in RbpA ERR motif on the growth curve of *S. coelicolor*

A growth curve was generated for S129 (pSET $\Omega$ ::*rbpA*<sub>R89A</sub>), S129 (pSET $\Omega$ ::*rbpA*<sub>R90A</sub>), S129 ( $\Delta$ *rbpA*) and J1981 (w/t) vector only control strains as explained in Section 5.2.2.2.

S129 (pSET $\Omega$ ::*rbpA*<sub>R89A</sub>) had a relatively shorter lag phase compared to S129 and entered stationary phase at a similar OD<sub>450</sub> to that of J1981 (w/t) (Fig. 5.8). S129 (pSET $\Omega$ ::*rbpA*<sub>R90A</sub>), on the other hand, had a very short lag phase (~8 h) and grew almost comparable to that of J1981 entering stationary phase at approximately the same OD<sub>450</sub> of J1981 (Fig. 5.8). The results appear to be consistent with the phenotypic results obtained in Fig. 5.7, suggesting that individual alanine substitutions of R89 or R90 can be “rescued” by the other arginine. However, when both are mutated, the RbpA mutant loses its functionality (Section 5.3.2).



**Figure 5.8: A graphical representation of the growth curve of J1981 (w/t), S129 ( $\Delta$ *rbpA*), S129 (pSET $\Omega$ ::*rbpA*<sub>R89A</sub>) and S129 (pSET $\Omega$ ::*rbpA*<sub>R90A</sub>).**

The graph was generated using SigmaPlot. Absorbance at OD<sub>450</sub> was recorded at 2 h intervals for each *S. coelicolor* strain.



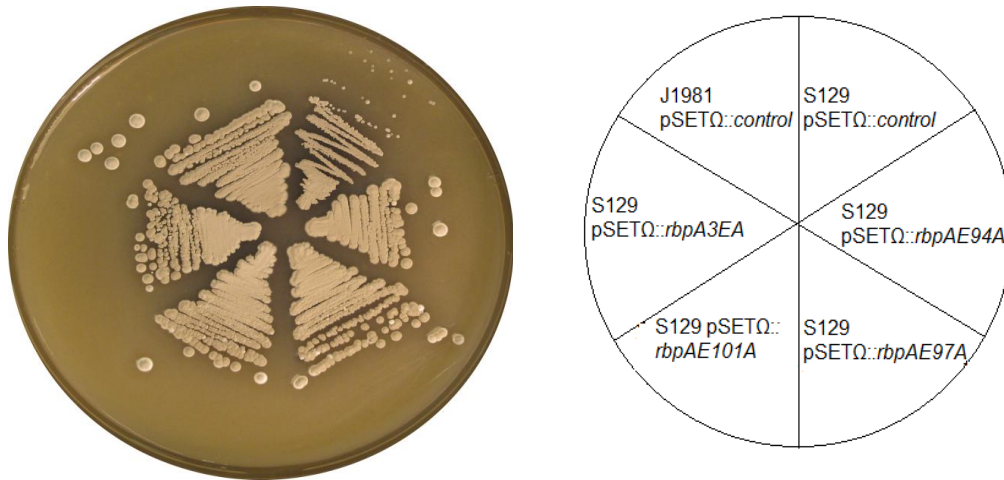
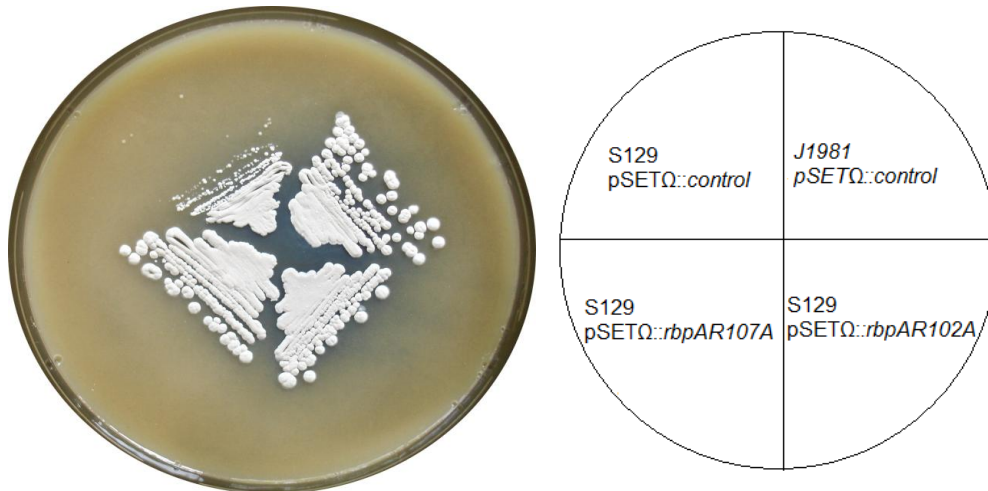
### 5.2.4 Helix 2

Helix 2 of RbpA is amphipathic and is defined by the presence of acidic, hydrophilic residues (E) on one side and non-polar, hydrophobic residues (L & V) on the other (Section 3.1.3.2). Three glutamates were selected, and mutated to alanine individually (E94A, E97A and E101A) and collectively (3EA). In addition, two conserved arginines were individually substituted to alanine (R102A and R107A).

The respective mutations were introduced into pMT3000::*rbpA\_NdeI* by inverse PCR, fragments extracted and subcloned into pSETΩ. pSETΩ::*rbpA* containing E94A, E97A, E101A, 3EA, R102A, and R107A substitutions were conjugated into *S. coelicolor* J1981 (w/t) and S129 ( $\Delta$ *rbpA*) as described (Section 5.2.2).

#### 5.2.4.1 Effect of mutations in RbpA helix 2 on the phenotype of *S. coelicolor*

Expression of RbpA<sub>E94A</sub>, RbpA<sub>E97A</sub>, RbpA<sub>E101A</sub>, RbpA<sub>3EA</sub>, RbpA<sub>R102A</sub> or RbpA<sub>R107A</sub> in S129 complemented the  $\Delta$ *rbpA* phenotype, producing appropriate sized colonies (Fig. 5.9).

**A****B**

**Figure 5.9: Complementation of *S. coelicolor* S129 strains by introducing mutated residues of Helix 2 of RbpA.**

The spores were plated on MS agar containing 50  $\mu\text{g}/\mu\text{l}$  spectinomycin and incubated at 30°C for 4-5 days. **(A)** Showing S129 (pSETΩ::*rbpA*<sub>E94A</sub>), S129 (pSETΩ::*rbpA*<sub>E97A</sub>), S129 (pSETΩ::*rbpA*<sub>E101A</sub>), and S129 (pSETΩ::*rbpA*<sub>3EA</sub>) together with S129 ( $\Delta$ *rbpA*) and J1981 (w/t) vector only controls. **(B)** Showing S129 (pSETΩ::*rbpA*<sub>R102A</sub>) and S129 (pSETΩ::*rbpA*<sub>R107A</sub>) together with S129 and J1981 vector only controls.

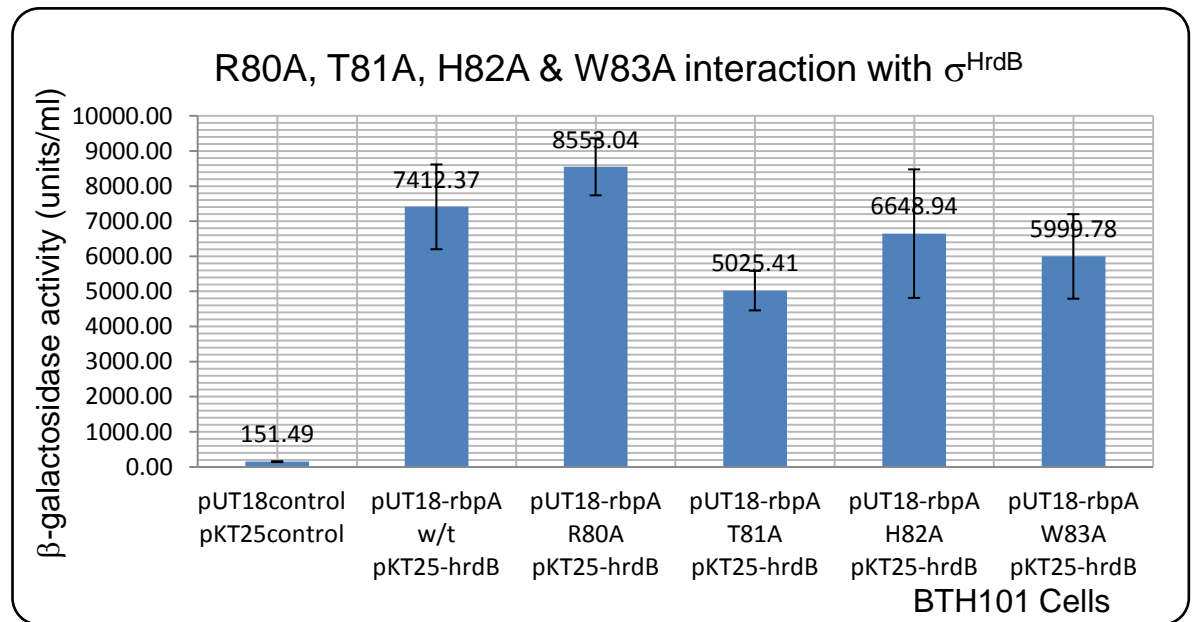
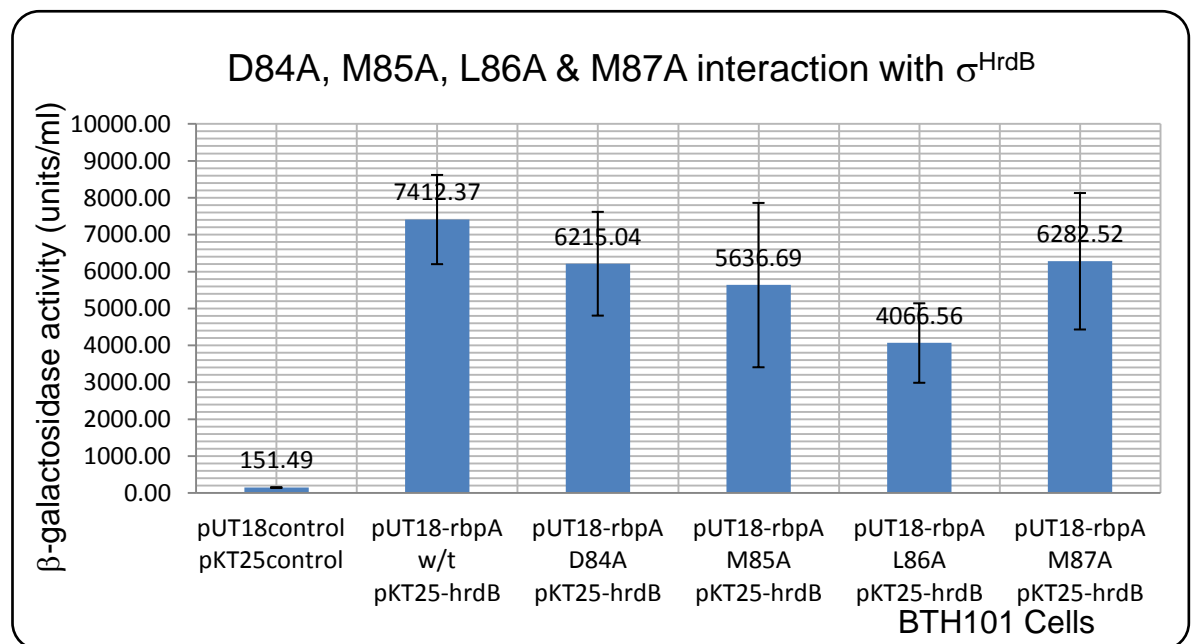
### **5.3 BACTH analysis of $\sigma^{\text{HrdB}}$ interaction with RbpA mutants**

Site-directed mutagenesis of RbpA has revealed that some residues in RbpA affect the phenotype, growth and antibiotic production of *S. coelicolor*. However, how these mutated residues affect the function of RbpA is not clear. Do these mutants still interact with the principal sigma factor,  $\sigma^{\text{HrdB}}$ ? If yes, does that mean that they sequester  $\sigma^{\text{HrdB}}$  and prevent it from binding to RNAP and thus hinder transcription? Therefore, the first step was to test whether these RbpA mutant proteins generated interact with  $\sigma^{\text{HrdB}}$  using BACTH assays.

#### **5.3.1 Helix 1**

All eight Helix 1 RbpA mutants were tested for interaction with the principal sigma factor,  $\sigma^{\text{HrdB}}$ . The engineered pMT3000::*rbpA* containing R80A, T81A, H82A, W83A, D84A, M85A, L86A and M87A substitutions were used to extract the mutant fragments as a *SacI*-*XhoI* fragment and then individually subcloned into pUT18-*rbpA* (Section 4.1.1), which was also digested with *SacI* and *XhoI*. The recombinant pUT18-*rbpA* containing R80A, T81A, H82A, W83A, D84A, M85A, L86A and M87A substitutions were sequenced and individually co-transformed with pKT25-*hrdB*<sup>(2-4)</sup> into BTH101 competent cells by electroporation. BTH101 vector only controls were also prepared. To determine and quantify any interactions, the  $\beta$ -galactosidase assays were performed (Section 2.6).

BTH101 containing pKT25-*hrdB*<sup>(2-4)</sup> and pUT18-*rbpA*, pUT18-*rbpA*<sub>R80A</sub>, pUT18-*rbpA*<sub>T81A</sub>, pUT18-*rbpA*<sub>H82A</sub>, pUT18-*rbpA*<sub>W83A</sub>, pUT18-*rbpA*<sub>D84A</sub>, pUT18-*rbpA*<sub>M85A</sub>, pUT18-*rbpA*<sub>L86A</sub>, and pUT18-*rbpA*<sub>M87A</sub> had 27- to 56- fold increase in  $\beta$ -galactosidase activity compared to the vector only control (Fig. 5.10). These assays indicated that RbpA<sub>R80A</sub>, RbpA<sub>T81A</sub>, RbpA<sub>H82A</sub>, RbpA<sub>W83A</sub>, RbpA<sub>D84A</sub>, RbpA<sub>M85A</sub>, RbpA<sub>L86A</sub> and RbpA<sub>M87A</sub> proteins interact with  $\sigma^{\text{HrdB}}$ .

**A****B**

**Figure 5.10: BACTH assay of the interaction of  $\sigma^{\text{HrdB}}$  with the mutated residues in Helix 1 of RbpA.**

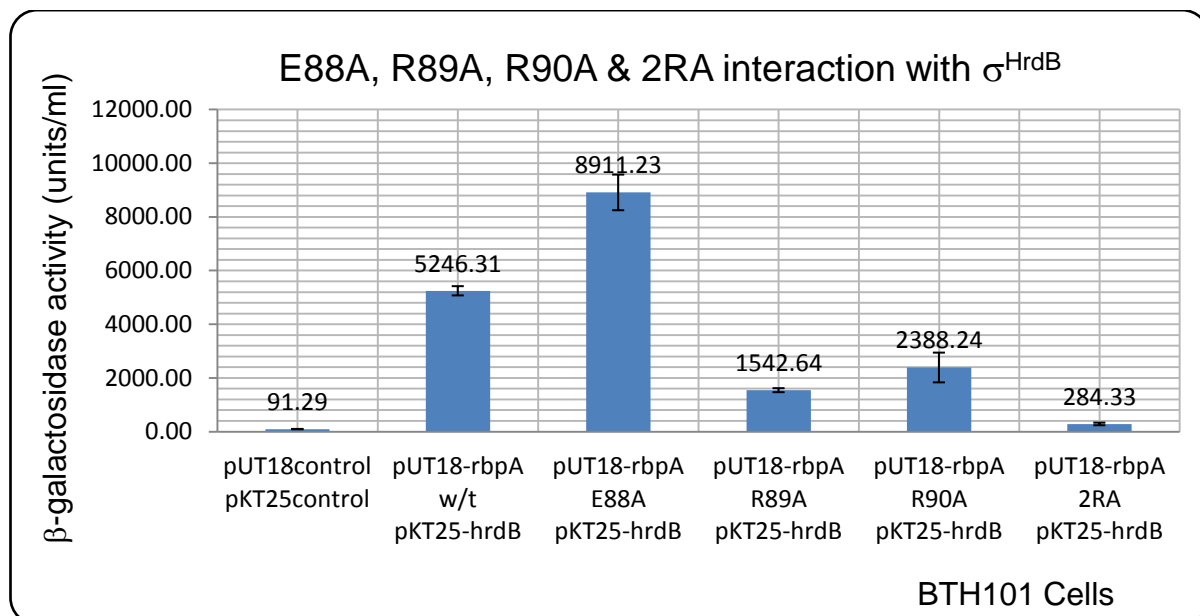
**(A)** Showing the interaction of RbpA<sub>w/t</sub>, RbpA<sub>R80A</sub>, RbpA<sub>T81A</sub>, RbpA<sub>H82A</sub> and RbpA<sub>W83A</sub> with  $\sigma^{\text{HrdB}}$ . **(B)** Showing the interaction of RbpA<sub>w/t</sub>, RbpA<sub>D84A</sub>, RbpA<sub>M85A</sub>, RbpA<sub>L86A</sub>, and RbpA<sub>M87A</sub> with  $\sigma^{\text{HrdB}}$ . These assays were repeated using three separate colonies in triplicate with error bars showing standard deviation.

### 5.3.2 The ERR motif

The two arginines, R89 and R90, in the ERR motif appeared to have a significant effect on the phenotype and growth of *S. coelicolor* (Fig. 5.2.3). BACTH assays were used to determine interactions of RbpA<sub>E88A</sub>, RbpA<sub>R89A</sub>, RbpA<sub>R90A</sub> and RbpA<sub>2RA</sub> proteins with  $\sigma^{\text{HrdB}}$  *in vivo*.

A SacI-XhoI fragment of *rbpA*<sub>E88A</sub>, *rbpA*<sub>R89A</sub>, *rbpA*<sub>R90A</sub> and *rbpA*<sub>2RA</sub> were extracted from pMT3000::*rbpA* containing the respective mutations, individually subcloned into pUT18-*rbpA* and recombinants were co-transformed separately into BTH101 competent cells together with pKT25-*hrdB*<sup>(2-4)</sup> as explained in 5.3.1. The  $\beta$ -galactosidase assays were performed to test whether the RbpA mutant proteins bind to  $\sigma^{\text{HrdB}}$  (Section 2.6).

BTH101 containing pKT25-*hrdB*<sup>(2-4)</sup> and pUT18-*rbpA*, pUT18-*rbpA*<sub>E88A</sub>, pUT18-*rbpA*<sub>R89A</sub>, and pUT18-*rbpA*<sub>R90A</sub> had 57-, 98-, 16-, and 26-fold higher  $\beta$ -galactosidase activity compared to the vector only control, respectively. However, BTH101 containing pUT18-*rbpA*<sub>2RA</sub> and pKT25-*hrdB*<sup>(2-4)</sup> had no significant difference in  $\beta$ -galactosidase activity compared to the vector only control. These assays show that RbpA<sub>E88A</sub>, RbpA<sub>R90A</sub> and to some extent RbpA<sub>R89A</sub> proteins interact with  $\sigma^{\text{HrdB}}$ , whilst the 2RA substitution abolished RbpA- $\sigma^{\text{HrdB}}$  interaction (Fig. 5.11).



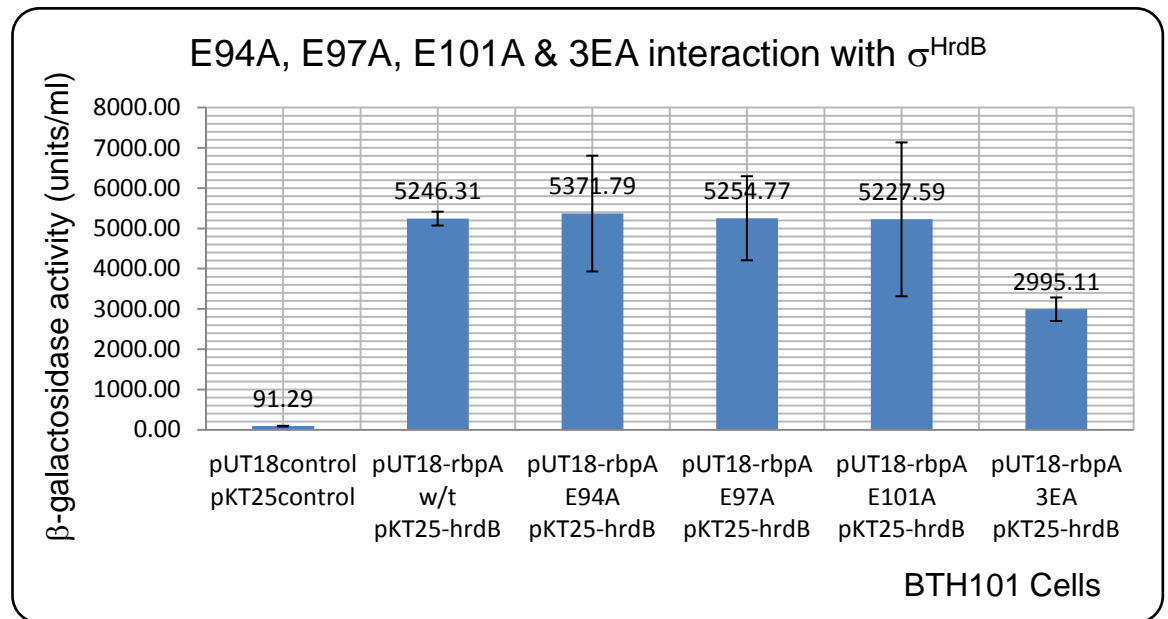
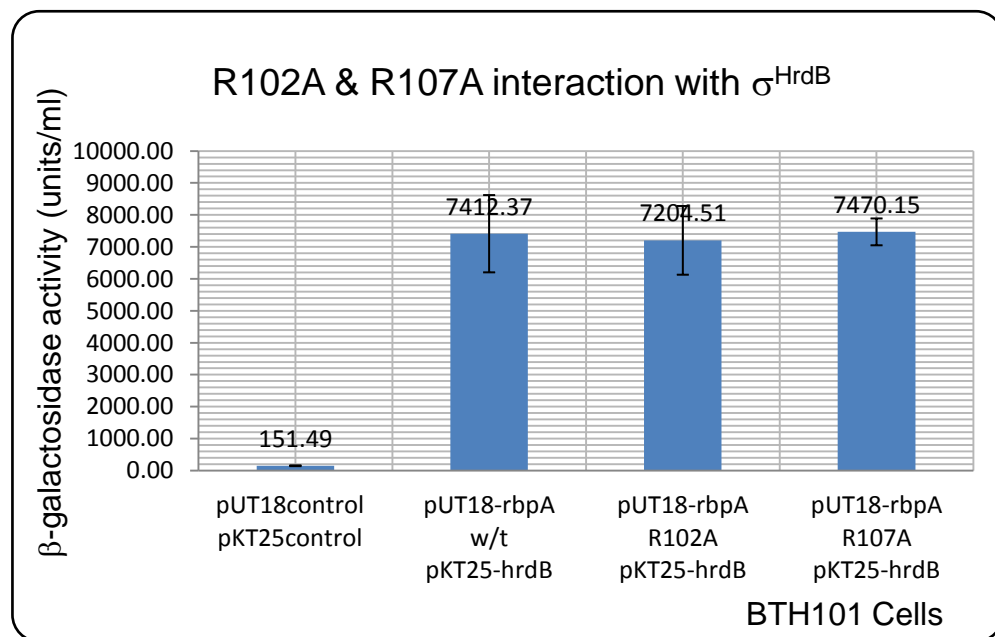
**Figure 5.11: BACTH assay of the interaction of  $\sigma^{\text{HrdB}}$  with the mutated residues of the ERR motif of RbpA.**

The  $\beta$ -galactosidase assays were performed on BTH101 containing pKT25-*hrdB*<sup>(2-4)</sup> and pUT18-*rbpA*<sub>w/t</sub>, pUT18-*rbpA*<sub>E88A</sub>, pUT18-*rbpA*<sub>R89A</sub>, pUT18-*rbpA*<sub>R90A</sub>, pUT18-*rbpA*<sub>2RA</sub> and vector only control. These assays were repeated for three separate colonies each in triplicate with error bars showing the standard deviation.

### 5.3.3 Helix 2

The residues mutated in Helix 2 did not show any relevant phenotypic effect in *S. coelicolor* (Fig. 5.9). Nonetheless, the interaction of these mutated RbpA residues with the principal sigma factor,  $\sigma^{\text{HrdB}}$ , was tested using BACTH. RbpA<sub>E101A</sub>, RbpA<sub>3EA</sub>, RbpA<sub>R102A</sub> and RbpA<sub>R107A</sub> were amplified using *T18\_rbpA* forward and reverse primers (Table 5- Section 2.1.5.1) with pMT3000::*rbpA* containing E101A, 3EA, R102A and R107A mutation, respectively. A *SacI*-*XhoI* fragment of RbpA<sub>E94A</sub> and RbpA<sub>E97A</sub> were extracted from pMT3000::*rbpA*<sub>E94A</sub> and pMT3000::*rbpA*<sub>E97A</sub>, respectively. All six fragments were then subcloned into pUT18-*rbpA* and the recombinants were co-transformed individually into BTH101 competent cells together with pKT25-*hrdB*<sup>(2-4)</sup> as explained in 5.3.1. To determine the interaction of these mutant proteins, the  $\beta$ -galactosidase assays were performed (Section 2.6).

The  $\beta$ -galactosidase activity for BTH101 containing pKT25-*hrdB*<sup>(2-4)</sup> and pUT18-*rbpA*, pUT18-*rbpA* containing E94A, E97A, E101A, R102, and R107A substitutions were around 50-fold higher compared to vector only control, whilst BTH101 containing pUT18-*rbpA*<sub>3EA</sub> and pKT25-*hrdB*<sup>(2-4)</sup> was 32-fold higher than vector only control (Fig. 5.12). The results indicate that RbpA<sub>E94</sub>, RbpA<sub>E97A</sub>, RbpA<sub>E101A</sub>, RbpA<sub>3EA</sub>, RbpA<sub>R102A</sub> and RbpA<sub>R107A</sub> proteins interact with  $\sigma^{\text{HrdB}}$ .

**A****B**

**Figure 5.12: BACTH assay of the interaction of  $\sigma^{\text{HrdB}}$  with mutated RbpA residues in Helix 2.**

**(A)** Showing the interaction of RbpA<sub>w/t</sub>, RbpA<sub>E94</sub>, RbpA<sub>E97A</sub>, RbpA<sub>E101A</sub>, and RbpA<sub>3EA</sub> proteins with  $\sigma^{\text{HrdB}}$ . **(B)** Showing the interaction of RbpA<sub>w/t</sub>, RbpA<sub>R102A</sub>, and RbpA<sub>R107A</sub> proteins with  $\sigma^{\text{HrdB}}$ . These assays were repeated using three separate colonies each in triplicate with error bars showing standard deviation.



#### **5.4 $\sigma^{\text{HrdB}}$ interaction with selected RbpA mutants *in vitro***

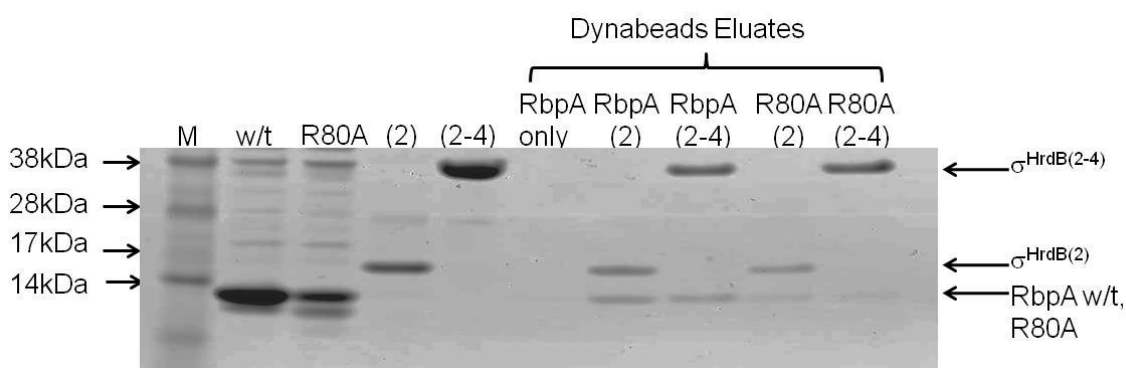
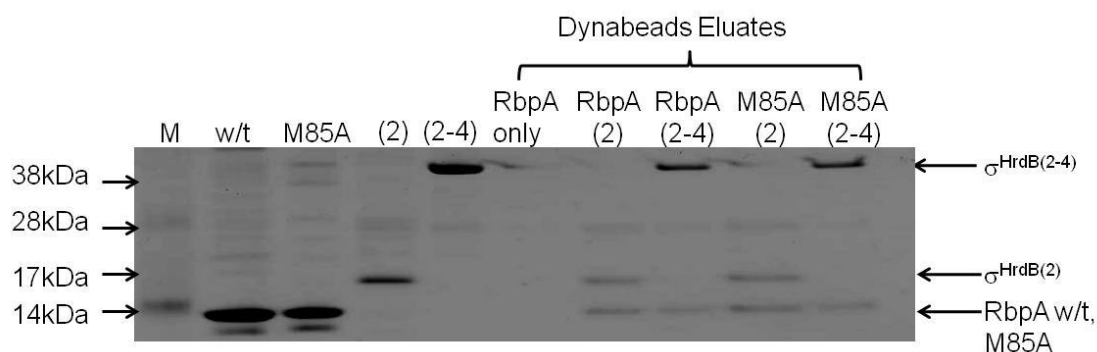
Some of the RbpA mutants generated appeared to have an effect on the phenotype of *S. coelicolor* strain and/or affect sigma interaction. RbpA<sub>R80A</sub>, RbpA<sub>M85A</sub>, RbpA<sub>R90A</sub>, and to some extent RbpA<sub>R89A</sub> proteins were shown to interact with  $\sigma^{\text{HrdB}}$  using BACTH (Section 5.3.1 & 5.3.2) yet appeared to affect to some degree the *S. coelicolor* growth (Section 5.2.2 & 5.2.3). On the other hand, RbpA<sub>2RA</sub> protein was shown to have lost its interaction with  $\sigma^{\text{HrdB}}$  (Section 5.3.2). Thus, these RbpA mutant proteins were selected to further confirm the BACTH results.

##### **5.4.1 RbpA<sub>R80A</sub> and RbpA<sub>M85A</sub>**

RbpA<sub>R80A</sub> and RbpA<sub>M85A</sub> proteins were shown to interact with the principal sigma factor,  $\sigma^{\text{HrdB}}$ , despite only partially complementing the  $\Delta rbpA$  phenotype (Fig. 5.10 & 5.5). To verify the results of the RbpA mutant- $\sigma^{\text{HrdB}}$  interactions, *in vitro* pull down assays were performed. The *rbpA*<sub>R80A</sub> and *rbpA*<sub>M85A</sub> alleles were subcloned into pET20b from pMT3000::*rbpA*<sub>R80A</sub> and pMT3000::*rbpA*<sub>M85A</sub> as NdeI-BamHI fragments in order to produce non-tagged RbpA mutant proteins. pET20b::*rbpA*<sub>R80A</sub> and pET20b::*rbpA*<sub>M85A</sub> were transformed separately into BL21 (pLysS) competent cells and RbpA<sub>R80A</sub> and RbpA<sub>M85A</sub> proteins were purified by Q.F.F anion-exchange chromatography and Mono Q anion-exchange chromatography (Fig. 5.13A & B, lane 3) (Section 2.3.2.5 & 2.3.2.6). The non-tagged native RbpA was already obtained in Section 4.2.2.

The *in vitro* pull-down assays were performed by mixing the purified native RbpA<sub>R80A</sub> and RbpA<sub>M85A</sub> with purified His-tagged  $\sigma^{\text{HrdB}(2-4)}$  and  $\sigma^{\text{HrdB}(2)}$  (Fig. 5.13A & B, lanes 4 & 5) (Section 4.2.2) and left on ice for 15 min before passing them through magnetic Ni-affinity Dynabeads (Section 2.3.2.4). As a negative control, RbpA alone was mixed with the magnetic Dynabeads (Fig. 5.13A & B, lane 6). Positive controls were also included by passing native RbpA together with His-tagged  $\sigma^{\text{HrdB}(2)}$  and  $\sigma^{\text{HrdB}(2-4)}$  (Fig. 5.13A & B, lane 7 & 8). The Dynabeads were washed four times with buffer and any bound protein eluted using 100  $\mu$ l elution buffer.

As shown before, RbpA by itself did not bind to the Ni-affinity Dynabeads (Fig. 5.13A & B, lane 6). RbpA<sub>w/t</sub>, RbpA<sub>R80A</sub> and RbpA<sub>M85A</sub> each co-eluted from the beads together with  $\sigma^{\text{HrdB}(2)}$  and  $\sigma^{\text{HrdB}(2-4)}$ . These results confirm that RbpA<sub>R80A</sub> and RbpA<sub>M85A</sub> proteins interact with  $\sigma^{\text{HrdB}}$  at region 1.2-2.4 (Fig. 5.13A & B, lane 9 & 10).

**A****B**

**Figure 5.13: The interaction of RbpA<sub>R80A</sub> and RbpA<sub>M85A</sub> with  $\sigma^{\text{HrdB}(2)}$  and  $\sigma^{\text{HrdB}(2-4)}$ .**

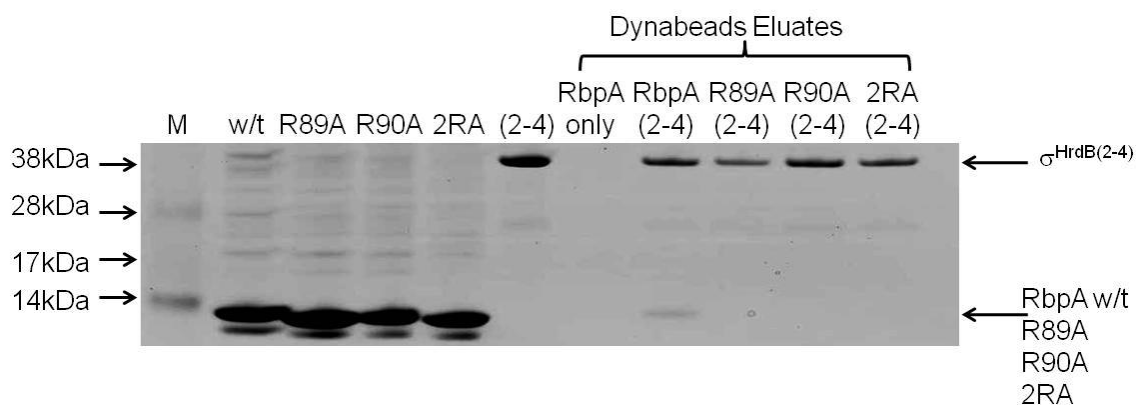
The pull down-assays were assessed by separating the proteins on SDS polyacrylamide gel. The Marker used was SeeBlue<sup>®</sup>Plus2 (Lane 1). **(A)** Lanes 2 to 5 show RbpA<sub>w/t</sub> (~14kDa), RbpA<sub>R80A</sub> (~14kDa),  $\sigma^{\text{HrdB}(2)}$  (~17kDa), and  $\sigma^{\text{HrdB}(2-4)}$  (~38kDa), respectively with their respective concentrations. Lanes 6 to 10 show the eluates from the dynabeads column. Lane 6: RbpA<sub>w/t</sub> only, Lane 7: RbpA<sub>w/t</sub> and  $\sigma^{\text{HrdB}(2)}$ , Lane 8: RbpA<sub>w/t</sub> and  $\sigma^{\text{HrdB}(2-4)}$ , Lane 9: RbpA<sub>R80A</sub> and  $\sigma^{\text{HrdB}(2)}$  and Lane 10: RbpA<sub>R80A</sub> and  $\sigma^{\text{HrdB}(2-4)}$ . **(B)** Lanes 2 to 5 show RbpA<sub>w/t</sub> (~14kDa), RbpA<sub>M85A</sub> (~14kDa),  $\sigma^{\text{HrdB}(2)}$  (~17kDa), and  $\sigma^{\text{HrdB}(2-4)}$  (~38kDa), respectively with their respective concentrations. Lanes 6 to 10 show the eluates from the dynabeads column. Lane 6: RbpA<sub>w/t</sub> only, Lane 7: RbpA<sub>w/t</sub> and  $\sigma^{\text{HrdB}(2)}$ , Lane 8: RbpA<sub>w/t</sub> and  $\sigma^{\text{HrdB}(2-4)}$ , Lane 9: RbpA<sub>M85A</sub> and  $\sigma^{\text{HrdB}(2)}$  and Lane 10: RbpA<sub>M85A</sub> and  $\sigma^{\text{HrdB}(2-4)}$ . The proteins were separated on a 4-12% Bis-Tris polyacrylamide gel at 120 V for 80 min, and visualised by Coomassie staining.

#### 5.4.2 RbpA<sub>R89A</sub> and RbpA<sub>R90A</sub>

Using BACTH, RbpA<sub>R90A</sub> and to some extent RbpA<sub>R89A</sub> proteins were shown to interact with  $\sigma^{\text{HrdB}}$  whereas RbpA<sub>2RA</sub> protein did not interact with  $\sigma^{\text{HrdB}}$  (Fig. 5.11). Therefore, to verify the results of RbpA mutant- $\sigma^{\text{HrdB}}$  interactions, *in vitro* pull-down assays were performed. NdeI-BamHI fragments of *rbpA*<sub>R89A</sub>, *rbpA*<sub>R90A</sub> and *rbpA*<sub>2RA</sub> from pMT3000::*rbpA* containing R89A, R90A and 2RA were subcloned into pET20b expression vector and RbpA mutants were purified as described in Section 5.4.1 (Fig. 5.14, lanes 3 to 5) (Section 2.3.2.5 & 2.3.2.6).

The *in vitro* pull-down assays were performed by separately mixing purified native RbpA<sub>R89A</sub>, RbpA<sub>R90A</sub> and RbpA<sub>2RA</sub> with His-tagged  $\sigma^{\text{HrdB}(2-4)}$ , left on ice for 15 min and then passed through Ni-affinity beads. The beads were washed four times with buffer and then any bound protein was eluted with 100  $\mu$ l elution buffer (Section 2.3.2.4). Controls were also prepared as mentioned in Section 5.4.1 (Fig. 5.14, lane 7 & 8).

RbpA<sub>w/t</sub> co-eluted from the beads together with  $\sigma^{\text{HrdB}(2-4)}$  (Fig. 5.14, lane 8), whereas, RbpA<sub>R89A</sub>, RbpA<sub>R90A</sub> and RbpA<sub>2RA</sub> were not sequestered by  $\sigma^{\text{HrdB}(2-4)}$  (Fig. 5.14, lane 9-11). This confirms that RbpA<sub>2RA</sub> protein does not interact with  $\sigma^{\text{HrdB}(2-4)}$ . Nevertheless, RbpA<sub>R89A</sub> and RbpA<sub>R90A</sub> proteins may possibly interact with  $\sigma^{\text{HrdB}}$  but with a lower affinity.



**Figure 5.14: The interaction of RbpA<sub>R89A</sub>, RbpA<sub>R90A</sub> and RbpA<sub>2RA</sub> with  $\sigma^{\text{HrdB}(2-4)}$ .**

The pull down-assays were assessed by separating the proteins on SDS polyacrylamide gel. The Marker used was SeeBlue<sup>®</sup> Plus2 (Lane 1). Lanes 2 to 6 show RbpA<sub>w/t</sub> (~14kDa), RbpA<sub>R89A</sub> (~14kDa), RbpA<sub>R90A</sub> (~14kDa), RbpA<sub>2RA</sub> (~14kDa), and  $\sigma^{\text{HrdB}(2-4)}$  (~38kDa), respectively, with their respective concentrations. Lanes 7 to 11 show the eluates from the dynabeads column. Lane 7: RbpA<sub>w/t</sub> only, Lane 8: RbpA<sub>w/t</sub> and  $\sigma^{\text{HrdBF4}}$ , Lane 9: RbpA<sub>R89A</sub> and  $\sigma^{\text{HrdB}(2-4)}$ , Lane 10: RbpA<sub>R90A</sub> and  $\sigma^{\text{HrdB}(2-4)}$ , Lane 11: RbpA<sub>2RA</sub> and  $\sigma^{\text{HrdB}(2-4)}$ . The proteins were separated on 4-12% Bis-Tris polyacrylamide gel at 120 V for 80 min and visualised by Coomassie staining.

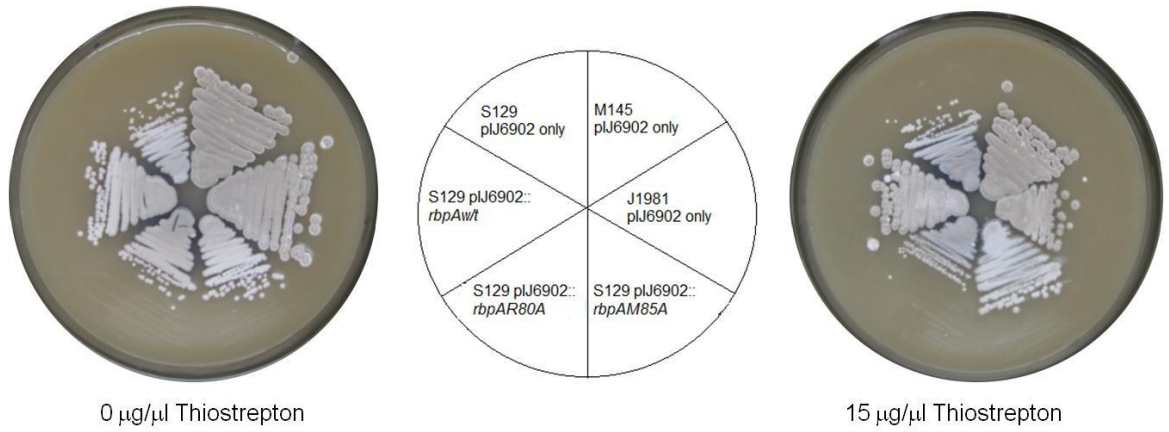
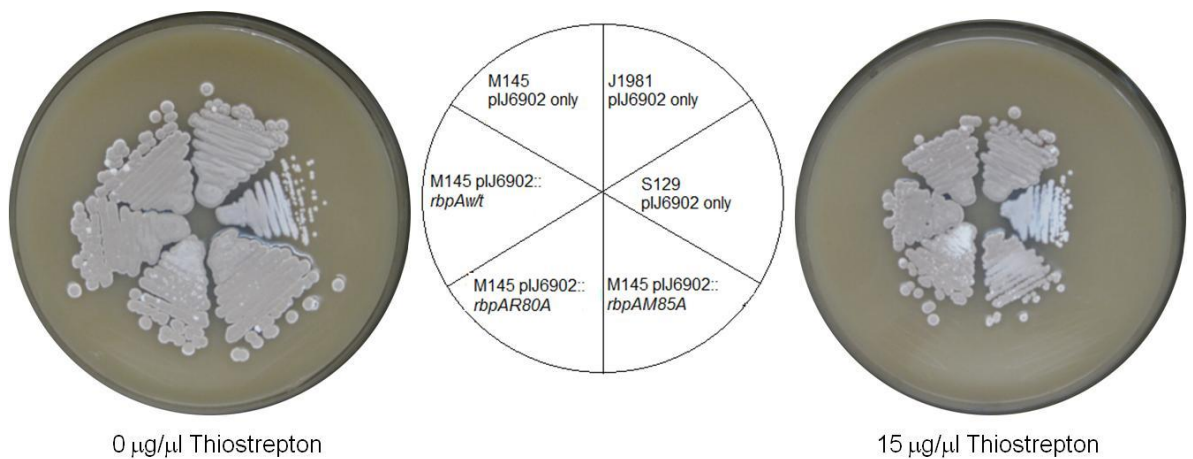
### **5.5 Testing RbpA<sub>R80A</sub> and RbpA<sub>M85A</sub> for dominant-negative effects**

As previously shown, S129 (pSET $\Omega$ ::*rbpA*<sub>R80A</sub>) and S129 (pSET $\Omega$ ::*rbpA*<sub>M85A</sub>) partially complemented the  $\Delta$ *rbpA* phenotype (Fig. 5.5). The mutant proteins were, however, shown to interact with  $\sigma^{\text{HrdB}}$  both *in vivo* (Section 5.3.1) and *in vitro* (Section 5.4.1). In theory, if the partially-active mutant proteins bind to  $\sigma^{\text{HrdB}}$  *in vivo*, they might compete with wild-type RbpA and thereby prevent normal transcription, leading to the inhibition of growth. Therefore, these RbpA mutants were expressed in a wild-type background to test for this dominant-negative effect.

NdeI-BamHI fragments of *rbpA*<sub>R80A</sub> and *rbpA*<sub>M85A</sub> obtained from pMT3000::*rbpA*<sub>R80A</sub> and pMT3000::*rbpA*<sub>M85A</sub>, respectively, were subcloned under the control of the *tipA* promoter in the pIJ6902 vector (Table 1- Section

2.1.2). Recombinant pIJ6902::*rbpA*<sub>R80A</sub> and pIJ6902::*rbpA*<sub>M85A</sub> were conjugated into S129 ( $\Delta$ *rbpA*), J1981 (w/t) and M145 (J1981 parental) *S. coelicolor* strains (Section 2.2.5.1). The ex-conjugants of S129, J1981 and M145 were streaked into single colonies and then the spores harvested (Section 2.2.5.2). The spores from each stock were then streaked onto MS agar plates containing 0  $\mu$ g/ $\mu$ l and 15  $\mu$ g/ $\mu$ l thiostrepton and left to grow 3-5 days at 30°C (Fig. 5.15). When thiostrepton was added (15  $\mu$ g/ $\mu$ l), the cloned *rbpA* mutant genes were induced from the *tipA* promoter.

In the presence of thiostrepton, pIJ6902::*rbpA*<sub>R80A</sub> did not complement the S129 mutant whilst pIJ6902::*rbpA*<sub>M85A</sub> partially complemented the  $\Delta$ *rbpA* phenotype (Fig. 5.15A). When J1981 and M145 containing pIJ6902::*rbpA*<sub>R80A</sub> were streaked in the presence of thiostrepton, both grew normally as with RbpA<sub>w/t</sub> (Fig. 5.15B & C) showing that RbpA<sub>R80A</sub> is not a dominant mutation. However, J1981 and M145 containing pIJ6902::*rbpA*<sub>M85A</sub> produced partially reduced colony sizes in the presence of thiostrepton (Fig. 5.15B & C) showing that RbpA<sub>M85A</sub> has a dominant effect.

**A****B****C**

**Figure 5.15: Introducing pIJ6902::*rbpA<sub>w/t</sub>*, pIJ6902::*rbpA<sub>R80A</sub>*, and pIJ6902::*rbpA<sub>M85A</sub>* into *S. coelicolor* S129, J1981 and M145 strains.**

The *S. coelicolor* spores were plated on MS agar containing 0 µg/µl and 15 µg/µl thiostrepton and incubated at 30°C for 3-5 days. **(A)** Showing S129

(pIJ6902::control), S129 (pIJ6902::*rbpA*<sub>R80A</sub>), S129 (pIJ6902::*rbpA*<sub>M85A</sub>), S129 (pIJ6902::*rbpA*<sub>w/t</sub>), and M145 and J1981 vector only controls. **(B)** Showing J1981 (pIJ6902::control), J1981 (pIJ6902::*rbpA*<sub>R80A</sub>), J1981 (pIJ6902::*rbpA*<sub>M85A</sub>), J1981 (pIJ6902::*rbpA*<sub>w/t</sub>), and M145 and S129 vector only controls. **(C)** Showing M145 (pIJ6902::control), M145 (pIJ6902::*rbpA*<sub>R80A</sub>), M145 (pIJ6902::*rbpA*<sub>M85A</sub>), M145 (pIJ6902::*rbpA*<sub>w/t</sub>), and S129 and J1981 vector only controls.

## **5.6 Transcription activation by RbpA<sub>M85A</sub> and RbpA<sub>2RA</sub> proteins**

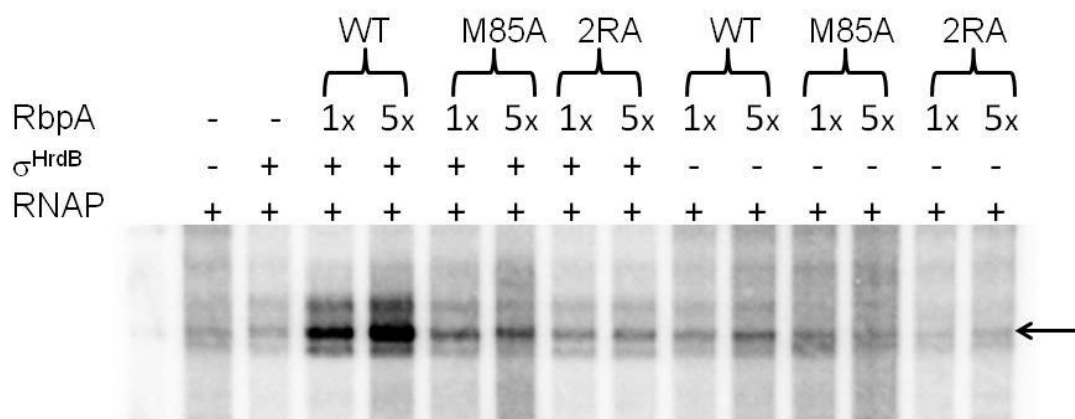
RbpA<sub>M85A</sub> protein binds to  $\sigma^{\text{HrdB}}$  (Section 5.3.1 & 5.4.1) and partially restores growth of *S. coelicolor* (Section 5.2.2.1), whereas RbpA<sub>2RA</sub> protein did not interact with  $\sigma^{\text{HrdB}}$  (Section 5.3.2) and did not complement the  $\Delta\textit{rbpA}$  strain (Section 5.2.3.1). Therefore, it was decided to find out whether these two distinct RbpA mutants had any effect on the activation of transcription from an RbpA-dependent promoter. *In vitro* transcription assays were performed using RbpA<sub>M85A</sub> and RbpA<sub>2RA</sub> proteins and were compared to RbpA<sub>w/t</sub> proteins, which served as a positive control.

Three genes shown to be regulated by RbpA from the microarray data were selected; SCO4652, SCO0527 and *rpoB* (Table A2- Appendices). These were analysed by *in vitro* transcription assays to determine whether they were activated by RbpA (In collaboration with M. Paget, data not shown). SCO4652 was then chosen as the desired template for further studies. The reaction mixture was set up containing 10 x alternative transcription buffer (Section 2.1.6.4), 10 x NTP mix (Section 2.1.6.4), 50 nM SCO4652 template, 500 nM *S. coelicolor* core RNAP (a gift from P. Doughty), 2.5  $\mu\text{M}$   $\sigma^{\text{HrdB}}$  (a gift from P. Doughty), 5 and 25  $\mu\text{M}$  RbpA. The protocol was followed as detailed in Section 2.5.2.

Transcription was not stimulated in the presence of core RNAP only (Fig. 5.16, lane 1). Addition of  $\sigma^{\text{HrdB}}$  to core RNAP was not sufficient to activate transcription from the RbpA-dependent promoter (Fig. 5.16, lane 2). When RbpA<sub>w/t</sub> was added, transcription was activated (Fig. 5.16, lane 3 & 4). However, when RbpA<sub>M85A</sub> was added, transcription was only partially activated



compared to RbpA<sub>w/t</sub> (Fig. 5.16, lane 5 & 6). RbpA<sub>2RA</sub> did not activate transcription initiation (Fig. 5.16, lane 7 & 8). As expected, RbpA did not stimulate transcription in the absence of added  $\sigma^{\text{HrdB}}$ , (Fig. 5.16, lanes 9-14) further verifying that RbpA stimulates transcription in a  $\sigma^{\text{HrdB}}$ -dependent manner. The slight increase in transcription caused by the addition of RbpA<sub>w/t</sub> is probably the result of contaminating  $\sigma^{\text{HrdB}}$  in the core RNAP preparation.



**Figure 5.16: *In vitro* transcription assays showing transcription activation in the presence of RbpA<sub>w/t</sub>, RbpA<sub>M85A</sub>, and RbpA<sub>2RA</sub>.**

The samples were loaded on the urea gel and separated at 600 V, 50 W at 55°C for 1 h. Lane 1: RNAP core only, Lane 2: RNAP and  $\sigma^{\text{HrdB}}$ , Lane 3: RNAP,  $\sigma^{\text{HrdB}}$  and RbpA<sub>w/t</sub> (1x), Lane 4: RNAP,  $\sigma^{\text{HrdB}}$  and RbpA<sub>w/t</sub> (5x), Lane 5: RNAP,  $\sigma^{\text{HrdB}}$  and RbpA<sub>M85A</sub> (1x), Lane 6: RNAP,  $\sigma^{\text{HrdB}}$  and RbpA<sub>M85A</sub> (5x), Lane 7: RNAP,  $\sigma^{\text{HrdB}}$  and RbpA<sub>2RA</sub> (1x), Lane 8: RNAP,  $\sigma^{\text{HrdB}}$  and RbpA<sub>2RA</sub> (5x), Lane 9: RNAP and RbpA<sub>w/t</sub> (1x), Lane 10: RNAP and RbpA<sub>w/t</sub> (5x), Lane 11: RNAP and RbpA<sub>M85A</sub> (1x), Lane 12: RNAP and RbpA<sub>M85A</sub> (5x), Lane 13: RNAP and RbpA<sub>2RA</sub> (1x), Lane 14: RNAP and RbpA<sub>2RA</sub> (5x).

## **5.7 Discussion**

### **5.7.1 The C-terminal region of RbpA**

The C-terminal region of RbpA was shown to interact with  $\sigma^{\text{HrdB}}$  (Section 5.1.1) and thus, it contains the  $\sigma^{\text{HrdB}}$ -binding region. It is possible that other regions of RbpA might be important *in vivo*, however, based on the assays in the chapter, the C-terminal region appears to be the region involved in  $\sigma^{\text{HrdB}}$  interaction. To identify possible residues that may be involved in this protein-protein interaction, specific residues were selected and mutated. The C-terminal region of RbpA is composed of Helix 1 and Helix 2, which are connected by the ERR motif. Helix 1 is present in all the RbpA and RbpB homologues with a highly conserved sequence. However, this sequence conservation differs in the RbpA and RbpB homologues (Fig. 3.1 & 3.2). In addition, the ERR motif is completely conserved amongst all the RbpA and RbpB homologues (Fig. 3.1 & 3.2). Helix 2 is also present in the RbpA and RbpB homologues, with few leucine, arginine and glutamate residues being well conserved amongst the RbpA and RbpB homologues. Helix 2 is amphipathic in RbpA homologues but this is not the case in the RbpB homologues. Therefore, all the residues within Helix 1 and the ERR motif were selected for mutagenesis, whereas, in Helix 2, the conserved glutamate residues and arginine residues were selected for mutagenesis.

### **5.7.2 R80 and M85 are important in RbpA functionality**

Site-directed mutagenesis studies of RbpA Helix 1 identified two key residues, R80 and M85, which seem to be indispensable for the full role of RbpA. When alanine was substituted for R80 and M85 and introduced into *S. coelicolor*  $\Delta rbpA$  strain, they partially complemented the  $\Delta rbpA$  phenotype (Section 5.2.2.1). The effect of R80A and M85A was further confirmed by determining the growth curve of these mutants, which were shown to rescue the growth relative to the mutant but still less than that of the wild-type (Section 5.2.2.2). The mutants did not fully restore the full functionality of RbpA showing that both these residues are necessary for its correct functioning.

Despite the inability of the R80A and M85A mutants to complement the  $\Delta rbpA$  phenotype, each protein interacted with  $\sigma^{\text{HrdB}}$  (Fig. 5.10 & Fig. 5.13). Alignment of RbpA and Crl showed that R80 and M85 were not conserved in these proteins (Fig. A3- Appendices), suggesting that the possible interaction with domain 2 of  $\sigma^{\text{HrdB}}$  might be by other residues. Therefore, how do these mutants exert their effect? They might bind to  $\sigma^{\text{HrdB}}$  and sequester it, preventing its binding to core RNAP. Alternatively, they might trap RNAP in an inactive complex, perhaps blocking closed complex formation or open complex formation. Further analysis is required to investigate these possibilities.

It was considered that a mutant RbpA that could bind the essential sigma factor,  $\sigma^{\text{HrdB}}$ , but not activate transcription, might be a dominant-negative mutant.  $rbpA_{\text{R80A}}$  did not cause such an effect. The reason why R80A did not cause this effect is not known. It is possible that it cannot compete with  $RbpA_{\text{w/t}}$  due to a lower affinity or that the protein is unstable *in vivo* and gets degraded. However,  $rbpA_{\text{M85A}}$  was a dominant mutation, which when induced in wild-type *S. coelicolor* strains, inhibited growth (Fig. 5.15B & C). Therefore, the M85A mutant protein might compete with  $RbpA_{\text{w/t}}$  for interaction with  $\sigma^{\text{HrdB}}$  thereby affecting transcription.

*In vitro* transcription assays performed on  $RbpA_{\text{M85A}}$  revealed that the mutant protein partially activates transcription initiation (Fig. 5.16) showing that the mutant protein can interact productively with  $\sigma^{\text{HrdB}}$  and stimulate transcription to some extent.

### 5.7.3 The ERR motif of RbpA is essential for $\sigma^{\text{HrdB}}$ interaction

Generating mutations in the ERR motif of RbpA revealed that R89 and R90 are important residues for maintaining the functional role of RbpA. When alanine was substituted for both R89 and R90 (2RA) and introduced into *S. coelicolor*  $\Delta rbpA$  strain, it did not complement the phenotype of the deletant (Section 5.2.3.1). However, when the alanine substitutions for R89 or R90 were generated individually and introduced into *S. coelicolor*  $\Delta rbpA$  strain, it

appeared that the presence of either R89 or R90 was sufficient to restore RbpA functionality to some extent *in vivo* (Section 5.2.3.1).

The ERR motif is very highly conserved amongst all the RbpA and RbpB homologues in the Actinobacteria (Fig. 3.1 & 3.2). Hence, it was not surprising when R89 and R90 (2RA) were shown to be necessary for  $\sigma^{\text{HrdB}}$  interaction (Section 5.3.2 & 5.4.2). This is consistent with the *in vitro* transcription assays which showed that RbpA<sub>2RA</sub> protein did not activate transcription (Fig. 5.16). On the other hand, RbpA<sub>R89A</sub> and RbpA<sub>R90A</sub> were shown to interact with  $\sigma^{\text{HrdB}}$  using BACTH assays (Section 5.3.2). However, it was observed that the interaction between these mutant proteins and  $\sigma^{\text{HrdB}}$  was weak as the  $\beta$ -galactosidase activity fold decreased from 57 for RbpA<sub>w/t</sub> to about 26 and 16 for RbpA<sub>R90A</sub> and RbpA<sub>R89A</sub>, respectively. The *in vitro* pull down assays showed that RbpA<sub>R89A</sub> and RbpA<sub>R90A</sub> did not to interact with  $\sigma^{\text{HrdB}}$  (Section 5.4.2), which is possibly due to a weak interaction between the two proteins as observed in BACTH (see above). Thus, R89 and R90 may also form specific interactions with residues in  $\sigma^{\text{HrdB}}$ .

Alternatively, R89 and R90 may be structurally important in maintaining the C-terminal region of RbpA. Nevertheless, when the two arginines were replaced with two alanines and the secondary structure analysed by PSIPRED, it was predicted that the secondary structure of the C-terminal region of RbpA would not be disrupted (data not shown). This can only be verified when the structure of RbpA- $\sigma^{\text{HrdB}(2)}$  is solved (Section 6.5).

#### 5.7.4 Helix 2 of RbpA is involved in interaction with $\sigma^{\text{HrdB}}$

Mapping the region where RbpA binds to  $\sigma^{\text{HrdB}}$  revealed that RbpA construct containing the C-terminal region only (RbpA <sub>$\Delta\beta$</sub> ) interacted with  $\sigma^{\text{HrdB}}$ , whereas RbpA construct lacking Helix 2 but containing the ERR motif (RbpA <sub>$\Delta\text{H2}$</sub> ) did not bind to the sigma factor (Section 5.1.1). This provides evidence that Helix 2 is required for the interaction with  $\sigma^{\text{HrdB}}$ .

However, the mutation of the conserved glutamate and arginine residues located in Helix 2, did not have any effect on the *S. coelicolor* phenotype (Section 5.2.4) and the mutant proteins were shown to interact with  $\sigma^{\text{HrdB}}$  (Section 5.3.3). This demonstrates that these residues do not individually play a critical role in mediating RbpA function. However, they may operate collectively with the conserved hydrophobic residues (L & V) in Helix 2 to mediate RbpA- $\sigma^{\text{HrdB}}$  interactions and possibly RbpA-RNAP interactions. It is also worth noting that introducing a charge reversal, from glutamate (-ve) to arginine or lysine (+), or a hydrophobic side group (L & V), may produce a significant effect on RbpA. However, these hydrophobic residues have not been analysed yet.

Even though the C-terminal region of RbpA was shown to be important for RbpA functionality, the N-terminal region of RbpA might also be important *in vivo*. The N-terminal region appears to be involved in maintaining the structural integrity of RbpA (Chapter 6).

## **6- Structural analysis of Rv2050/RbpA**

## **6.0 Overview**

The structure of RbpA has not yet been solved and there is no available structure for any RbpA homologue. It was hoped that a structure of RbpA or one of its orthologues (e.g. Rv2050) would help us to better understand the biological function of these proteins and so X-ray crystallographic and nuclear magnetic resonance (NMR) approaches were used to try to solve RbpA/Rv2050 structure. The secondary structure prediction of RbpA suggested that RbpA might have four  $\beta$ -strands interconnected by loops in the N-terminal region followed by two  $\alpha$ -helices in the C-terminal region (Section 3.1.1). This section focuses on determining whether this prediction is true. Also, as RbpA/Rv2050 were shown to interact with  $\sigma^{\text{HrdB}}/\sigma^{\text{A}}$  at domain 2 (Section 4.2 & 4.3.2), approaches were used to purify the RbpA/Rv2050-Sigma complex for structural studies. This complex would reveal important residues in the interaction of RbpA with its sigma factor.

### **6.1 X-ray crystallography of Rv2050**

Initially, X-ray crystallography was used to try to solve the structure of Rv2050, the orthologue of RbpA in *M. tuberculosis*. Native Rv2050 (Section 4.3.2.2) was over-expressed and purified by anion-exchange purification (data not shown) followed by gel filtration (data not shown) to obtain a final concentration of 8.62 mg/ml. This was then used for setting up crystal trials.

A pre-crystallization test (PCT) was performed to determine the appropriate concentration of Rv2050 to use for the crystal trials (Qiagen: EasyXtal® and NeXtal® Protein Crystallization Handbook) using the Pre-Screen Assay Suite (Qiagen: EasyXtal and Nextal products). These solutions were placed into 24-well plate and the “hanging drop vapour diffusion” method (Section 2.3.7) was used with two different concentrations of Rv2050 (3 mg/ml and 6 mg/ml). It was left at 4°C and after a few days, the Rv2050 6 mg/ml was chosen to be the appropriate concentration for further trials as it produced very small crystals at some buffer conditions.

Crystallization trials were then performed using the “Structure Screen 2 MD1-02” from Molecular Dimensions and “PACT premier™ box 1 of 2 and box 2 of 2 MD1-29” from Molecular dimensions. The “hanging drop vapour diffusion” was performed and after few weeks, very small rod-like crystals were seen in few wells. However, when X-ray beams were passed through them, they did not diffract well. The crystals were either poor or small crystals of Rv2050 or they were salt precipitates.

At this point it was decided to temporarily abandon the X-ray crystallisation approach in favour of a nuclear magnetic resonance-based approach. Although, the crystallisation approach had not been exhaustive, the small size of RbpA/Rv2050 is suitable for NMR-based studies.

## **6.2 NMR spectroscopy of Rv2050**

Unlike X-ray crystallization, NMR is performed in solution. It relies on the nuclei possessing a small electromagnetic charge called spin. Any element that has a nuclear spin can give an NMR signal.  $^{13}\text{C}$  and  $^{15}\text{N}$  are isotopes used for protein labelling since they have a net nuclear spin of  $1/2$ . This makes them suitable for NMR spectroscopy studies. Pulses of electromagnetic energy are sent to the sample, which generates multiple resonance frequencies that are used to assign chemical shifts. The structure calculations are determined using the chemical shifts and the distance restraints. Labelled  $^{13}\text{C}$   $^{15}\text{N}$  Rv2050 was over-produced in minimal medium, purified by anion exchange-chromatography followed by gel filtration and sent to the laboratory of Professor Stephen Matthews in Imperial College London for NMR analysis.

### **6.2.1 Native $^{13}\text{C}$ & $^{15}\text{N}$ Rv2050 over-expression**

*E. coli* BL21 (pLysS) (pET20b::Rv2050) was grown in 200 ml LB media to an  $\text{OD}_{600}$  of 3-5. The cells were harvested by centrifugation, and then re-suspended in optimized high-cell density IPTG-induction minimal medium (Section 2.1.4.1). This minimal medium included  $^{13}\text{C}$ -glucose as carbon source and  $^{15}\text{N}$ - $\text{NH}_4\text{Cl}$  as nitrogen source, thereby labelling Rv2050. The culture was equilibrated in minimal media for 1.5-2 h, then induced with IPTG and the cells harvested (Section 2.3.1.2) (Sivashanmugam *et al.*, 2009).



## 6.2.2 Purification of $^{13}\text{C}$ $^{15}\text{N}$ Rv2050

### 6.2.2.1 Anion-exchange - QFF column

The first anion-exchange step was performed using a low resolution Hi-Trap QFF column. Following the preparation of the cleared cell lysate (Section 2.3.2.1) using the anion-exchange binding buffer (Section 2.1.6.2), Rv2050 was loaded onto the column then eluted using a 10 ml gradient of 0 to 100% elution buffer (Section 2.3.2.5). Rv2050 eluted at ~ 40% elution buffer *i.e* 0.4 M NaCl (4-9 ml; Fig. 6.1A).

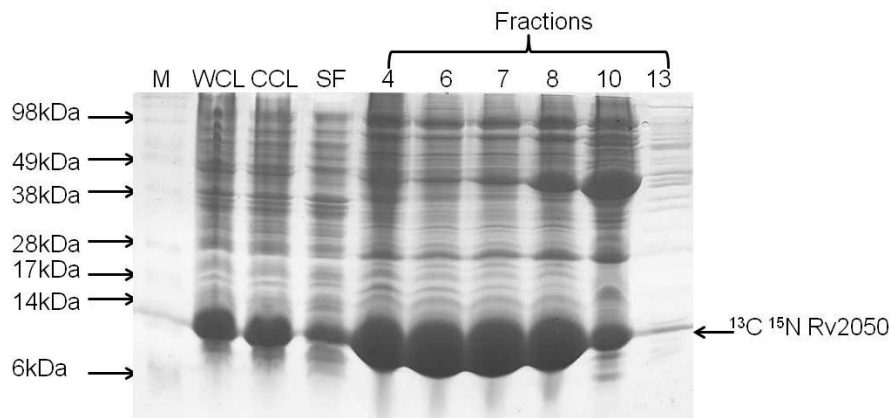
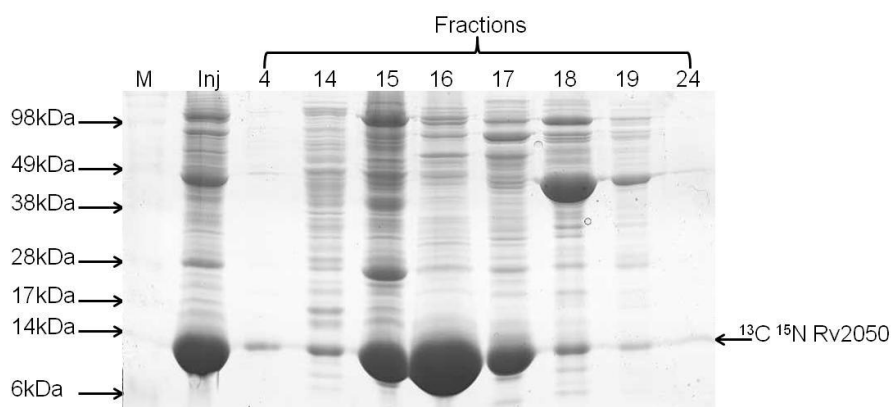
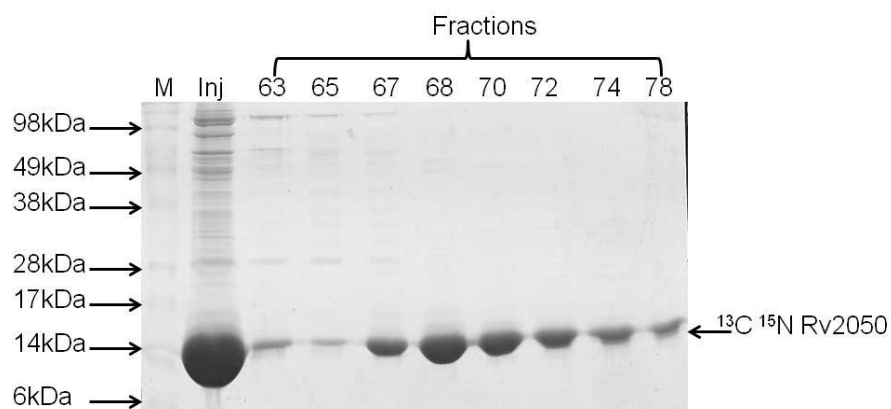
### 6.2.2.2 Anion-exchange - Mono Q column

Fractions 4 to 8 (Fig. 6.1A) were dialysed into anion-exchange binding buffer then further purified by anion-exchange using a Mono Q column 5/50 GL column. The gradient was set at 0 to 80% for 20 ml followed by 80 to 100% for 5 ml (Section 2.3.2.5). Rv2050 eluted between fractions 15 to 17 (~ 0.4-0.5 M NaCl). However, fraction 16 was the most pure fraction (Fig. 6.1B).

### 6.2.2.3 Gel filtration

Fraction 16 (Fig. 6.1B) was dialysed into gel filtration buffer (using 20 mM Tris pH 7.9, 50 mM NaCl, 5% (v/v) glycerol, 5 mM  $\beta$ -mercaptoethanol) at 4°C overnight, filtered, and passed through a pre-equilibrated gel filtration column (HiLoad<sup>TM</sup> 16/60 Superdex<sup>TM</sup> 200 column) (Section 2.3.2.6). Eluates were collected as 1 ml fractions and Rv2050 eluted at around 65 to 72 ml. When these fractions were analysed by SDS polyacrylamide gel, the fractions were considered sufficiently pure for NMR analysis (Fig. 6.1C).

Fractions 67 to 78 ml (Fig. 6.1C) were concentrated to 42 mg/ml in a volume of 500  $\mu$ l and then dialysed into a standard NMR buffer (Section 2.1.6.2) overnight, snap frozen in liquid nitrogen after the addition of 0.02% sodium azide, and finally sent for NMR spectroscopy.

**A****B****C**

**Figure 6.1: The purification of  $^{13}\text{C } ^{15}\text{N}$  Rv2050.**

The samples were assessed by separating the proteins on 4-15% Bis-Tris polyacrylamide gel at 200 V for 70 min and visualized by Coomassie Blue staining. The Marker used was SeeBlue®Plus2. **(A)** QFF anion-exchange chromatography. 10  $\mu\text{l}$  of the whole cell lysate (WCL), cleared cell lysate (CCL), supernatant flow-through (SF) and 15  $\mu\text{l}$  of different 1 ml fractions were loaded on the gel. **(B)** Mono Q anion-exchange purification. 5  $\mu\text{l}$  of injected Rv2050 (Inj) together with 10  $\mu\text{l}$  of different 1 ml fractions loaded on gel. **(C)** Gel filtration purification of  $^{13}\text{C } ^{15}\text{N}$  Rv2050. The same volumes of samples loaded as in **B**.

### 6.2.3 NMR structure of Rv2050

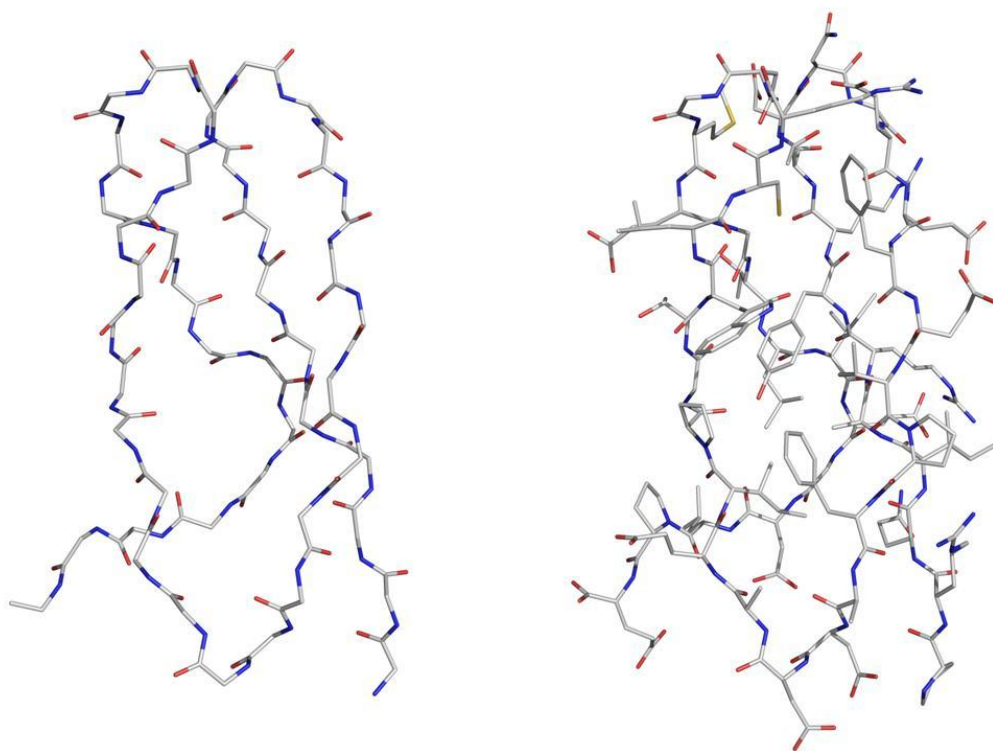
NMR was performed in collaboration with Professor Stephen Matthews (Imperial College London). The analysis of the backbone and allocation of the side chains was performed by Marie-Laure Parsy and Dr. Peter Simpson, and are briefly described here. Initially, unlabelled Rv2050 was assessed by one-dimensional (1D)  $^1\text{H}$  magnetic spectrum, which revealed that it was structured and could be analyzed for NMR. The 2-dimensional  $^{15}\text{N}$ - $^1\text{H}$  Heteronuclear Single Quantum Coherence Spectrum of labelled Rv2050 showed the H-N correlations of the protein and produced dispersed peaks that would allow for assignment of  $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\beta$ , and  $^{13}\text{C}'$  resonance. The backbone resonance, chemical shift Index Script and side chain resonance were then obtained. Rv2050 structure was then calculated and validated (M.L. Parsy, P. Simpson, & S. Matthews- unpublished observations). Rv2050 was not aggregated and was free of contaminants but appeared to have only 40 to 60 structured residues (Fig. 6.2: Bold with light grey shading).

It was noticed that the C-terminus of Rv2050 was very dynamic and its structure could not be solved. However, the N-terminal region of Rv2050 was solved and was analysed and visualized using PyMOL software. The N-terminal region of Rv2050 backbone structure, charge surface, and cartoon illustrations were generated using PyMOL (Fig. 6.3 & 6.4). The N-terminal region of Rv2050 has four  $\beta$ -strands that run anti-parallel and are interconnected by loops (Fig. 6.4- Right). The  $\beta$ -strands of Rv2050 are located at the exact regions as denoted by the predicted secondary structure software - PSIPRED (Section 3.1.1).

MADRVLRGSR LGAVSYETDRNHDLA**PRQIARYRTDNGEEFEVPFADDAEIPG**  
**TWLCRNGMEGT**LI**EGDLPE**PKKVKKPPRTHWDMLLERRSIEELEELLKERLELI  
 RSRRRG

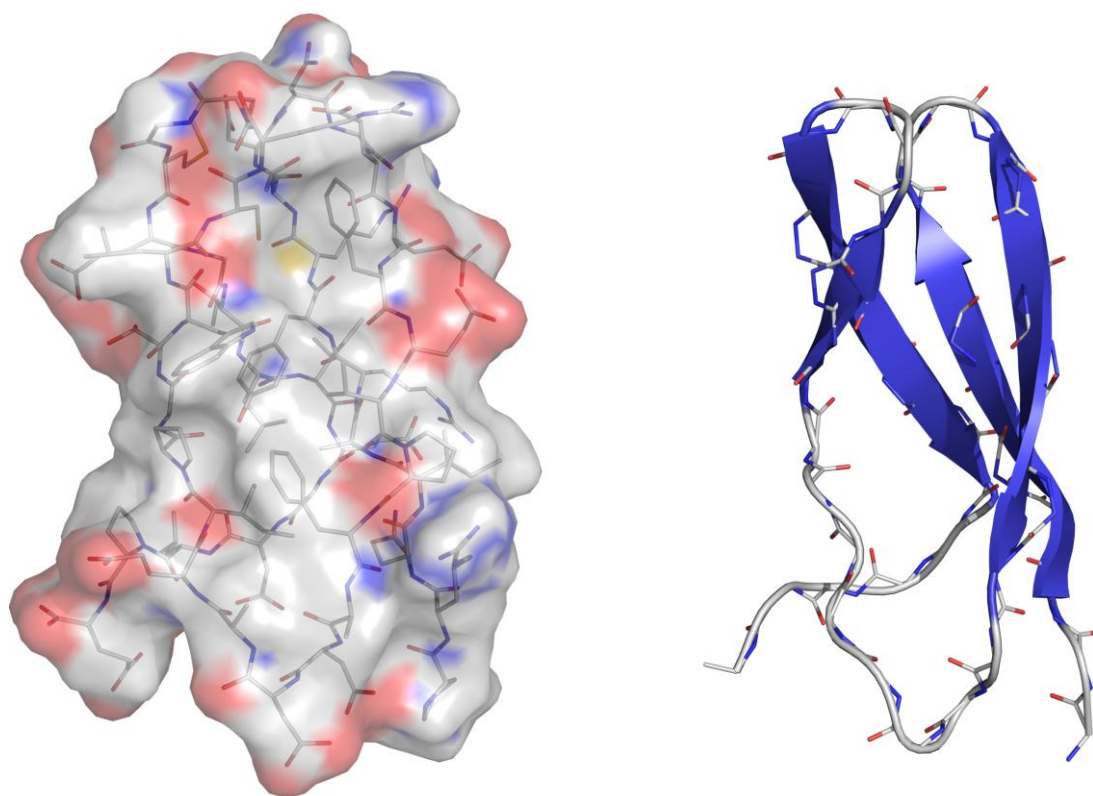
**Figure 6.2: The amino acid sequence of Rv2050 indicating its structured region.**

The amino acid residues with bold grey shading represent the structured region of Rv2050.



**Figure 6.3: Backbone structure of N-terminus of Rv2050.**

**Left** - Using PyMOL software to show the carbon ( $C\alpha$ ), Nitrogen (N) and Oxygen (O) backbone of the N-terminal region of Rv2050 excluding the hydrogens and side chains. **Right** - including the side chains (right). Residues of Rv2050 included are: PRQIARYRTDNGEEFEVVPFADDAEIPGTWLCR NGMEGTLIEGDLPE.



**Figure 6.4: The surface charge and the cartoon representation of the N-terminus of Rv2050.**

**Left** - Surface charge: The red regions denotes the negative charge, the blue regions represent the positive charge, whilst the white regions denote the neutral charge amino acids. **Right** - Cartoon representation: The blue arrows represent the  $\beta$ -stands whilst the grey lines represent the loops connecting the  $\beta$ -strands. PyMOL software was used to generate these images. Residues of Rv2050 included are: PRQIARYRTDNGEEFEVFPFADDAEIPGTWLCRNGMEGTLIEGDLPE.

#### 6.2.4 Homology search of Rv2050

Using the pdb file of the N-terminal region of Rv2050, a DALI search was performed. Most of the proteins in the search belonged to ribosomal proteins: 50S Ribosomal proteins (Table 13). The top hit was a Peptide Chain Release Factor subunit 1 (RF1) (Section 6.6.1).

**Table 13: DALI Search hits using the N-terminal region of Rv2050.**

<b>Protein</b>	<b>Z Score</b>
Peptide Chain Release factor Subunit 1: PDB entry 3IR9	4.2
50S Ribosomal Protein L27: PDB entry 2Y0Z	4.1
50S Ribosomal Protein L1: PDB entry 3KIR	3.9
50S Ribosomal protein L2: PDB entry 3I1N	3.4
Desulfoferrodoxin: PDB entry 1Y07	2.8
Replicase polyprotein 1AB: PDB entry 2FE8	2.8
Aminomethyltransferase: PDB entry 3A8I	2.8
Prokaryotic transcription elongation factor GreA/GreB: PDB entry 2PN0	2.2

The Z-score is a measure of the accuracy of the prediction, values < 2 are false.

### **6.3 Rv2050 is not a Zn metalloprotein**

As mentioned in Section 3.1.1, RbpA homologues exist in the Zn and non-Zn form. It was shown that *S. coelicolor* RbpA is a Zn-binding protein with its Zn coordinated by three cysteine residues (C34, C56 & C59) and probably one histidine residue (H38) (Fig. 3.1) (P. Doughty & M. Paget, personal communication). However, Rv2050 only has a single cysteine and two histidine residues, neither of which is equivalent to RbpA H38 (Fig. 3.1), which suggests that this protein does not coordinate a metal. Therefore, inductively coupled plasma mass spectrometry (ICP-MS) was used to analyse the presence of trace metals in Rv2050.

#### **6.3.1 ICP-MS analysis of Rv2050**

A pure fraction of 0.5 mg Rv2050 purified by anion-exchange chromatography was sent for ICP-MS (University of Sussex). For a 1:1 stoichiometry of Rv2050 complexed with Zn, Mg, or Mn, the concentrations would be 946 ppb, 352 ppb, and 795 ppb, respectively. However, the sample contained Zn, Mg and Mn at 34.25 ppb, 5.814 ppb and 0.03938 ppb, respectively (Table 14), indicating that Rv2050 is not a metalloprotein.

**Table 14: ICP-MS data for Zn, Mg and Mn trace metals of Rv2050.**

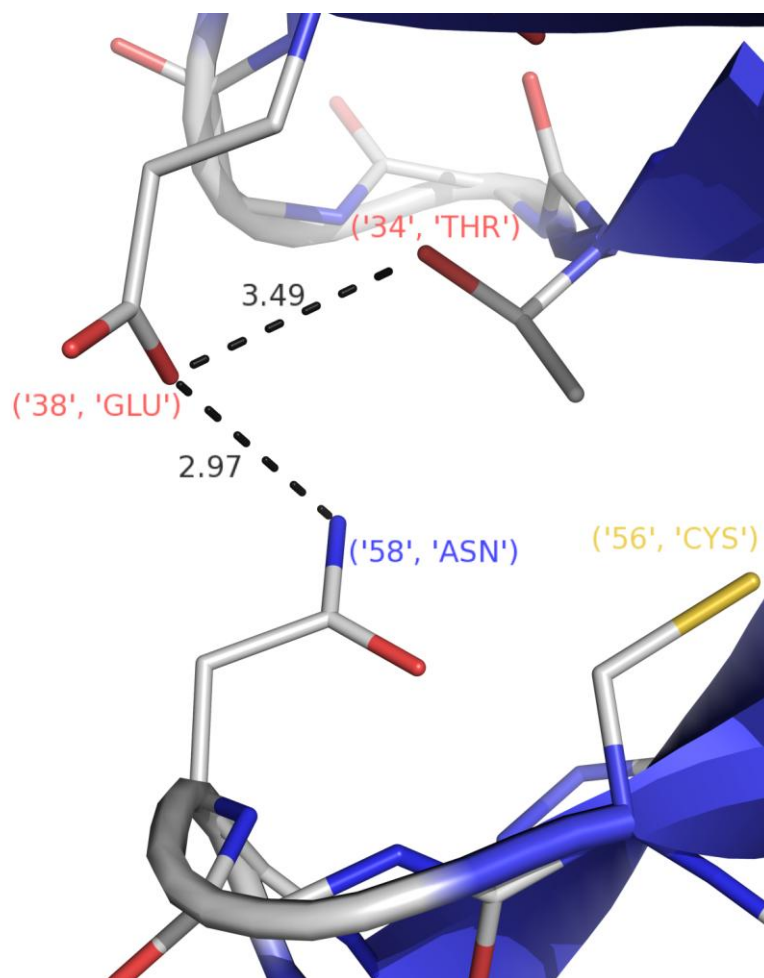
File	001_ STD.D		002_ STD.D		003 SMPL.D		004 SMPL.D		005 SMPL.D	
Sample	0 ppb		1000 Ppb		rinse blank		Buffer		Rv2050 sample	
Dilution	1		1		1		1		1	
	conc. (ppb)	% RSD	conc. (ppb)	% RSD	conc. (ppb)	% RSD	conc. (ppb)	% RSD	conc. (ppb)	% RSD
<b>Zn / 66</b>	-6.7E-08	>100	1000	0.64	7.383	4.11	<b>3.168</b>	8.84	<b>34.25</b>	<b>0.6</b>
<b>Mg / 24</b>	1.8E-07	>100	1000	0.94	0.2267	78.12	<b>2.332</b>	<b>1.27</b>	<b>5.814</b>	<b>0.66</b>
<b>Mn / 55</b>	1.7E-09	>100	1000	0.5	0.4237	34.25	<b>-0.1494</b>	<b>3.15</b>	<b>0.03938</b>	<b>22.06</b>

“Buffer” represents concentration of metals present in the solution Rv2050 was dialysed in.

“RV2050 Sample” represents the concentration of the different metals present in 0.5 mg Rv2050.

### 6.3.2 Key residues that maintain the structural integrity of Rv2050

The alignment in Fig. 3.1 revealed that the predicted Zn-ligands in RbpA (C34, H38, C56 and C59) are “replaced” by the following conserved residues in Rv2050: T34, E38, C56 and N58. Strikingly, these residues appear to play an important role in maintaining the structural fold. For example, there appears to be a hydrogen/salt bridge formed between the glutamate (E38) and asparagine (N58). Furthermore, the PyMOL program predicted a hydrogen bond between E38 and T34 (Fig. 6.5). It follows that the proximity of T34, E38, C56 and N58 side chains in Rv2050 suggests that the equivalent residues in RbpA ligate Zn, as predicted.



**Figure 6.5: Residues in Rv2050 that are thought to be interacting instead of coordinating Zn.**

This illustration was generated using PyMOL software. Broken lines represent hydrogen bonds and/or salt bridges forming between GLU 38 and THR 34, and GLU 38 and ASN 58 together with the distance of the bonds (black).

#### **6.4 Attempts to solve the structure of RbpA using NMR**

The overall structure of Rv2050 could not be solved due to the flexible/degraded C-terminal region. Therefore, it was decided to repeat the experiments using *S. coelicolor* RbpA, which is likely to have a virtually identical fold. Initially, native RbpA labelled with  $^{15}\text{N}$  was over-produced, purified by anion-exchange chromatography followed by gel filtration, and sent for NMR analysis to determine if the protein was structured. RbpA produced dispersed peaks showing that it was structured, possibly more so than Rv2050, and therefore RbpA was produced labelled with  $^{13}\text{C}$  and  $^{15}\text{N}$ .



Following purification using anion-exchange chromatography and gel filtration,  $^{13}\text{C}$   $^{15}\text{N}$  RbpA was dialysed into phosphate buffer and sent for NMR spectroscopy.

#### 6.4.1 Over-expression of $^{13}\text{C}$ $^{15}\text{N}$ RbpA

*E. coli* BL21 (pLysS) (pET20b::*rbpA*) was grown in 500 ml LB media to an  $\text{OD}_{600}$  of 3-5. The culture was harvested by centrifugation and the pellet re-suspended in optimized high-cell density IPTG-induction minimal media containing  $^{13}\text{C}$ -glucose and  $^{15}\text{N}$ - $\text{NH}_4\text{Cl}$  (Section 2.1.4.1). The cultures were cold shocked, 0.03 mM  $\text{ZnSO}_4$  added as RbpA is a Zn metalloprotein, and RbpA expression was induced with IPTG. The cultures were induced at 30°C with vigorous shaking for 3 h (Section 2.3.1.2).

#### 6.4.2 Purification of $^{13}\text{C}$ $^{15}\text{N}$ RbpA

##### 6.4.2.1 Anion-exchange - QFF column

Anion-exchange chromatography was performed as described in Section 6.2.2.1, with some minor modifications. RbpA was also eluted at ~0.4 M NaCl (21-27 ml). A lower band appeared beneath the RbpA band suggesting that RbpA was degrading during the steps of purification (Fig. 6.6A).

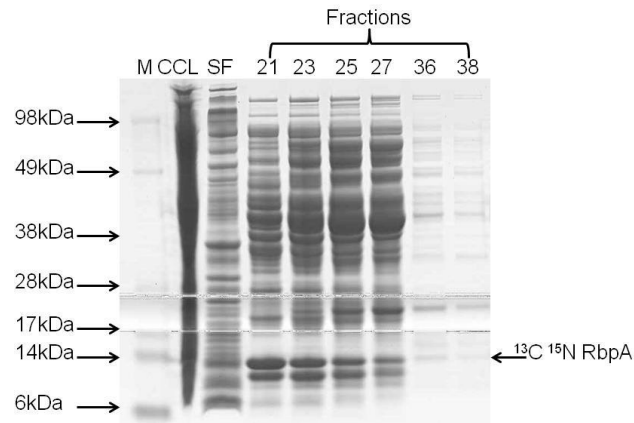
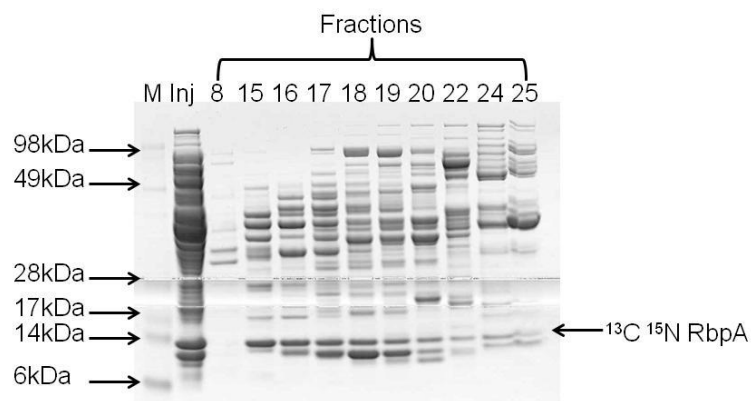
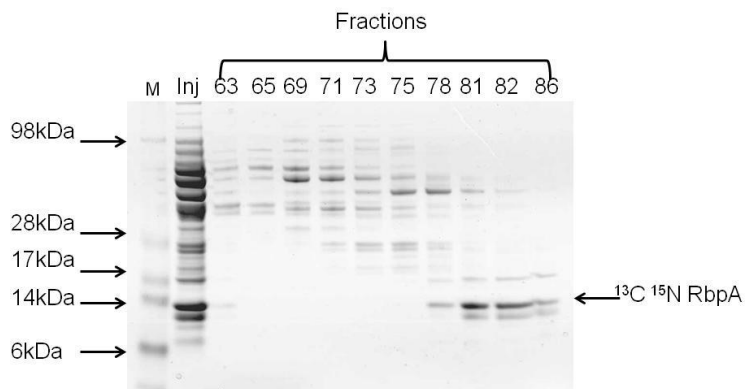
##### 6.4.2.2 Anion-exchange - Mono Q column

Fractions 21 to 25 (Fig. 6.6A) were dialysed into anion-exchange binding buffer then further purified by anion-exchange using a Mono Q column 5/50 GL column. The gradient was set at 0 to 40% for 30 ml followed by 40 to 100% for 5 ml. RbpA eluted between fractions 15 to 19 (~0.4 M NaCl) (Fig. 6.6B). However, fraction 15 was the most pure fraction lacking the lower degraded band (Fig. 6.6B).

##### 6.4.2.3 Gel filtration

Gel filtration chromatography was used as the final step of purification of  $^{13}\text{C}$   $^{15}\text{N}$  RbpA. Fraction 15 (Fig. 6.6B) was dialysed into gel filtration buffer (see above - 100 mM NaCl) at 4°C overnight, filtered, and passed through a pre-equilibrated gel filtration column (HiLoad<sup>TM</sup> 16/60 Superdex<sup>TM</sup> 200 column) (Section 2.3.2.6). Eluates were collected as 1 ml fractions and RbpA was concentrated up (Fig. 6.6C). The

concentration of  $^{13}\text{C}$   $^{15}\text{N}$  RbpA was low due to its gradual degradation (the lower band that appeared below RbpA- Fig. 6.6C). Therefore, an overall of 2 L of native  $^{13}\text{C}$   $^{15}\text{N}$  RbpA was expressed and purified (see above) to obtain 13.18 mg/ml in a final volume of 200  $\mu\text{l}$ .  $^{13}\text{C}$   $^{15}\text{N}$  RbpA was then dialysed into a standard NMR buffer (Section 2.1.6.2) overnight, snap-frozen in liquid nitrogen after the addition of 0.02% sodium azide, and finally sent for NMR spectroscopy.

**A****B****C**

**Figure 6.6: The purification of  $^{13}\text{C } ^{15}\text{N}$  RbpA.**

The samples were assessed by separating the proteins on 4-12% Bis-Tris polyacrylamide gel at 120 V for 80 min and visualized by Coomassie Blue staining. The Marker used was SeeBlue®Plus2. **(A)** QFF anion-exchange chromatography. 10  $\mu\text{l}$  of the cleared cell lysate (CCL), supernatant flow-through (SF) and different 1 ml fractions were loaded on the gel. **(B)** Mono Q anion-exchange purification. 10  $\mu\text{l}$  of injected Rv2050 (Inj) and different 1 ml fractions loaded on gel. **(C)** Gel filtration purification of  $^{13}\text{C } ^{15}\text{N}$  Rv2050. The same volumes of samples loaded as in **B**.

### 6.4.3 The C-terminal region

Throughout the purification of RbpA a common problem faced was the presence of a band below RbpA that might be a degradation product of RbpA (e.g. Fig. 6.6). To confirm this, the two separate bands were excised separately from an SDS polyacrylamide gel and analysed by mass spectroscopy following digestion with trypsin (L. Bowler, University of Sussex). The results indicated that the upper band contained fragments that covered the whole length of RbpA, whilst the lower band lacked fragments from the C-terminal region of RbpA (data not shown). This suggests that the lower band was RbpA lacking its C-terminal region. This provides further evidence that the C-terminal region is critical for interaction with sigma, as the lower band that is RbpA protein lacking C-terminal region did not interact with  $\sigma^{\text{HrdB}}$  (Fig. 4.7).

### 6.4.4 NMR analysis

As with Rv2050, the C-terminal region of  $^{13}\text{C}$   $^{15}\text{N}$  RbpA was not visible by NMR, probably due to degradation. However, the N-terminal region contained 50 amino acids in the structured region ranging from residues 21 to 71 (Fig. 6.7). The grey highlighted residues were successfully assigned (K. O'Dwyer, P. Simpson and S. Matthews, personal communication). However, a constant problem faced was the continued degradation of RbpA during the acquisition of the NMR data.

MSERALRGTRLVVTSYETDR**GIDLAPRQAVEYACEKGHRFEMPFSVEAEIPPEWE**  
**CKVCGAQALLVDGDG**PEEKKAKPARTHWDMLMERRTREELEEVLLEERLAVLRSG  
 AMNIAVHPRDSRKSA

**Figure 6.7: The amino acid sequence of RbpA indicating its structured region.**

The amino acid sequence of RbpA showing the structured region (in bold) and the residues that have been assigned (in grey highlight).

The structure of RbpA was observed to be highly similar to that of Rv2050, as predicted by PSIPRED (Section 3.1.1). RbpA contains four  $\beta$ -strands interconnected by loops at the N-terminal region (S. Matthews, personal communication). RbpA was

also shown to co-ordinate a metal. Using 1D NMR, it was noticed that the addition of EDTA to the RbpA solution showed a reduction in the intensity of the structured region and a rise in the flexibility of this region. Therefore, as expected, the bound Zn is likely to be important for maintaining the structural integrity of RbpA (K. O'Dwyer, P. Simpson and S. Matthews, personal communication).

Similar to Rv2050, it was impossible to obtain the complete structure of RbpA due to the rapid degradation of the C-terminal region. To test whether a faster purification would help obtain intact/full length RbpA, His-tagged RbpA which is partially soluble, was purified at Imperial College London to be analysed immediately by NMR (K. O'Dwyer and S. Matthews, ICL). Following nickel-affinity chromatography, the protein was purified by gel filtration chromatography. Also, to minimise degradation, the purification was done at 4°C. High yield of pure His-tagged  $^{13}\text{C}$   $^{15}\text{N}$  RbpA is in the process of being purified together with further NMR analysis (R. Lewis, M. Paget, P. Simpson, and S. Matthews, Work in progress).

## **6.5 Attempts to solve the structure of an RbpA/Rv2050-sigma complex**

The structure obtained so far of Rv2050 only contained the N-terminal region but not the C-terminal region. NMR and mass spectrometry analysis of RbpA also showed that the C-terminal region was missing. It is possible that the C-terminal region is flexible and that this enhances its susceptibility to degradation. It was shown in Chapter 5, that this is the key region that interacts with the principal sigma factor. It therefore follows that sigma might stabilise the C-terminal region in a complex between the two proteins, thereby allowing the structure to be solved.

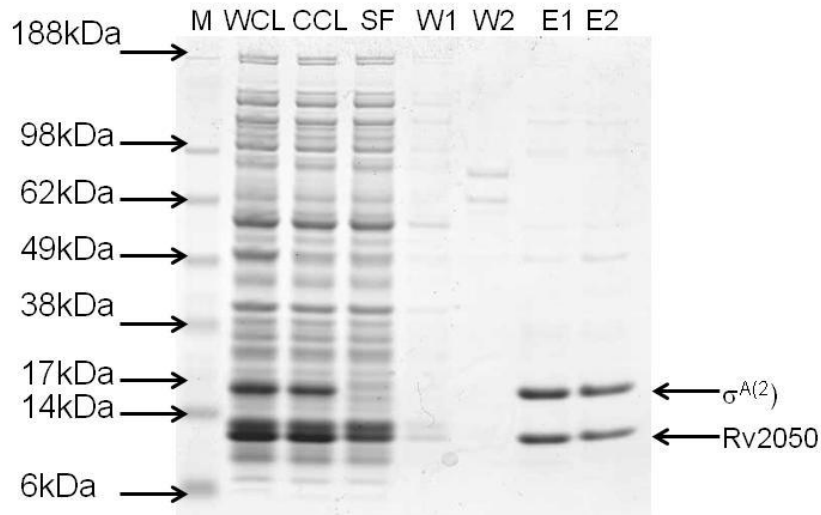
### **6.5.1 Isolation of Rv2050- $\sigma^{A(2)}$**

Rv2050 was shown to interact with  $\sigma^{A(2)}$  using BACTH (Section 4.3.2.1). It was important to show that this interaction occurs *in vitro*, which would allow the complex to be purified before further structural studies. Different approaches were taken to obtain this complex. The first attempt was to over-express and purify Rv2050 and  $\sigma^{A(2)}$  separately and then mixed together to check for interaction. Another approach was to clone Rv2050 and  $\sigma^{A(2)}$  together into the same expression vector, pET15b, allowing their co-expression and co-purification.

*sigA*<sup>(2)</sup> was amplified using *sigA*(2) primers introducing an NdeI site at the start codon and a BamHI site downstream of a newly introduced stop codon (Table 7- Section 2.1.5.3). This amplified fragment was cloned into pBluescript II SK+, sequenced, then sub-cloned into pET15b and pET20b (Table 1- Section 2.1.2). Using *E. coli* BL21 (pLysS) (pET20b::*sigA*<sup>(2)</sup>),  $\sigma^{A(2)}$  was not detected in both soluble and insoluble fractions (data not shown).  $\sigma^{A(2)}$  was highly expressed from *E. coli* BL21 (pLysS) (pET15b::*sigA*<sup>(2)</sup>), but unfortunately was absent from a cleared cell lysate (data not shown), indicating that it is insoluble.

It was decided to follow published methods to solubilise  $\sigma^{A(2)}$ . The first approach was by performing Ni-affinity purification using 6 M Urea buffers (Section 2.1.6.2 & 2.3.2.2), followed by step dialysis, involving the gradual removal of urea, allowing the protein to re-nature slowly. However,  $\sigma^{A(2)}$  precipitated out during the removal of urea. Another approach, adapted from Ignatova and co-workers (Ignatova & Gierasch, 2006, Fisher, 2006), involved over-expressing  $\sigma^{A(2)}$  in the presence of high salt and proline, which is an osmoprotectant. However,  $\sigma^{A(2)}$  remained insoluble showing that proline has no effect on  $\sigma^{A(2)}$  solubility.

A different approach was then taken by co-expressing  $\sigma^{A(2)}$  and Rv2050 in pET15b. Rv2050 was isolated from pET20b::*Rv2050* as a HindIII/BglII fragment and cloned into HindIII/BamHI-digested pET15b::*sigA*<sup>(2)</sup>. *E. coli* BL21 (pLysS) competent cells were transformed with the resulting plasmid, pET15b::*sigA*<sup>(2)</sup>-*Rv2050*. Rv2050 and  $\sigma^{A(2)}$  were both over-expressed and soluble as both proteins were present in the cleared cell lysate fraction (CCL). When the CCL was passed through the Ni-NTA column, both proteins bound to the column and were co-eluted, even though Rv2050 was not His-tagged. This further indicates that Rv2050 and  $\sigma^{A(2)}$  interact with each other (Fig. 6.8).



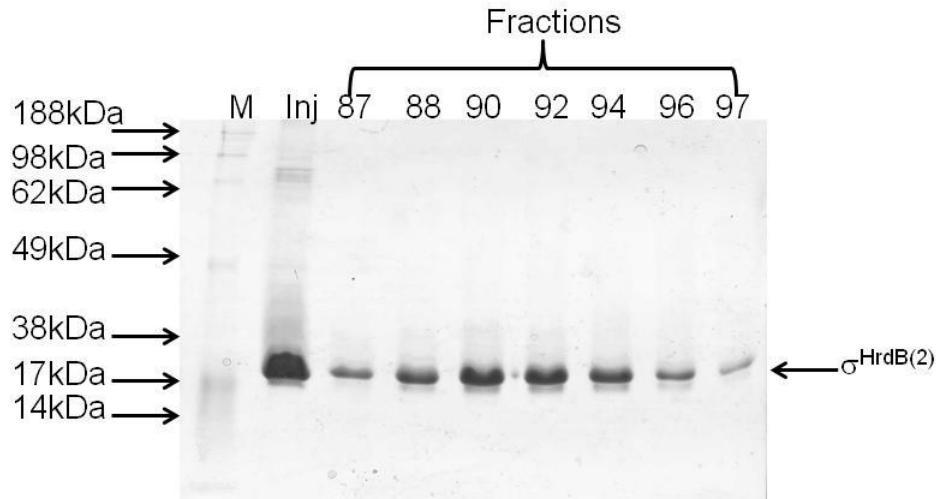
**Figure 6.8: The expression, solubility and interaction of Rv2050 and  $\sigma^{A(2)}$ .**

13  $\mu$ l of the whole cell lysate (WCL), cleared cell lysate (CCL), supernatant flow-through (SF), first wash (W1), second wash (W2), first elution (E1) and second elution (E2) were assessed by separating the proteins on 4-12% Bis-Tris polyacrylamide gel at 120 V for 80 min. The protein was visualized by staining with Coomassie Blue. The marker (M) used was SeeBlue<sup>®</sup>Plus2.

Unfortunately, when the Rv2050 and  $\sigma^{A(2)}$  elution fractions were dialysed into gel filtration buffer (Section 2.1.6.2) for a final purification step, the proteins precipitated out of solution. Further work is ongoing to try to maintain protein solubility. Following this, the complex will be sent for NMR analysis or for X-ray crystallography (R. Lewis, and M. Paget, work in progress).

### 6.5.2 Isolation of RbpA- $\sigma^{\text{HrdB(2)}}$

In Section 4.3.2.1,  $\sigma^{\text{HrdB(2)}}$  was shown to interact both with RbpA and its orthologue, Rv2050 in *M. tuberculosis*. Since  $\sigma^{\text{HrdB(2)}}$  appears to be more soluble than the equivalent fragment in  $\sigma^A$ , it was decided to attempt to study the structure of a complex between  $\sigma^{\text{HrdB(2)}}$  and RbpA. Hence,  $\sigma^{\text{HrdB(2)}}$  was purified by Ni-affinity chromatography followed by gel filtration (Fig. 6.9) and the appropriate buffer conditions were identified to keep the protein in solution.



**Figure 6.9: Gel filtration purification of  $\sigma^{\text{HrdB}(2)}$ .**

13  $\mu\text{l}$  of the fraction injected into gel filtration column (Inj) together with the different 1 ml fractions collected from the AKTA purifier were assessed by separating the protein samples on 5-15% Bis-Tris polyacrylamide gel at 200 V for 70 min. The protein bands were visualised using Coomassie Blue staining. The Marker (M) used was SeeBlue<sup>®</sup>Plus2.

Pure  $\sigma^{\text{HrdB}(2)}$  (fractions 87 to 97, Fig. 6.9) was concentrated to 1.6 mg/ml, sodium azide added and then snap frozen in liquid nitrogen in 200  $\mu\text{l}$  aliquots. In collaboration (K. O'Dwyer, P. Simpson, and S. Matthews, ICL),  $\sigma^{\text{HrdB}(2)}$  was titrated into <sup>15</sup>N-labelled RbpA. Preliminary data obtained also suggested an interaction between the C-terminal flexible/degradable region of RbpA and  $\sigma^{\text{HrdB}(2)}$  due to observed peak broadening followed by peak disappearance (Section 5.1.1) (K. O'Dwyer, P. Simpson, and S. Matthews, unpublished observations).

Further work is ongoing to express high yields of labelled  $\sigma^{\text{HrdB}(2)}$ , labelled RbpA, and unlabelled RbpA. These will then be titrated into each other to start assigning residues within the two proteins and thus, obtain the overall structure of RbpA- $\sigma^{\text{HrdB}(2)}$  complex (R. Lewis, M. Paget, P. Simpson and S. Matthews, personal communications).



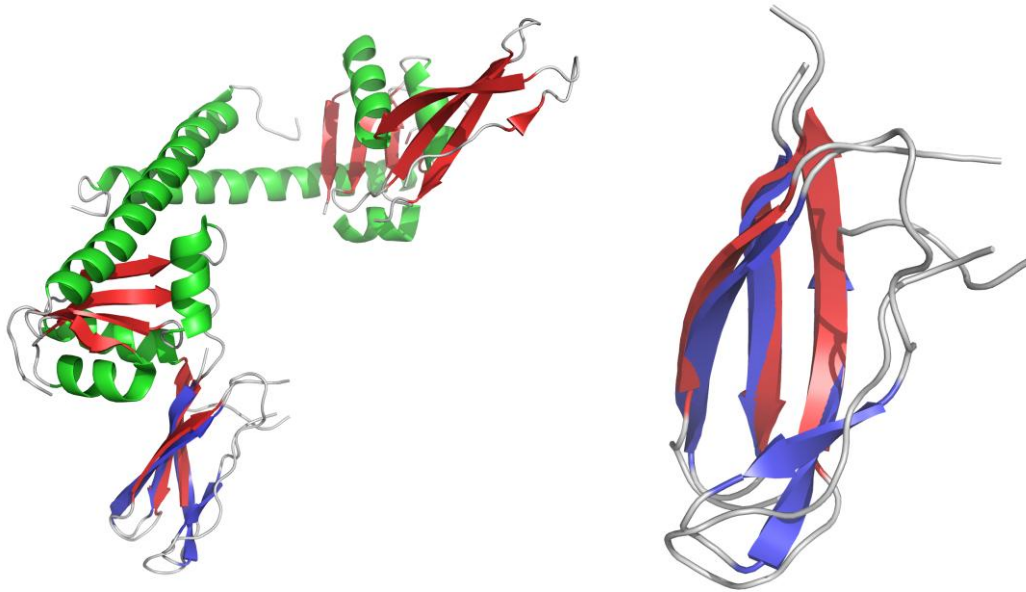
## **6.6 Discussion**

Solving the structure of Rv2050 using X-ray crystallography failed as Rv2050 produced poor crystals that did not diffract. However, using NMR spectroscopy, the structure of the N-terminal region of Rv2050 has been solved and this region is composed of four anti-parallel  $\beta$ -strands interconnected by flexible loops. The C-terminal region of Rv2050 appeared to be flexible and/or degradable. The structure of RbpA is also underway to be solved and its N-terminal region is likely to be very similar to that of Rv2050, however, RbpA co-ordinates a Zn (II) ion. The C-terminal region of RbpA also appeared to be flexible and/or degradable and since this region is involved in the interaction with the sigma factor (Section 5.1.1), methods have been developed to allow the purification of a complex (RbpA- $\sigma^{\text{HrdB}(2)}$  and/or Rv2050- $\sigma^{\text{A}(2)}$ ), which will allow the structure of the complex to be solved in the future.

### **6.6.1 Protein homology of the N-terminal region of Rv2050**

NMR spectroscopy revealed that the N-terminal region of Rv2050 is composed of four anti-parallel  $\beta$ -strands interconnected by loops (Fig. 6.4 - Right). Using the DALI server, the N-terminal region of Rv2050 was shown to share protein homology with the C-terminal domain of Peptide Chain Release Factor subunit 1 (RF1) from the archaeobacterium, *Methanosarcina mazei* (Fig. 6.10), and some small ribosomal proteins (Table 13). RF1 was shown to co-ordinate Zn using four cysteine residues located at either end of the monomers by the  $\beta$ -strands (Fig. 6.10).

The N-terminal region of Rv2050 is similar to the C-terminal domain of RF1 as they both have anti-parallel  $\beta$ -strands inter-connected by flexible loops (Fig. 6.4- Right & 6.10). Interestingly, whereas Rv2050 lacks Zn, RbpA contains Zn (Section 6.3). The Zn (II) ion present in RF1 appears to be in an equivalent place as the predicted Zn (II) ion co-ordinated in RbpA. Apart from the structural similarity of these regions, Rv2050/RbpA has no apparent functional relation with RF1 or any of the DALI search hits (Table 13).



**Figure 6.10: The X-ray crystal structure of the C-terminal domain of the peptide chain release factor 1 (RF1) dimer from *Methanosarcina mazei* overlaid with Rv2050 N-terminal region.**

**Left-** The structure of the C-terminal domain of RF1 dimer from *Methanosarcina mazei* overlaid with Rv2050 N-terminal region. The red ribbons denotes the  $\beta$ -strands of RF1 and the blue ribbons denote the  $\beta$ -strands of Rv2050. RF1 crystal structure was obtained from the RSCB PDB database no. 3IR9 (Osipiuk et al., to be published). **Right-** The overlay region of RF1 and Rv2050. The red ribbons denote  $\beta$ -strands of RF1 whilst the blue ribbons denote the  $\beta$ -strands of Rv2050. PyMOL software was used and alignment done using cealign. The RMSD value is 2.67 (a value of 0 means perfect alignment) and alignment is over 40 residues.

### 6.6.2 Zn metal as a structural component of RbpA

Zn is an essential metal required in all living organisms and can play a structural, catalytic, regulatory or substrate role (Andreini *et al.*, 2011). However, excess Zn in cells can be toxic. Zn (II) ions, co-ordinated by four amino acid residues, are

frequently found to play structural roles in protein folding. The most common occurring Zn co-ordinating residues are cysteine (80%) and histidine (19%), with the few rare exceptions being aspartate, glutamate and serine (Andreini *et al.*, 2011). Also, most proteins that co-ordinate Zn for proper folding have at least two cysteine residues. Cysteine is favoured above other amino acid residues as it makes stronger bonds with the Zn (II) ion due to its ability to transmit negative charge to the Zn (II) ion (Andreini *et al.*, 2011). Many structural Zn-clusters exist as “Zinc ribbons” with two  $\beta$ -strands parallel to two other  $\beta$ -strands and the Zn (II) ion present in the middle (Andreini *et al.*, 2011).

RbpA appears to co-ordinate a Zn metal in its N-terminal region using its three cysteine residues and one histidine residue (Section 3.1.1). This arrangement is consistent with the role of Zn in RbpA being structural. Also, from the PSIPRED RbpA secondary structure prediction (Section 3.1.1) and initial NMR analysis data (Section 6.4.4), RbpA was shown to have a “Zn ribbon”, which contains the amino acid residues co-ordinating the Zn (II) ion. There is no data that suggests that RbpA is a redox-sensing protein and it is thought that the Zn plays a purely structural role. The paralogue of RbpA in *S. coelicolor*, RbpB, also has the three cysteine residues (C35, C57 and C60) and the one histidine (H39) residue, which suggests that RbpB may also be a Zn-binding protein. However, Rv2050, the orthologue of RbpA in *M. tuberculosis*, does not co-ordinate any metal and two of the three cysteine ligands and the histidine ligand are absent (Section 6.3). Genome sequencing is revealing increasing numbers of examples where homologous proteins occur in alternative Zn-binding or Zn-free forms, depending on lineage.

### **6.6.3 Protein homologues that exist in Zn and non-Zn form**

#### **6.6.3.1 Ribosomal proteins**

Ribosomal proteins stabilise rRNA and may play a functional role in the ribosomes. *S. coelicolor* has seven ribosomal proteins (S14, S18, L28, L31, L32, L33 and L36), which exist in both the Zn and non-Zn form (Owen *et al.*, 2007). In most cases the Zn (II) ion is co-ordinated by four cysteine residues and has a structural role (Owen *et al.*, 2007), whereas, the non-Zn form has apparently lost its “Zn ribbon” (Makarova *et*

*al.*, 2001). It was shown that L31 can partly replace the Zn form of L31 in *S. coelicolor* (Owen *et al.*, 2007). However, other organisms have only the non-Zn form of L31 implying that this form is fully functional.

The reason why organisms might evolve Zn-free forms of r-proteins became clear when the regulation of these genes was studied. The non-Zn forms of these ribosomal proteins were induced during Zn limitation, on account of their control by the Zn regulator Zur (Panina *et al.*, 2003, Owen *et al.*, 2007). It is hypothesised that this reduces cellular demand for Zn, and possibly stimulates the release of the Zn (II) from Zn-binding r-proteins, which may be incorporated by other Zn metalloproteins (Owen *et al.*, 2007).

#### 6.6.3.2 DksA

DksA, an RNAP binding protein, was shown to coordinate Zn using four cysteine residues (Section 1.2.6.1). The Zn (II) ion is chelated by two cysteine residues in the globular domain and two cysteine residues in the coiled-coil domain. The role of Zn in DksA is to ensure proper folding of the globular domain and is essential for DksA functionality (Blaby-Haas *et al.*, 2011). DksA was shown to exist in both the Zn and non-Zn (DksA2) form in *P. aeruginosa* (Blaby-Haas *et al.*, 2011).

DksA2, the paralogue of DksA, is fully functional and behaves in a similar manner to DksA inhibiting open complex formation (Section 1.2.6.1). DksA2 is thought to be expressed in poor Zn conditions and repressed in conditions of Zn availability and this regulation is controlled by Zur (Blaby-Haas *et al.*, 2011).

*P. aeruginosa* is a common pathogen that infects and causes death in many cystic fibrosis patients. Calprotectin, a neutrophil protein that sequesters Zn, is present in high levels in cystic fibrosis sputum. Thus, this protein leads to a low Zn environment in the sputum thereby expression of DksA2 is stimulated, which aids the survival of the bacterium in such conditions (Blaby-Haas *et al.*, 2011).

#### 6.6.3.3 Superoxide dismutase (SodC)

Superoxide dismutase co-ordinate metals and catalyse the dismutation of superoxide radicals ( $O_2^-$ ) into molecular oxygen ( $O_2$ ) and hydrogen peroxide ( $H_2O_2$ )

(Spagnolo *et al.*, 2004). *sodC* encodes a Cu-, Zn-dependent superoxide dismutase. The Zn metal present in *sodC*-encoded enzyme is coordinated by three histidine residues and one aspartate residue and is important to maintain the structural integrity of the protein. However, in *M. tuberculosis* the *sodC*-encoded superoxide dismutase exists in the non-Zn form yet is fully functional (Spagnolo *et al.*, 2004).

SodC is required during *M. tuberculosis* infection. When the bacterium invades the macrophages, *sodC* expression is stimulated and is involved in the defence against the host-derived superoxide (Spagnolo *et al.*, 2004). It was proposed that Zn might be limiting in the host which might account for the evolution of this Zn-free form of SodC (Spagnolo *et al.*, 2004, Riccardi *et al.*, 2008). An analogous explanation can be proposed for the apparent evolution of Rv2050, which is an essential protein in *M. tuberculosis*, to a Zn-free form. The RbpA orthologue in *M. smegmatis* also lacks Zn suggesting that the lost of Zn may have evolved to aid these microorganisms in Zn limiting conditions and gives them a competitive advantage against others.

## **7- General Discussions and Summary**

## **7.0 Overview**

RbpA, an RNAP binding protein, was discovered when it co-eluted with RNAP during gel filtration and is partially regulated by  $\sigma^R$  (Paget *et al.*, 2001b). *S. coelicolor*  $\Delta rbpA$  strains grow slowly producing smaller sized colonies than wild-type and overproduce the antibiotic actinorhodin (Newell *et al.*, 2006). These strains were also shown to be sensitive to the antibiotic rifampicin (Newell *et al.*, 2006), however, it is not thought that RbpA confers rifampicin resistance. RbpA was shown to interact with  $\sigma^{HrdB}$ , the principal sigma factor in *S. coelicolor*, and to activate  $\sigma^{HrdB}$ -dependent transcription (P. Doughty and M. Paget, unpublished observations). It was shown in this thesis that RbpA specifically interacts with domain 2 of  $\sigma^{HrdB}$  using its C-terminal region. Furthermore, the orthologue of RbpA in *M. tuberculosis*, Rv2050, was also shown to interact in a similar manner to RbpA by binding to domain 2 of its principal sigma factor,  $\sigma^A$ .

Previous work showed that RbpA appears to be involved in an early stage of transcription initiation (P. Doughty and M. Paget, personal communication). However, the exact mechanism of action of RbpA has not yet been determined. Work done in this thesis suggests that RbpA is not involved in holoenzyme formation, but may be involved at some point within closed-open complex formation.

Using NMR spectroscopy, the N-terminal region of Rv2050 has been solved and the equivalent region in RbpA is also underway to be solved. The C-terminal region appeared to be flexible and susceptible to degradation therefore, as this region interacts with sigma, the underlying work of obtaining a soluble sigma fragment that interacts with RbpA and its purification was completed. The overall structure of RbpA- $\sigma^{HrdB(2)}$  is currently subject to investigation (In collaboration with R. Lewis and M. Paget, work in progress).

## **7.1 The RbpA family**

The RbpA family is present in only the Actinobacteria and is divided into three distinct classes. Two of these classes are likely to include orthologous proteins that differ by the presence or absence of putative Zn ligands e.g. RbpA in *S. coelicolor*, which co-ordinates Zn (II) using its three cysteine residues (C34, C56 and C59) and

possibly one histidine residue (H38), and Rv2050 in *M. tuberculosis* that lacks the putative Zn ligands (C34, C59 and H38). The third class are the RbpA paralogues, designated RbpB, which are less conserved amongst the Actinobacteria but are present in all of the *Streptomyces* and most of the *Frankia* genomes sequenced to date. Each RbpB orthologue has putative Zn ligands (C35, H39, C57, and C60 in *S. coelicolor* RbpB) suggesting that each co-ordinates Zn.

### 7.1.1 Why is RbpA only present within the Actinobacteria?

There appears to be no orthologous or even analogous proteins to RbpA outside of the Actinobacteria, raising the important question of why these proteins are required in this family. There is no clear answer to this question at present, but there are several possibilities. For example, the Actinobacteria contain a high G+C content, and therefore RbpA might aid in the melting of DNA during transcription initiation. However, this hypothesis is probably not true as RbpA is absent from some Actinobacteria containing a high G+C content (*e.g. Conexibacter woessii*, 72.7% G+C, Pukall *et al.*, 2009) yet present in others with a lower G+C content (*e.g. M. leprae*, 57.8% G+C content, Cole *et al.*, 2001). Nonetheless, a detailed analysis of G+C content at transcription initiation sites is yet to be performed, and so it is not clear whether the overall high G+C content is reflected by local high G+C content in promoter regions.

Another possibility why RbpA is only found within the Actinobacteria might be due to the presence of some difference in the principal sigma factors or other subunits of the RNAP of the Actinobacteria that causes a requirement for RbpA. Multiple sequence alignment of  $\sigma^{\text{HrdB}}$  with other  $\sigma^{\text{HrdB}}$ -like sigma factors in the Actinobacteria containing and lacking RbpA did not show any evident dissimilarity. Many  $\sigma^{\text{HrdB}}$ -like sigma factors have a strong positive (Lysine, K-rich) and negative (Aspartate, D- and Glutamate, E-rich) region 1.1 followed by well conserved domains 2-4 (Fig.A2-Appendices). However, some organisms lacked the highly basic region (Lysine, K).

RbpA is not the only protein to be conserved and limited to the Actinobacteria. Other proteins (233 proteins) have been identified to be specific for the Actinobacteria (Gao *et al.*, 2006). However, 29 of these proteins are found in most of the Actinobacteria



with some of these proteins absent within *R. xylanophilus* and/or *T. whipplei* (Gao *et al.*, 2006), which is similar to the RbpA proteins (Table A1- Appendices). Is it possible that RbpA interacts with one of these proteins that are specific to the Actinobacteria and functions through this? This might explain why RbpA is conserved only within the Actinobacteria.

### **7.1.2 Is the RbpA family essential in the Actinobacteria?**

RbpA is essential for the normal growth of *S. coelicolor*, whilst RbpB is non-essential and appears to be important only in the absence of RbpA (Newell *et al.*, 2006). Since the construction of a *S. coelicolor*  $\Delta rbpA \Delta rbpB$  double mutant has not been successful, it is thought that RbpA and RbpB are partially redundant and *rbpB* is able to compensate for the absence of *rbpA*. In support of this, the overexpression of *rbpB* in a  $\Delta rbpA$  strain partially suppresses the slow-growth phenotype. Therefore, the combined function of RbpA and RbpB may be essential for the survival of *S. coelicolor*.

The application of global transposon mutagenesis and microarray-based analysis of mutant libraries first provided some evidence that Rv2050, the homologue of RbpA in *M. tuberculosis*, might be a core gene required for optimal growth *in vitro* (Sassetti *et al.*, 2003, Marmiesse *et al.*, 2004). This was confirmed recently by the construction of a conditional mutant in *M. tuberculosis* (Forti *et al.*, 2011).

The importance of RbpA is further suggested by its presence in the highly reduced genome of *M. leprae* (Cole *et al.*, 2001). It was shown here that Rv2050 interacts with the principal sigma factor of *M. tuberculosis* and therefore, is likely to play a key role in transcription initiation in this organism. Therefore, Rv2050 may be a possible target for drug design in the combat against mycobacterial infections such as tuberculosis, which is a persistent infection claiming millions of lives.

### **7.2 RbpA as a distinct transcriptional activator**

RbpA stimulates transcription initiation although the mechanism involved is unclear. The evidence presented in this thesis suggests that this activation occurs through an

interaction with the principal sigma factor, although the additional interaction with one or more other subunits of RNAP cannot be discounted.

Most of the described activators in the literature have been shown to interact with DNA in addition to a subunit of the RNAP holoenzyme. Activators that bind close to or overlapping the -35 promoter region often interact with the sigma factor, while those binding further upstream tend to interact with the  $\alpha$  subunit (Rhodius & Busby, 1998). Some activators can interact with several sites on RNAP. For example, the most studied transcriptional activator in the last few decades is the *E. coli* catabolite activator protein (CAP). CAP activates transcription from different classes of promoters differently. At some promoters CAP binds upstream from the RNAP DNA binding site (e.g. *lac* operon) and contacts the  $\alpha$ -CTD of RNAP stabilising closed complex (Igarashi & Ishihama, 1991, Busby & Ebright, 1999). At other promoters CAP overlaps the -35 region and contacts the  $\alpha$ -CTD and  $\alpha$ -NTD using distinct activating regions (Niu *et al.*, 1996). At this position, it is also thought to contact the sigma factor domain 4 (Rhodius & Busby, 2000). The interaction with the  $\alpha$ -CTD stimulate RNAP holoenzyme to bind to promoters forming closed complex, whilst the interaction with the  $\alpha$ -NTD leads to isomerisation from closed to open complex (Niu *et al.*, 1996).

A particularly well-studied example of an activator that interacts with sigma is the  $\lambda$ cl repressor that activates its own expression, while repressing rightward lytic gene expression.  $\lambda$ cl was shown to make specific contacts with the C-terminal region of *E. coli*  $\sigma^{70(4)}$  (Li *et al.*, 1994, Dove *et al.*, 2000). From the structure of  $\sigma^{70(4)}$  and  $\lambda$ cl it was shown that residues in  $\sigma^{70(4)}$  interact with  $\lambda$ cl, thereby stabilising the interaction of  $\sigma^{70}$  region 4 with the -35 element (Nickels *et al.*, 2002).

More unusual examples of activators interacting with other subunits of RNAP include: DnaA, which in addition to its role in DNA replication in *E. coli*, is also involved in transcription activation through interactions with the  $\beta$ -subunit of RNAP at the  $P_R$  promoter of bacteriophage  $\lambda$  (Szalewska-Palasz *et al.*, 1998);  $N_{4SSB}$ , encoded by the bacteriophage  $N_4$ , which activates transcription of the  $N_4$  *late* promoter by

interacting with the C-terminal region of the  $\beta'$  subunit of RNAP- $\sigma^{70}$  (Miller *et al.*, 1997).

RbpA, however, is distinct from the above transcriptional activators in two key ways. Firstly, it does not appear to interact with DNA. Secondly, it specifically interacts with domain 2 of  $\sigma^{\text{HrdB}}$  (Section 4.2). This is particularly unusual because most other described proteins that interact with sigma domain 2 are anti-sigma factors that have a negative effect on transcription. Some of the common examples include: FlgM- $\sigma^{28}$  in *Salmonella typhimurium* (Section 1.2.7), Rsd- $\sigma^{70}$  in *E. coli* (Section 1.2.7.1), and SpoIIAB- $\sigma^F$  in *B. subtilis* (Campbell & Darst, 2000). These sigma factors appear to block sigma from associating with core RNAP (Rsd; Jishage & Ishihama, 1999, and SpoIIAB; Campbell & Darst, 2000) or destabilise the holoenzyme (FlgM; Chadsey *et al.*, 1998) as it binds to domain 2, which make major interactions with core RNAP. Clearly, RbpA must interact with this region in a distinct way that does not block core interactions.

Recently, Crl has been described as a transcriptional activator that interacts with sigma factor but does not interact with DNA (Section 1.2.9.1). It directly binds to the stationary phase sigma factor,  $\sigma^S$  and increases  $\sigma^S$ -dependent transcription in *E. coli* (Gaal *et al.*, 2006). Furthermore, Crl was shown to also interact with domain 2 of  $\sigma^S$  in *Salmonella enteric* Typhi with a 1:1 stoichiometry (England *et al.*, 2008, Monteil *et al.*, 2010b). The addition of Crl to  $\sigma^S$  prior to core RNAP increased transcription activity showing that Crl might act by facilitating RNAP- $\sigma^S$  holoenzyme formation (Gaal *et al.*, 2006). Crl was also shown to increase the affinity of  $\sigma^S$  for RNAP by 7-fold (England *et al.*, 2008), thereby giving  $\sigma^S$  an advantage to bind to core RNAP during sigma competition (Typas *et al.*, 2007).

### **7.3 The role of RbpA in transcription initiation**

RbpA appears to be involved in an early stage of transcription initiation as shown by its stimulation of transcription in single-round *in vitro* transcription assays and the formation of open-complexes (P. Doughty and M. Paget, personal communications). However, its exact mechanism of action is yet to be determined. As mentioned above, RbpA interacts with domain 2 of  $\sigma^{\text{HrdB}}$ , which is composed of sub-regions:

1.2, 2.1, 2.2, 2.3 and 2.4. These regions are well conserved amongst the  $\sigma^{70}$ -like sigma factors and the regions are involved in forming major interactions with  $\beta'$ -coiled-coil subunit of RNAP (region 1.2, 2.1 and 2.2) (Malhotra *et al.*, 1996, Zenkin *et al.*, 2007), melting of the -10 element to form the transcription bubble (region 2.3) (Feklistov & Darst, 2011), and the -10 element promoter recognition (region 2.4) (Murakami & Darst, 2003). In addition, the 1.2 region was recently shown to interact with the “discriminator region” (Haugen *et al.*, 2008).

Formation of an active holoenzyme can be seen as the first step in initiation and is the step at which Crl is thought to act (see above). A possible role of RbpA in RNAP- $\sigma^{\text{HrdB}}$  formation was tested using *in vivo* pull-down assays (Section 3.2.4). However, the level of  $\sigma^{\text{HrdB}}$  in affinity purified RNAP fractions was similar in wild-type and  $\Delta rbpA$  mutant backgrounds. In addition, preliminary results obtained using surface plasmon resonance assays *in vitro* suggest that RbpA does not stimulate holoenzyme formation (P. Doughty and M. Paget, personal communication). A possible role of RbpA in formation of the closed complex at  $\sigma^{\text{HrdB}}$ -dependent promoters has been difficult to test because of the natural affinity of  $\sigma^{\text{HrdB}}$  for non-specific DNA, probably due to its positively charged repeat region.

RbpA might alternatively play a role in isomerisation from closed to open complex. Preliminary data obtained from  $\text{KMnO}_4$  assays suggested that RbpA stimulates open complex formation although the stage at which it acts could not be defined (P. Doughty and M. Paget, personal communication). If RbpA was involved in isomerisation it might aid in the melting of the -10 element directing it to transcription start site or it might stabilise the transcription bubble. For example, RbpA might bind to  $\sigma^{\text{HrdB}}$  within the RNAP holoenzyme and affect the conformation of the sigma factor exposing region 2.3, hence, allowing initiation of DNA melting of the non-template -10 element (Feklistov & Darst, 2011) and/or positioning region 2.4 to make specific interactions with bases in the non-template -10 element (Feklistov & Darst, 2011), thereby shifting the equilibrium towards the formation of the open complex.

An example of a transcriptional activator that increases the rate of isomerisation from closed to open complex is  $\lambda\text{cI}$  (Hawley & McClure, 1982). It was proposed that  $\lambda\text{cI}$

stabilises the binding of  $\sigma^{70(4)}$  to the -35 element leading to formation of a stable intermediate, which directs the equilibrium towards the open complex formation (Dove *et al.*, 2000). However, as RbpA interacts with domain 2 instead of domain 4 its mechanism of action would be quite distinct. Interestingly, Crl also showed an increase in promoter melting using KMnO<sub>4</sub> assay suggesting a possible role for Crl in transcription bubble formation during open complex (England *et al.*, 2008).

#### **7.4 RbpA structure and region function**

RbpA appears to be divided into two distinct regions: the N-terminal region and the C-terminal region. The structure of the N-terminal region of Rv2050 has been solved using NMR spectroscopy (Section 6.2) and is composed of four anti-parallel  $\beta$ -strands interconnected by loops. Determination of the structure of RbpA is also underway, and from an initial assessment, the N-terminal region was noted to be virtually identical to that of its homologue, Rv2050, as predicted by the PSIPRED (Section 3.1.1). This region of RbpA appears to be involved in maintaining the structural integrity of the protein by co-ordinating a Zn (II) ion using its three cysteine residues (C34, C56 and C59) and possibly one histidine (H38). As mentioned previously, Rv2050, however, lacks Zn and is a non-metalloprotein (Section 6.3).

The C-terminal region of RbpA was not solved by NMR as this region appears to be susceptible to degradation and also possibly flexible. However, from the PSIPRED prediction, the C-terminal region of RbpA seems to be composed of two  $\alpha$ -helices, which are connected by a very short turn (ERR motif). Both helices together with the ERR motif were shown to be highly conserved amongst the RbpA homologues. Both helices in the C-terminal region play a key role in the interaction of RbpA with  $\sigma^{\text{HrdB}}$  (Section 5.1).

Site-directed mutagenesis studies on the C-terminal region of RbpA revealed that the arginine residues in the ERR motif are essential for the interaction with  $\sigma^{\text{HrdB}}$  and when mutated the RbpA<sub>2RA</sub> is non-functional in *S. coelicolor*. Another interesting mutation within Helix 1 of RbpA appeared to be M85. An RbpA M85 mutant protein still interacted with  $\sigma^{\text{HrdB}}$ , however, only retained partial function both *in vivo* (Section 5.2.2) and *in vitro* (Section 5.6). The expression of the M85A allele in a wild-type

background revealed that it is a dominant negative mutation. Presumably, this is because it interacts with the essential  $\sigma^{\text{HrdB}}$  and either prevents its association with core RNAP, or somehow blocks transcription initiation at a later stage. The future analysis of this allele and other dominant mutations is likely to play a key role in the dissection of RbpA function.

Another group has shown that the RbpA homologue in *M. smegmatis* is involved in interacting with RNAP  $\beta$ -subunit using cross-linking experiments (Dey *et al.*, 2010). However, it should be noted that cross-linking reactions could generate non-specific products, which might account for the results. A recent paper showed that RbpA homologue in *M. tuberculosis* interacts with the  $\beta$ -subunit of RNAP at the 'sandwich-barrel-hybrid motif' that is located outside the active site cleft of RNAP (Hu *et al.*, 2012). This is distinct from the interaction predicted by Dey *et al.*, 2010. While our data indicate that sigma is a major binding partner for RbpA, they do not rule out an additional interaction with other subunits. For example, it might be possible that RbpA interacts with core subunits via its N-terminal region.

### **7.5 Future directions**

RbpA is thought to be a general positive regulator activating  $\sigma^{\text{HrdB}}$ -dependent promoters. Is RbpA specific towards a particular promoter sequence or does it activate all  $\sigma^{\text{HrdB}}$ -dependent genes? Using global approaches such as ChIP-Seq, it may be possible to identify whether RbpA might act as a promoter specific activator. However, since RbpA is partly expressed by stress (Paget *et al.*, 2001b), it might be possible that it reprogram's transcription and not act as a RNAP subunit but rather as a modulator.

Rv2050 was shown to be essential for *M. tuberculosis* (Forti *et al.*, 2011) and thus, the notion that RbpA together with RbpB are essential in *S. coelicolor* needs to be tested. As RbpA is predicted to be involved in transcription initiation, determining at what stage RbpA is involved in is crucial in understanding transcription in the medically and industrially important Actinobacteria. Biochemical assays together with the structure of the complex: RbpA- $\sigma^{\text{HrdB}(2)}$  would shed light into the role of RbpA. The overall structure of RbpA within the holoenzyme is a major goal, which might

answer many questions regarding the mechanism of action of RbpA. RNAP is a major proven target for anti-tuberculosis drugs, therefore novel essential components of RNAP such as Rv2050 that interact with RNAP and is present in *M. tuberculosis* can be potential drug targets to combat such diseases.

## **Appendices**

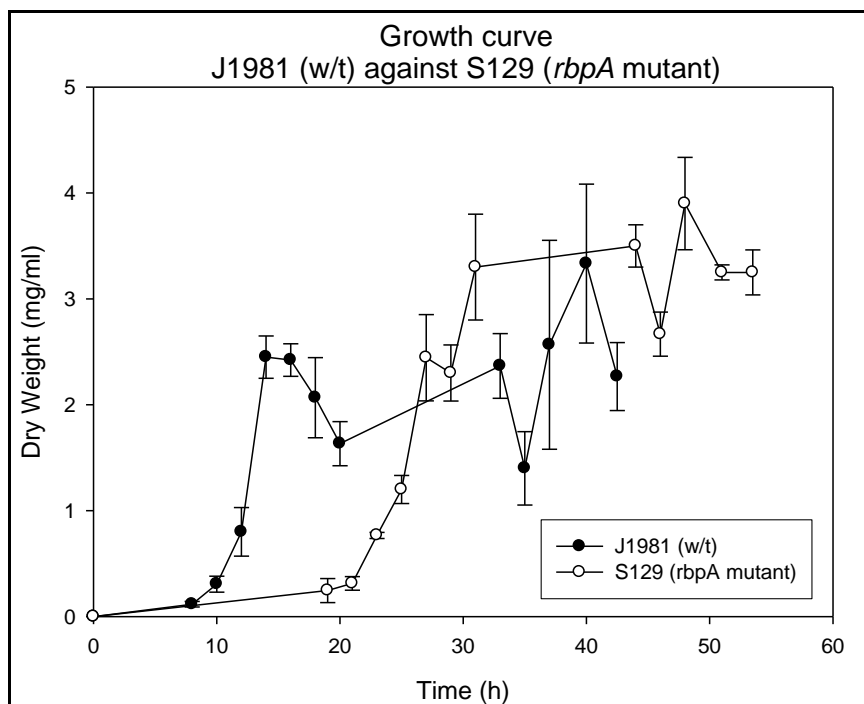


**Table A1: The level of conservation of RbpA and RbpB in the Actinobacteria.**

<b>Actinobacteria</b>	<b>RbpA</b>	<b>RbpB</b>
<i>Acidimicrobium ferrooxidans</i>		
<i>Acidothermus cellulolyticus</i>	P	
<i>Actinomyces coleocanis</i>	P	
<i>Actinosynnema mirum</i>	P	P
<i>Aeromicrobium marinum</i>	P	P
<i>Agreia</i> sp.	P	
<i>Amycolatopsis mediterranei</i>	P	P
<i>Arcanobacterium haemolyticum</i>	P	
<i>Arthrobacter arilaitensis</i>	P	
<i>Atopobium parvulum</i>		
<i>Beutenbergia cavernae</i>	P	
<i>Bifidobacterium adolescentis</i>	P	
<i>Brachybacterium faecium</i>	P	
<i>Brevibacterium linens</i>	P	
<i>Catenulispora acidiphila</i>	P	
<i>Cellulomonas flavigena</i>	P	
<i>Clavibacter michiganensis michiganensis</i>	P	
<i>Collinsella aerofaciens</i>		
<i>Conexibacter woesei</i>		
<i>Corynebacterium diphtheriae gravis</i>	P	
<i>Cryptobacterium curtum</i>		
<i>Dermacoccus</i> sp.	P	
<i>Dietzia cinnamea</i>	P	
<i>Eggerthella lenta</i>		
<i>Frankia alni</i> ACN14a	P	P
<i>Gardnerella vaginalis</i>	P	
<i>Geodermatophilus obscurus</i>	P	P
<i>Gordonia bronchialis</i>	P	
<i>Intrasporangium calvum</i>	P	
<i>Janibacter</i> sp.	P	
<i>Jonesia denitrificans</i>	P	
<i>Kineococcus radiotolerans</i>	P	P
<i>Kocuria rhizophila</i>	P	
<i>Kribbella flavida</i>	P	P
<i>Kytococcus sedentarius</i>	P	P
<i>Leifsonia xyli xyli</i>	P	
<i>Micrococcus luteus</i> Fleming	P	
<i>Micromonospora aurantiaca</i>	P	P
<i>Mobiluncus curtisii</i>	P	
<i>Mycobacterium tuberculosis</i> H37Rv	P	
<i>Nakamurella multipartita</i>	P	P
<i>Nocardia farcinica</i>	P	

<b>Actinobacteria</b>	<b>RbpA</b>	<b>RbpB</b>
<i>Nocardioides</i> sp.	P	P
<i>Nocardiopsis dassonvillei</i>	P	P
<i>Olsenella uli</i> VPI		
<i>Parascardovia denticolens</i>	P	
<i>Propionibacterium acnes</i>	P	P
<i>Pseudonocardia</i> sp.	P	P
<i>Renibacterium salmoninarum</i>	P	
<i>Rhodococcus equi</i>	P	
<i>Rothia dentocariosa</i>	P	
<i>Rubrobacter xylanophilus</i>		
<i>Saccharomonospora viridis</i>	P	P
<i>Saccharopolyspora erythraea</i>	P	P
<i>Salinispora arenicola</i>	P	
<i>Sanguibacter keddiei</i>	P	
<i>Scardovia inopinata</i>	P	
<i>Segniliparus rotundus</i>	P	
<i>Slackia heliotrinireducens</i> RHS 1		
<i>Stackebrandtia nassauensis</i>	P	P
<i>Streptomyces coelicolor</i> A3(2)	P	P
<i>Streptosporangium roseum</i>	P	
<i>Thermobifida fusca</i>	P	P
<i>Thermobispora bispora</i>	P	P
<i>Thermomonospora curvata</i>	P	P
<i>Tropheryma whipplei</i>		
<i>Verrucosipora maris</i>	P	
<i>Tsukamurella paurometabola</i>	P	
<i>Xylanimonas cellulolytica</i>	P	P

The list of the Actinobacteria only includes a single representative of the species. RbpA and RbpB are either present (P) in all or absent in all.



**Figure A1: A graph representation of the growth curve of J1981 and S129.**

Samples were taken at early exponential phase and then at 2 h intervals until stationary phase and the dry weights were measured. The error bars shown represent the standard deviation of triplicate samples. The graph was plotted using SigmaPlot.

**Table A2: Genes that are downregulated in *S. coelicolor*  $\Delta rbpA$  strain, S101, relative to wild-type strain, J1915.**

Gene Name	Log <sub>2</sub> Average Cond1/Cond2	inv. Log	Description
SCO0527	-1.079	0.47	cold shock protein
SCO0753	-2.686	0.16	hypothetical protein
SCO1086	-1.488	0.36	hypothetical protein
SCO1139	-0.764	0.59	integral membrane protein
SCO1341	-0.995	0.50	Lipoprotein
SCO1421	-1.363	0.39	<i>rbpA</i> gene
SCO1578	-1.185	0.44	acetylglutamate kinase
SCO1579	-1.002	0.50	bifunctional ornithine acetyltransferase/N-acetylglutamate synthase protein
SCO2628	-0.831	0.56	amino acid permease
SCO2698	-0.210	0.87	small hydrophilic protein
SCO2779	-0.690	0.63	acyl-CoA dehydrogenase
SCO2920	-0.761	0.59	secreted protease
SCO2968	-0.727	0.60	cell division protein
SCO3483	-0.714	0.61	integral membrane transport protein
SCO3484	-0.777	0.58	secreted sugar-binding protein
SCO3527	-0.375	0.77	hypothetical protein
SCO3553	-0.588	0.67	small membrane protein
SCO3661	-0.738	0.60	ATP-dependent protease ATP-binding subunit
SCO3662	-0.923	0.53	hypothetical protein
SCO3663	-0.861	0.55	hypothetical protein
SCO3731	-0.989	0.50	cold-shock protein
SCO3781	-0.751	0.59	secreted protein
SCO3795	-1.103	0.47	aspartyl-tRNA synthetase
SCO3897	-0.721	0.61	hypothetical protein
SCO3961	-1.243	0.42	seryl-tRNA synthetase
SCO4078	-1.017	0.49	phosphoribosyl formylglycinamide synthase I
SCO4079	-0.692	0.62	phosphoribosylformylglycinamide synthase subunit II
SCO4089	-0.761	0.59	valine dehydrogenase
SCO4178	-1.079	0.47	small membrane protein
SCO4179	-0.801	0.57	hypothetical protein
SCO4614	-0.724	0.61	nucleotide-binding protein
SCO4652	-0.430	0.74	50S ribosomal protein L10
SCO4660	-0.789	0.58	30S ribosomal protein S7
SCO4716	-0.514	0.70	30S ribosomal protein S8
SCO4717	-0.392	0.76	50S ribosomal protein L6
SCO4718	-0.466	0.72	50S ribosomal protein L18
SCO4719	-0.541	0.69	30S ribosomal protein S5
SCO4720	-0.430	0.74	50S ribosomal protein L30
SCO4721	-0.460	0.73	50S ribosomal protein L15
SCO4945	-1.150	0.45	Dehydrogenase

Gene Name	Log <sub>2</sub> Average Cond1/Cond2	inv. Log	Description
SCO5049	-0.682	0.62	hypothetical protein
SCO5189	-0.833	0.56	hypothetical protein
SCO5190	-1.351	0.39	DNA-binding protein
SCO5191	-1.035	0.49	hypothetical protein
SCO5209	-1.028	0.49	tetR-family transcriptional regulator
SCO5272	0.126	0.92	hypothetical protein
SCO5519	-0.821	0.57	hypothetical protein
SCO5520	-0.757	0.59	delta-1-pyrroline-5-carboxylate dehydrogenase
SCO5537	-0.660	0.63	ATP/GTP binding protein
SCO5680	-0.806	0.57	cytidine deaminase
SCO6592	-0.663	0.63	secreted protein
SCO6624	-0.842	0.56	hypothetical protein
SCO7036	-0.909	0.53	argininosuccinate synthase
SCO7209	-1.089	0.47	hypothetical protein
SCO7463	-0.654	0.64	sensor histidine kinase
SCO7464	-0.505	0.71	hypothetical protein
SCO7708	-0.911	0.53	hypothetical protein
SCO7793	-1.366	0.39	hypothetical protein

All these genes had a pfp value < 0.15. Hence, there is > 95 % probability that the difference in expression has not occurred by chance.

**Table A3: Genes that are upregulated in *S. coelicolor*  $\Delta$ *rbpA* strain, S101, relative to wild-type strain, J1915.**

Gene Name	Log <sub>2</sub> Average Cond1/Cond2	inv. Log	Description
SCO0049	3.595	12.08	hypothetical protein
SCO0168	1.006	2.00	regulator protein
SCO0169	0.975	1.96	hypothetical protein
SCO0190	0.787	1.73	Methyltransferase
SCO0201	0.864	1.82	integral membrane protein SCJ1213c
SCO0204	1.019	2.03	luxR family two-component response regulator
SCO0212	1.305	2.47	hypothetical protein
SCO0226	0.684	1.61	hypothetical protein
SCO0231	1.400	2.64	small hydrophobic hypothetical protein
SCO0409	1.204	2.30	spore-associated protein precursor
SCO0426	0.750	1.68	hypothetical protein
SCO0596	1.114	2.17	DNA-binding protein
SCO0600	0.712	1.64	RNA polymerase sigma factor sig8
SCO0608	0.921	1.89	regulatory protein
SCO0650	0.891	1.85	gas vesicle synthesis-like protein
SCO0678	2.111	4.32	hypothetical protein
SCO0694	2.270	4.83	hypothetical protein
SCO0736	1.739	3.34	secreted protein
SCO0761	1.171	2.25	hypothetical protein
SCO0779	0.743	1.67	hypothetical protein
SCO0931	0.780	1.72	secreted proline-rich protein
SCO0973	0.898	1.86	integral membrane protein
SCO0999	1.281	2.43	superoxide dismutase
SCO1023	0.981	1.97	hypothetical protein
SCO1236	1.268	2.41	urease gamma subunit
SCO1364	1.247	2.37	hypothetical protein
SCO1385	1.081	2.12	hypothetical protein
SCO1550	2.212	4.63	small membrane protein
SCO1603	1.120	2.17	Transposase
SCO1604	1.261	2.40	hypothetical protein
SCO1605	0.895	1.86	hypothetical protein
SCO1627	1.257	2.39	ATP-GTP binding protein
SCO1659	1.079	2.11	glycerol uptake facilitator protein
SCO1660	0.774	1.71	glycerol kinase
SCO1661	0.702	1.63	glycerol-3-phosphate dehydrogenase
SCO1674	1.059	2.08	secreted protein
SCO1675	2.867	7.30	small membrane protein
SCO1695	1.340	2.53	hypothetical protein
SCO1700	3.094	8.54	hypothetical protein
SCO1793	0.748	1.68	secreted protein

Gene Name	Log <sub>2</sub> Average Cond1/Cond2	inv. Log	Description
SCO1800	2.107	4.31	small secreted protein
SCO1839	0.925	1.89	transcriptional regulator
SCO1865	0.855	1.81	diaminobutyrate--2-oxoglutarate aminotransferase
SCO1866	2.129	4.38	L-ectoine synthase
SCO1867	1.536	2.90	Hydroxylase
SCO1875	1.147	2.22	secreted penicillin binding protein
SCO1993	0.728	1.66	hypothetical protein
SCO1995	1.136	2.19	hypothetical protein
SCO2113	1.160	2.24	bacterioferritin: iron storage protein
SCO2367	0.859	1.81	hypothetical protein
SCO2582	1.343	2.54	hypothetical protein
SCO2633	1.931	3.81	superoxide dismutase [Fe-Zn]
SCO2717	1.992	3.98	small membrane protein
SCO2718	0.787	1.72	secreted protein
SCO2760	0.741	1.67	membrane protein
SCO2807	0.955	1.94	integral membrane protein
SCO2828	0.718	1.64	probable amino acid ABC transporter protein, solute-binding component
SCO2912	1.119	2.17	hypothetical protein
SCO3009	0.767	1.70	hypothetical protein
SCO3034	1.336	2.53	sporulation regulatory protein
SCO3082	1.740	3.34	hypothetical protein
SCO3089	2.249	4.75	ABC transporter ATP-binding protein
SCO3090	1.326	2.51	ABC transporter integral membrane protein
SCO3098	0.975	1.97	secreted protein
SCO3110	0.710	1.64	ABC transport system integral membrane protein
SCO3111	1.552	2.93	ABC transport system ATP-binding protein
SCO3202	1.180	2.27	RNA polymerase sigma factor, hrdD
SCO3218	0.748	1.68	small conserved hypothetical protein
SCO3323	0.965	1.95	RNA polymerase sigma protein, bldN
SCO3324	1.023	2.03	hypothetical protein
SCO3396	0.897	1.86	hypothetical protein
SCO3397	0.738	1.67	integral membrane lysyl-tRNA synthetase
SCO3561	0.720	1.65	secreted protein
SCO3660	0.688	1.61	hypothetical protein
SCO3814	1.196	2.29	DNA-binding protein
SCO3855	1.259	2.39	hypothetical protein
SCO3857	0.740	1.67	regulatory protein
SCO3862	1.508	2.84	hypothetical protein
SCO3924	0.799	1.74	hypothetical protein
SCO3929	0.915	1.88	hypothetical protein
SCO3937	0.917	1.89	integrase /recombinase
SCO3945	1.646	3.13	cytochrome oxidase subunit I
SCO3946	1.119	2.17	cytochrome oxidase subunit II

Gene Name	Log <sub>2</sub> Average Cond1/Cond2	inv. Log	Description
SCO3956	2.978	7.88	ABC transporter ATP-binding protein
SCO3957	1.941	3.84	integral membrane protein
SCO3958	2.671	6.37	ABC transporter ATP-binding protein
SCO3992	0.754	1.68	hypothetical protein
SCO3993	1.002	2.00	hypothetical protein
SCO4004	1.652	3.14	small membrane protein
SCO4063	0.875	1.83	secreted protein
SCO4080	1.156	2.23	hypothetical protein
SCO4148	0.704	1.63	ABC transport system ATP-binding protein
SCO4174	3.140	8.81	integral membrane protein
SCO4187	1.357	2.56	hypothetical protein
SCO4198	0.758	1.69	DNA-binding protein
SCO4226	0.933	1.91	hypothetical protein
SCO4247	0.793	1.73	hypothetical protein
SCO4248	0.911	1.88	hypothetical protein
SCO4251	1.183	2.27	secreted protein
SCO4252	1.076	2.11	hypothetical protein
SCO4253	1.012	2.02	hypothetical protein
SCO4257	1.057	2.08	hydrolytic protein
SCO4258	1.188	2.28	hydrolytic protein
SCO4261	0.988	1.98	response regulator
SCO4277	0.802	1.74	tellurium resistance protein: responds to bacterial stress
SCO4317	2.312	4.97	hypothetical protein
SCO4347	1.590	3.01	hypothetical protein
SCO4350	0.964	1.95	Integrase
SCO4471	1.125	2.18	secreted protein
SCO4562	0.865	1.82	NADH dehydrogenase alpha subunit
SCO4584	1.481	2.79	hypothetical protein
SCO4585	0.823	1.77	ABC transporter ATP-binding protein
SCO4610	1.348	2.55	integral membrane protein
SCO4635	0.771	1.71	50S ribosomal protein L33
SCO4636	1.412	2.66	hypothetical protein
SCO4675	0.920	1.89	hypothetical protein
SCO4767	1.314	2.49	regulatory protein
SCO4807	0.984	1.98	hypothetical protein
SCO4861	1.163	2.24	hypothetical protein
SCO4934	0.901	1.87	Lipoprotein
SCO4994	1.582	3.00	hypothetical protein 2SCK3617
SCO5071	1.469	2.77	hydroxylacyl-CoA dehydrogenase
SCO5074	2.262	4.80	Dehydratase
SCO5075	1.253	2.38	Oxidoreductase
SCO5077	0.985	1.98	hypothetical protein
SCO5078	1.275	2.42	hypothetical protein



Gene Name	Log <sub>2</sub> Average Cond1/Cond2	inv. Log	Description
SCO5079	1.033	2.05	hypothetical protein
SCO5080	1.028	2.04	hydrolase: in actinorhodin cluster
SCO5084	1.127	2.18	hypothetical protein: in actinorhodin cluster
SCO5085	2.048	4.14	actinorhodin cluster activator protein, actIIORFIV
SCO5086	0.824	1.77	ketoacyl reductase: in actinorhodin cluster
SCO5088	1.473	2.78	actinorhodin polyketide beta-ketoacyl synthase beta subunit: in actinorhodin cluster
SCO5091	0.885	1.85	cyclase: in actinorhodin cluster
SCO5130	0.944	1.92	ABC transporter integral membrane protein
SCO5145	0.692	1.62	hypothetical protein
SCO5174	0.910	1.88	Transferase
SCO5198	0.699	1.62	hypothetical protein
SCO5204	1.205	2.31	integral membrane protein
SCO5496	0.741	1.67	small membrane hydrophobic protein
SCO5590	1.411	2.66	hypothetical protein
SCO5644	0.787	1.73	hypothetical protein
SCO5741	1.179	2.26	hypothetical protein
SCO5790	0.967	1.96	hypothetical protein
SCO5826	1.117	2.17	hypothetical protein
SCO5862	1.014	2.02	two-component regulator CutR
SCO5877	1.078	2.11	transcriptional regulator RedD: in RED cluster
SCO5885	1.503	2.83	hypothetical protein: in RED cluster
SCO5890	1.236	2.36	8-amino-7-oxononanoate synthase
SCO5897	1.210	2.31	Oxidase
SCO5898	1.328	2.51	probable membrane protein
SCO6073	0.894	1.86	Cyclase
SCO6239	1.117	2.17	sigma factor
SCO6282	0.946	1.93	3-oxoacyl-[acyl-carrier protein] reductase
SCO6377	0.932	1.91	Lipoprotein
SCO6380	0.944	1.92	hypothetical protein
SCO6382	0.783	1.72	secreted protein
SCO6394	1.276	2.42	IS element ATP binding protein
SCO6400	0.999	2.00	IS117 transposase
SCO6401	0.936	1.91	hypothetical protein
SCO6402	1.349	2.55	hypothetical protein
SCO6404	0.837	1.79	hypothetical protein
SCO6499	4.066	16.75	gas vesicle synthesis protein
SCO6500	0.784	1.72	gas vesicle synthesis-like protein
SCO6501	0.815	1.76	gas vesicle synthesis protein
SCO6502	1.603	3.04	gas vesicle synthesis protein
SCO6509	3.416	10.67	hydrophobic protein
SCO6682	0.947	1.93	hypothetical protein
SCO6708	0.873	1.83	hypothetical protein
SCO6766	0.801	1.74	hypothetical protein

Gene Name	Log <sub>2</sub> Average Cond1/Cond2	inv. Log	Description
SCO6776	1.140	2.20	hypothetical protein
SCO6814	2.113	4.33	ABC transporter ATP-binding protein
SCO6815	0.812	1.76	ABC transporter permease protein
SCO6816	1.574	2.98	ABC transporter binding lipoprotein
SCO6817	1.557	2.94	hypothetical protein
SCO6818	1.124	2.18	Phosphoglyceromutase
SCO6819	0.888	1.85	3-phosphoshikimate 1-carboxyvinyltransferase
SCO6820	1.123	2.18	Oxidoreductase
SCO6821	2.153	4.45	hypothetical protein
SCO6831	0.850	1.80	hypothetical protein
SCO6925	1.003	2.01	hypothetical protein (membrane protein)
SCO6931	0.997	1.99	hypothetical protein
SCO6932	2.696	6.48	hypothetical protein
SCO6935	1.090	2.13	hypothetical protein
SCO6939	0.832	1.78	hypothetical protein
SCO6941	0.768	1.70	hypothetical protein
SCO6978	0.767	1.70	carbohydrate kinase
SCO6983	0.877	1.84	hypothetical protein
SCO7018	0.922	1.89	hypothetical protein
SCO7092	0.870	1.83	secreted protein
SCO7108	0.819	1.77	hypothetical protein
SCO7210	1.078	2.11	hypothetical protein
SCO7278	0.797	1.74	RNA polymerase sigma factor
SCO7325	0.897	1.86	Stage II sporulation protein: belongs to anti-sigma factor antagonist
SCO7330	0.898	1.86	membrane protein
SCO7383	0.239	1.18	hypothetical protein
SCO7405	0.771	1.71	acetyltransferase (fragment)
SCO7536	1.025	2.04	integral membrane protein
SCO7646	1.829	3.55	hypothetical protein
SCO7657	1.242	2.37	secreted protein
SCO7658	1.429	2.69	hypothetical protein
SCO7756	1.715	3.28	hypothetical protein
SCO7781	0.681	1.60	pseudo gene
SCO7805	0.548	1.46	hypothetical protein
SCO7820	0.696	1.62	hypothetical protein

The genes had a pfp value < 0.15. Hence there is > 95% probability that the difference in expression has not occurred by chance.

```

Sig_Rx      1 -----
Sig_Af      1 -----
Sig_Cc      1 ---MAAASRTSTT-----
Sig_Ap      1 ---MAAKKANDEV-----
Sig_El      1 ---MAQASSKQAO-----
Sig_Se      1 -----
Sig_Ou      1 ---MTTNKAASDL-----
Sig_Mt      1 ---MAATKASTATDEPVKRTATKSP-----
Sig_Cd      1 MANAGEASSATVKKAAKKTAAKKTARK-----
Sig_Am      1 ---MAAAKTTOAAKADAEETPKATS-----
Sig_Sv      1 ---MDETATPETKSSGRKTAAKKAP-----
Sig_Kf      1 ---MSSSSSENLPAESARAV-----
Sig_Ac      1 ---MAQNGGKEISVAVAKTTKSKGVSEVEAPEVLAES-----
Sig_Sc      1 ---MSASTSRTLPEIAESVSVMALIERGKAEGQIAGDDVRRAFEADQIPATQWKVNLRS
Sig_Fa      1 ---MPRNLGRSR-----
HrdA        1 -----
HrdC        1 -----
HrdD        1 -----

```

```

Sig_Rx      1 -----MAAEDTMMLADRAYELLSKGRSQGFLSAEDVADLVREGDLSPEAEAE
Sig_Af      1 -----MR AVLGRGSAPH
Sig_Cc     11 -----EKVPGKD-----GSSQPDEEIQNQAASDSEDSVDLEDLEAPD
Sig_Ap     11 -----KLSDELTKIVDDLVS-SSSKSNGSISED DIQVAIRDIDVNDDELS
Sig_El     11 -----ERNNAAV-----ASIEAEDILEEDALDDEPDVVDAGDGLDDD
Sig_Se      1 -----MDIEEG-ALNEESLGDD---LAAEDGKD--
Sig_Ou     11 -----EFSDELQVVVTKLIRSASSSDSSITEDDIQVAIAEMDVDDDELS
Sig_Mt     23 -----AASASGAKTGAKRTAAKSASGSPPA-----KRATKPAARS
Sig_Cd     28 -----VARKVTAVKVAATPTVTSSAESATAEPAIELPSSEPVKKAAKKTAKKT
Sig_Am     23 -----ARKAPVRKPAAKTAAAGKTTARAPR-----KTAAKAAAGS
Sig_Sv     23 -----AKSAKTTTRKTTTKSDKAKTTTRS-----KASKSTSKAK
Sig_Kf     18 -----AATARSTAPRTSARVPDP-----AAKKATVKKAAPAAKKAAYS
Sig_Ac     36 -----LAEKPAKKTATKKKAPAKKAAEATDEVAEKPKRTRKTKATESTEATA
Sig_Sc     58 LNQILEEEGVTLMVSAAEPKRTRKSVAAKSPAKRTATKAVAAKPVTSRKATAPAAPAAPA
Sig_Fa     10 -----EGSDEIP-----PMSPTVLPREAQVDEVKDLITRGKEIG-
HrdA        1 -----VRGGQRRASRLRPPTYRRRPPPAASILEVAPVQ
HrdC        1 -----MAPTARTPTAR
HrdD        1 -----MATRAVAR

```

```

Sig_Rx     48 LY-----ATLEEDITLVDDELEG
Sig_Af     12 RF-----VHPASQSSGGGE-----
Sig_Cc     48 TVG-----DDAIAGDKPEAESEEDLLEG
Sig_Ap     55 LYD-----SLRVKGIEITTASEATVSFAIEDDAEDDFSSDDDLN
Sig_El     48 -----KLESPLSDSDDEDLLEG
Sig_Se     25 -----VDANLSD-----EELLEG
Sig_Ou     56 IYE-----AVRARGVSIVASGESQPVSAVDDDEDGLDALEADG-KAA
Sig_Mt     58 VKPASAPQDTTSTIPKRKTRAAAKSAAAKAPSARGHATKPRAPKDAQHEAATDPEDALD
Sig_Cd     76 AK-----KVAKKAAKKTAKKTAKKSTRKSARKVAAPVPEEATQSSESESELE
Sig_Am     58 PK-----AKKAGEGDEP-----EETEGAPGDEDLVD
Sig_Sv     58 SK-----KSAELDEG-----FDSPELEDVDLSDLDVD
Sig_Kf     56 TK-----EDGIPAKKAAAKKAP-----AKKAVSEAGLAIDPKTG
Sig_Ac     84 EA-----TEEKPKKKRAPRKKK-----AEEEVASETPAPEADAE
Sig_Sc    118 TEP-----AAVEEEAPAKKAAAKKTT-----AKKATAKKTAKKAAAK
Sig_Fa     44 -----FLTTEDTVTAIQAAELPPE
HrdA       34 TQ-----TLTQTDTAAGGAEPDAE
HrdC       12 -----TRDDRRATTTRTAR-----
HrdD        9 -----RKSAAGETSGSAT-----

```

```

Sig_Rx 67 AAEVTEPER-----
Sig_Af 27 -----
Sig_Cc 76 IPE-----EELKAS--SSVSLP-----
Sig_Ap 96 SSDFEDEDDEDDFDVVNEAREVKEALRSVS-----
Sig_El 66 IPE-----EELKAT--VEVQLP-----
Sig_Se 38 IPE-----NELKAASTSEINLP-----
Sig_Ou 96 LDDEDDEDDDEISKARKEARVANEVLRSV-----
Sig_Mt 118 SVEELDAEPDLDVEPGEDLDLDAADLNLDDLEDDVAPDADDDLDSGDDEDHEDLEAEAAAV
Sig_Cd 125 SLAGEDSDEDEYDPHLDEFDDDDE-----
Sig_Am 84 EVEELIDEP--VEEEPAE-----
Sig_Sv 85 AVEVDVVDATVTEDTDES-----
Sig_Kf 90 KPRALDE-----VDEADFP-----
Sig_Ac 118 EIEDRHDELAEPDEFADEEDLDLDLVD-----
Sig_Sc 156 KTTAKKEDGELLEDEATEEP-----
Sig_Fa 63 QAET-----VLQVLNDEGIEVLEAG-----
HrdA 53 RGVLLAMPAQPGAGAALPH-----
HrdC 25 -----
HrdD 22 -----

```

```

Sig_Rx 75 -----AAGREGDNPASQLAHAV
Sig_Af 27 -----LVTKERQER
Sig_Cc 91 -----KVTASKSRIRQSSRSRAADTSITM
Sig_Ap 125 -----KAKVSKPKRSSRARARRSDTTTVM
Sig_El 81 -----KVAGK-SKVRSVRKRNADAS-VTM
Sig_Se 55 -----KLSS--SRASKVRRRSSDGSSAVM
Sig_Ou 125 -----KPKGRK--RSSRARAHKLDSSAVM
Sig_Mt 178 APGQTADDDEEIAEPTEKD-----KASGDFVWDEDESEALRQARKDAELT
Sig_Cd 148 --MHTDELGEDDSSDDSD-----EDEGSSVWDEDESAALRQARKDAELS
Sig_Am 99 -----EENAKPGE-----GDFVWDEEESALRQARKDAELT
Sig_Sv 103 -----DEESDAND-----PDFVWDEEESALRQARKDAELT
Sig_Kf 105 ---DAVKPLEKELTEDQ-----GFIVSEADEDEP--EQQVMVAG
Sig_Ac 147 --DDLEEDLESDDDEDSEGARRGKPDAAETQAVVKTGGFVVSEFDDDEP--VQKVTVAG
Sig_Sc 176 ---KAAT-EEPEGTENA-----GFVLSDEDEDDAP--AQQVAAAG
Sig_Fa 83 -----GENPDEADLLARRRRREEEELALKA
HrdA 73 -----GAPVDVPEHPEPPPPPTRTESGG
HrdC 25 -----LRTRIPEPD
HrdD 22 -----SVRANGGEL

```

```

Sig_Rx 92 ATGDSIRMYLAEIGRVLLLTHADEIRLAKAISRG-----
Sig_Af 36 EDEDLVRLYLGDIGSHELLTKDDEAELGRRMDR--ARDAQRQLDAG--VKEPAR--
Sig_Cc 115 LTGDPVRMYLKEIGKVLLLTAAEEIDLAMKIEAGVEAGEOLEAAEADN----AEELSRR
Sig_Ap 149 LTGDPVRMYLKEIGKVLLLTASEEVHLAMKIEAGTAASOKIEDFEDG----LIELNRA
Sig_El 103 LTGDPVRMYLKEIGKVLLLTAAEEIDLAMKIEAGVAAMEEIEKAEDE----GIELERR
Sig_Se 77 LTGDPVRMYLKEIGKVLLLTAAEEIDLAMKIEAGVAATAELERAEDE----GIELNRR
Sig_Ou 147 LTGDPVRMYLKEIGKVLLLTASEEVHLAMKIEAGTEATERLEAADRG----EIELTRA
Sig_Mt 223 ASADSVRAYLKQIGKVALLNAEEVEELAKRIEAGLYATQIMTELSE--RGEKLPAA
Sig_Cd 191 ASADSVRAYLKQIGKVALLNAEEVEELAKRIEAGLYATYRMEQMEEAFNNGDKEAKLTPA
Sig_Am 130 ASADSVRAYLKQIGKVALLNAEEVEELAKRIEAGLYAERIRRAED-----ESEKLTPQ
Sig_Sv 134 ASADSVRAYLKQIGKVALLNAEEVEELAKRIEAGLYAERIRRAEE-----EGEKLTSTQ
Sig_Kf 140 ATADPVRDYLKQIGKVLLLNAEQEVELAKRIEAGLEAEEQLSDESA-----KLKDK
Sig_Ac 203 ATADPVRDYLKQIGKVSLLTAADEVDLARRIEAGLYAEHKIKTEGD-----ELPSK
Sig_Sc 210 ATADPVRDYLKQIGKVLLLNAEQEVELAKRIEAGLEAEDKLAN-SD-----KLAPK
Sig_Fa 107 PTSDPVRMYLKEIGKVLLLTAAEEVDLAKRIEAGLEAEEKLAVATKK-----TSPQ
HrdA 95 PSSDLFRQYLREIGRIPLLSAAEEVDLARRVEAGLEAEEKLRCSPGL-----DDREAL-
HrdC 34 EEPDLTGQYLTQIGATPLLLTAADEVRLATRIEAGVRAREELETADTG-----EPAPTPR
HrdD 31 ADRLVGMYLDEIARTPLLDAAKEVELSQTIEAGVEARQVLEGYEET-----GADAT--

```

Sig\_Rx 126 -----CKRSKDKLVEANLRLVVSIAKKYRNRCVSFLDLIQEGNIGLIRAAEKFD  
 Sig\_Af 86 -RAEIEELVRQEEAKRAFIQANLRLVVSIAKRYQASGLPLDLIQEGNMGLIHAVEKFD  
 Sig\_Cc 170 EHRRILTRIEAVGLDAKQOLIEANLRLVVSIAKRYVGRGMLFLDLIQEGNLGLIRAVEKFD  
 Sig\_Ap 203 EQRRIMRIEISVGLDAKQOLISANLRLVVSIAKRYVGRGMLFLDLIQEGNLGLIRAVEKFD  
 Sig\_El 157 EKRRILGRIEQVGIDAKQOLIEANLRLVVSIAKRYVGRGMLFLDLIQEGNLGLIRAVEKFD  
 Sig\_Se 131 DKRRILTRIEAVGLDAKQOLIEANLRLVVSIAKRYVGRGMLFLDLIQEGNLGLIRAVEKFD  
 Sig\_Ou 201 EQRRIMRIEISVGLDAKQOLISANLRLVVSIAKRYVGRGMLFLDLIQEGNLGLIRAVEKFD  
 Sig\_Mt 277 QRRDMWICRDGDRAKNHLLLEANLRLVVSIAKRYTGRGMAFLDLIQEGNLGLIRAVEKFD  
 Sig\_Cd 251 VKRRILRAIARDGRKAKNHLLLEANLRLVVSIAKRYTGRGMAFLDLIQEGNLGLIRAVEKFD  
 Sig\_Am 184 MRRDLRWIVRDGERAKNHLLLEANLRLVVSIAKRYTGRGMAFLDLIQEGNLGLIRAVEKFD  
 Sig\_Sv 188 MRRDLKWIIVRDGERAKSHLLLEANLRLVVSIAKRYTGRGMAFLDLIQEGNLGLIRAVEKFD  
 Sig\_Kf 191 VRDEYDWIISEDGRRAKNHLLLEANLRLVVSIAKRYTGRGMLFLDLIQEGNLGLIRAVEKFD  
 Sig\_Ac 254 LRREIQIIVHDGQLSKNHLLLEANLRLVVSIAKRYTGRGMLFLDLIQEGNLGLIRAVEKFD  
 Sig\_Sc 260 LKRELEITIAEDGRRAKNHLLLEANLRLVVSIAKRYTGRGMLFLDLIQEGNLGLIRAVEKFD  
 Sig\_Fa 158 MRRDEATERDGOIAKRKLVEANLRLVVSIAKRYVGRGMLFLDLIQEGNLGLIRAVEKFD  
 HrdA 148 ---DLDRIVVLGRLAKRRLIEANLRLVVSIAKRYVGRGMLFLDLIQEGNLGLIRAVEKFD  
 HrdC 88 RRRITLEETVHDGOEAKDHMVRANLRLVVSIAKRYVGRGMLFLDLIQEGNLGLIRAVEKFD  
 HrdD 83 -REELQATIDESERAKDVFIIRSNLRLVVSIAKRYVGRGMLFLDLIQEGNLGLIRAVEKFD

Sig\_Rx 175 YRKGFKEFSTYATWWIRQAITRAIADQARTIRIPVHMVEKVNKYHRTQRMVQALGREPTD  
 Sig\_Af 145 WRKGFKEFSTYATWWIRQAITRGIANTRTIRLPVHAGDSLSRVQRAQVQLEVRILGRAPTIV  
 Sig\_Cc 230 YTKGFKEFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVRIQRQLLOELGREPTP  
 Sig\_Ap 263 YTKGFKEFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVRIQRQLLOELGREPTP  
 Sig\_El 217 YTKGFKEFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVRIQRQLLOELGREPTP  
 Sig\_Se 191 YTKGFKEFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVRIQRQLLOELGREPTP  
 Sig\_Ou 261 YTKGFKEFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVRIQRQLLOELGREPTP  
 Sig\_Mt 337 YTKGFKEFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLGRIQRELLQDLGREPTP  
 Sig\_Cd 311 YTKGFKEFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLGRIQRELLQDLGREPTP  
 Sig\_Am 244 YTKGFKEFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLGRIQRELLQDLGREPTP  
 Sig\_Sv 248 YTKGFKEFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLGRIQRELLQDLGREPTP  
 Sig\_Kf 251 YTKGFKEFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLARVQRQLLOELGREPTP  
 Sig\_Ac 314 YTKGFKEFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLARVQRQLLOELGREPTP  
 Sig\_Sc 320 YTKGFKEFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLARVQRQLLOELGREPTP  
 Sig\_Fa 218 YTKGFKEFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVRIQRQLLOELGREPTP  
 HrdA 205 YARGYKEFSTYATWWIRQAIMSALADQARTIRIPVHVVELINRVVRVQRRLQERCCEPTP  
 HrdC 148 YTKGFKEFSTYATWWIRQAITERGLATHARTVRLPVHVVELOKLAKVERKIRAGLDREPTT  
 HrdD 142 YRKGFKEFSTYATWWIRQAITRSIADQRTIRLPVHLVEELGRIRRVQREENREHGREPTP

Sig\_Rx 235 EEVAEEIGVPVEEVIRLQEIORSISLETPVGDESSSLGDFLEDASAVTETDAVSESLLL  
 Sig\_Af 205 RELAAEVDLPYDKLVEALRFRAEPVSLSEPLRDDGDAELGDVVEDSQSATPFDVAAEHLL  
 Sig\_Cc 290 EEIAEEMDLTPDRVREIQKISQEPVSLETPIGEEEDSQLGDFIEDDAVVPPDAASFMSL  
 Sig\_Ap 323 EEIGSEMGMSPDRVREIQKISQEPVSLETPIGEEEDSQLGDFIEDSTAVAPPEAASDSML  
 Sig\_El 277 EEIGKEMCLPAERVREIQKISQEPVSLETPIGEEEDSQLGDFIEDDAVVPPDAASFMSL  
 Sig\_Se 251 EEIGKEMCLSAERVREIQKISQEPVSLETPIGEEEDSQLGDFIEDDAVVPPDAASFMSL  
 Sig\_Ou 321 EEIGKEMGISADRVREIQKISQEPVSLETPIGEEEDSQLGDFIEDSSAVAPPEAASDSML  
 Sig\_Mt 397 EELAKEMDITPEKVLEIQYAREPISLDQTIGDEGDSQLGDFIEDSEAVVAVDVAVSFILL  
 Sig\_Cd 371 EELAKEMDITPEKVLEIQYAREPISLDQTIGDEGDSQLGDFIEDSEAVVAVDVAVSFILL  
 Sig\_Am 304 EELAKEMDITPEKVLEIQYAREPISLDQTIGDEGDSQLGDFIEDSEAVVAVDVAVSFILL  
 Sig\_Sv 308 EELAKEMDISPEKVLEIQYAREPISLDQTIGDEGDSQLGDFIEDSEAVVAVDVAVSFILL  
 Sig\_Kf 311 EELAKEIDMTPEKVLEVQYGREPISLHTPLGEDGDSSEFGLIEDSEAVVPAEAVSFILL  
 Sig\_Ac 374 EELAKEIDMTPEKVLEVQYGREPISLHTPLGEDGDSSEFGLIEDSEAVVPTDAVAVSFILL  
 Sig\_Sc 380 EELAKEIDMTPEKVLEVQYGREPISLHTPLGEDGDSSEFGLIEDSEAVVPADAVAVSFILL  
 Sig\_Fa 278 EELAKEMDLTPDKVREILKVSQEPVSLETPIGEEEDSQLGDFIEDDAVVPPDAASFILL  
 HrdA 265 QEVAEHLDLAPERVGEVRLAQEPVSLHAPVGEEDDVALGDLIEDDAASEVESAAFLILL  
 HrdC 208 EEVAEESGIDVDKVVWRREVGRDAVSLDTPVDETGDVVGDLIPDTEVIRAEVAEEQAL  
 HrdD 202 AEIAAEELGSTPERVTDVLDWARDPVSLNMSVDDEGETQEGDLIEDTSVAVSEEQSVLTLIR

Sig_Rx	295	KLHLREALDELPERERQIIEMRFGMKDDRPRTLEEVGREEDITRERVRQIQMKTINLLRE
Sig_Af	265	PSEVSRLLAPLDGREERELRLRYGLDRCEPRTLEEVGGEYHILTRERIRQIEARMSKLRH
Sig_Cc	350	QEQLTKVLDGLAERERKVIISLRFGLEDGHPRTLEEVGREFGVTRERIRQIESKTLAKLRH
Sig_Ap	383	REQLEQVLDGLADREKVIKFRFGLEDGHPRTLEEVGREFGVTRERIRQIESKTLAKLRH
Sig_El	337	QEQLSKVLDGLAERERKVIISLRFGLEDGHPRTLEEVGREFGVTRERIRQIESKTLAKLRH
Sig_Se	311	QEQLSKVLDGLAERERKVIISLRFGLEDGHPRTLEEVGREFGVTRERIRQIESKTLAKLRH
Sig_Ou	381	REQLEQVLDGLADREKVIKFRFGLEDGHPRTLEEVGREFGVTRERIRQIESKTLAKLRH
Sig_Mt	457	QDQLQSVLDITLSEREAGVVRLRFGLTDGQPTLDEIGQVYGVTRERIRQIESKTM SKLRH
Sig_Cd	431	QDQLQSVLDITLSEREAGVVRLRFGLTDGQPTLDEIGQVYGVTRERIRQIESKTM SKLRH
Sig_Am	364	QDQLQSVLDITLSEREAGVVRLRFGLTDGQPTLDEIGQVYGVTRERIRQIESKTM SKLRH
Sig_Sv	368	QDQLQSVLDITLSEREAGVVRLRFGLTDGQPTLDEIGQVYGVTRERIRQIESKTM SKLRH
Sig_Kf	371	QEQLHAVLDITLSEREAGVVS MRFGLEDGQPKTLDEIGKVIYGVTRERIRQIESKTM SKLRH
Sig_Ac	434	QEQLHKVLDITLSEREAGVVS MRFGLEDGQPKTLDEIGKVIYGVTRERIRQIESKTM SKLRH
Sig_Sc	440	QEQLHSVLDITLSEREAGVVS MRFGLEDGQPKTLDEIGKVIYGVTRERIRQIESKTM SKLRH
Sig_Fa	338	QEQLDSVLDITLSEREAGVVS MRFGLEDGQPKTLDEIGKVIYGVTRERIRQIESKTM SKLRH
HrdA	325	RQHLEAVLSTLGERERKVVQLRYGLADGRPRTLEEIGRIFGVTRERIRQIESKTL SKLRD
HrdC	268	AAELREAVGTLAPRESLILSLRYGLDGRPRTLQOVAQHVGILTRERVRQIEKESLAHLRA
HrdD	262	SEETDGLIGRIDPRTASTIKMRYGDDGRERTITEVGKEHGLTRERIRQIEKHAILELAK
Sig_Rx	355	QRRTONLREYILH-
Sig_Af	325	PSADIGARDLLITG
Sig_Cc	410	PSRSKLLDYLED
Sig_Ap	443	PSRSGRLKDYME
Sig_El	397	PSRSKLLDYLED
Sig_Se	371	PSRSKLLDYLED
Sig_Ou	441	PSRSGRLKDYME
Sig_Mt	517	PSRSQVLRDYLD-
Sig_Cd	491	PSRSQVLRDYLD-
Sig_Am	424	PSRSQVLRDYLD-
Sig_Sv	428	PSRSQVLRDYLD-
Sig_Kf	431	PSRSQVLRDYLD-
Sig_Ac	494	PSRSQVLRDYLD-
Sig_Sc	500	PSRSQVLRDYLD-
Sig_Fa	398	PSRSQVLRDYLD-
HrdA	385	HAYADQLRCGYLD-
HrdC	328	ENRERLLD WAS-
HrdD	322	LA-RDTGFEAAA-

**Figure A2: Multiple sequence alignment of  $\sigma^{\text{HrdB}}$  between RbpA and Non-RbpA containing Actinobacteria together with  $\sigma^{\text{HrdA}}$ ,  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$ .**

The Clustal W2 multiple sequence alignment tool and the BOXSHADE 3.21 server were used. Amino acid residues shown in white font with black highlight showed identical residues whilst the amino acid residues in white font with grey highlight showed similar residues. The homology was tested to the 0.5 fraction that must agree for shading. The first seven  $\sigma^{\text{HrdB}}$  amino acid sequences are from strains lacking RbpA: Sig\_Rx: *Rubrobacter xylanophilus* DSM 9941, Sig\_Af: *Acidimicrobium ferrooxidans* ICP, DSM 10331, Sig\_Cc: *Cryptobacterium curtum* 12-3, DSM 15641, Sig\_Ap: *Atopobium parvulum* IPP 1246, DSM 20469, Sig\_El: *Eggerthella lenta* VPI 0255, DSM 2243, Sig\_Se: *Slackia exigua* ATCC 700122 and Sig\_Ou: *Olsenella uli* VPI, DSM 7084. The following eight  $\sigma^{\text{HrdB}}$  sequences are from strains containing RbpA: Sig\_Mt: *Mycobacterium tuberculosis* H37Rv (lab strain), Sig\_Cd: *Corynebacterium diphtheriae* gravis NCTC 13129, Sig\_Am: *Actinosynnema mirum*

DSM 43827, Sig\_Sv: *Saccharomonospora viridis* DSM 43017, Sig\_Kf: *Kribbella flavida* DSM 17836, Sig\_Ac: *Actinomyces coleocanis* DSM 15436 contig00038, Sig\_Sc: *Streptomyces coelicolor* A3 (2), and Sig\_Fa: *Frankia alni* ACN14a. The sequences of  $\sigma^{\text{HrdA}}$ ,  $\sigma^{\text{HrdC}}$  and  $\sigma^{\text{HrdD}}$  of *Streptomyces coelicolor* A3 (2) were also aligned.

```

RbpA    1  MSERALRGTRLVVTSYETDRGIDLAPRQAVEYACEKGIRTEMPFSVEAEIPPEWECKVCG
Cr1     1  MTLPSGHPKSRLIKKET-----ALGP-YIREGQCEDNREFFDCLAVCVNPKPAPEKREFW

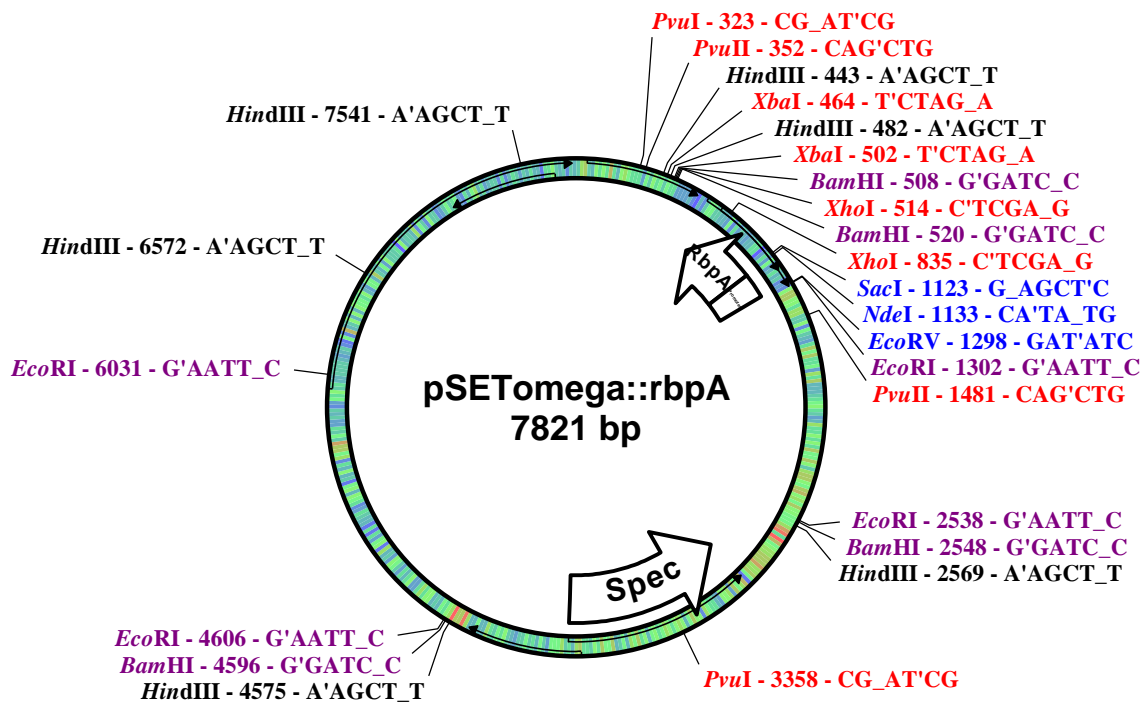
RbpA    61  AQAILLVGDGDP-----EEKKAKPARTHWDMLE--MERRTRELEEVLEERLAVLRSG--AM
Cr1     55  GWWMELEAQEKRFYRYQFGLFDKEGNWTVVPINTEVVERLEYTLREFHEKLRDLLISM

RbpA    112 NIAVHPRDSRKSA-----
Cr1     115 ELALEPSDDFNDEPVKLSA

```

**Figure A3: Alignment of RbpA and Cr1 amino acid sequences.**

The amino acids in white fonts and black highlights show identical residues whilst amino acids in white font and grey highlights show similar residues. The homology was tested to the 0.6 fraction that must agree for shading. This shows some conserved residues between RbpA and Cr1.



**Figure A4: The vector map of pSETΩ::rbpA.**

Spec; Spectinomycin resistance gene, RbpA; gene of interest including its promoter region, restriction enzyme sites mapped around the vector shown in black, red, purple and blue. The vector map was generated using pDRAW32 software.



## References

**Bioinformatics:**

BOXSHADE 3.21 server - [http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)

ExPASy-ProtParam tool - <http://web.expasy.org/protparam/>

StrepDB - <http://strepdb.streptomyces.org.uk>

BLAST SEARCH - <http://img.jgi.doe.gov/cgi-bin/w/main.cgi?section=FindGenesBlast&page=geneSearchBlast>

Clustal W Multiple Alignment Tool - <http://www.ebi.ac.uk/Tools/msa/clustalw2/>

BOXSHADE 3.21 server - [http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)

PSIPRED Secondary Structure Prediction - <http://bioinf.cs.ucl.ac.uk/psipred/>

Phyre (Protein Homology/analogy Recognition Engine) -

<http://www.sbg.bio.ic.ac.uk/phyre/>

Helical Wheel Representation -

<http://cti.itc.virginia.edu/~cmg/Demo/wheel/wheelApp.html>

UnitProTKB - <http://www.uniprot.org/uniprot>

Venn diagram - <http://www.pangloss.com/seidel/Protocols/venn.cgi>

ImageJ Gel band quantification obtained from - <http://rsbweb.nih.gov/ij/index.html>

PyMOL Software: [http://download.cnet.com/PyMOL/3000-2054\\_4-10914845.html](http://download.cnet.com/PyMOL/3000-2054_4-10914845.html)

PyMOL and cealign alignment: (Shindyalov & Bourne, 1998) (Jia *et al.*, 2004).

DALI: [http://ekhidna.biocenter.helsinki.fi/dali\\_server/start](http://ekhidna.biocenter.helsinki.fi/dali_server/start)

- Alting-Mees, M. A. & J. M. Short, (1989) pBluescript II: gene mapping vectors. *Nucleic Acids Res* **17**: 9494.
- Andreini, C., I. Bertini & G. Cavallaro, (2011) Minimal Functional Sites Allow a Classification of Zinc Sites in Proteins. *PLoS ONE* **6**: e26325.
- Artsimovitch, I., V. Patlan, S. Sekine, M. N. Vassilyeva, T. Hosaka, K. Ochi, S. Yokoyama & D. G. Vassilyev, (2004) Structural basis for transcription regulation by alarmone ppGpp. *Cell* **117**: 299-310.
- Artsimovitch, I., M. N. Vassilyeva, D. Svetlov, V. Svetlov, A. Perederina, N. Igarashi, N. Matsugaki, S. Wakatsuki, T. H. Tahirov & D. G. Vassilyev, (2005) Allosteric modulation of the RNA polymerase catalytic reaction is an essential component of transcription control by rifamycins. *Cell* **122**: 351-363.
- Babcock, M. J., M. J. Buttner, C. H. Keler, B. R. Clarke, R. A. Morris, C. G. Lewis & M. E. Brawner, (1997) Characterization of the rpoC gene of *Streptomyces coelicolor* A3(2) and its use to develop a simple and rapid method for the purification of RNA polymerase. *Gene* **196**: 31-42.
- Baldwin, N. E. & A. J. Dombroski, (2001) Isolation and characterization of mutations in region 1.2 of *Escherichia coli*  $\sigma$ 70. *Molecular Microbiology* **42**: 427-437.
- Bang, I.-S., J. G. Frye, M. McClelland, J. Velayudhan & F. C. Fang, (2005) Alternative sigma factor interactions in *Salmonella*:  $\sigma$ E and  $\sigma$ H promote antioxidant defences by enhancing  $\sigma$ S levels. *Molecular Microbiology* **56**: 811-823.
- Barne, K. A., J. A. Bown, S. J. W. Busby & S. D. Minchin, (1997) Region 2.5 of the *Escherichia coli* RNA polymerase [sigma]70 subunit is responsible for the recognition of the 'extended -10' motif at promoters. *EMBO J* **16**: 4034-4040.
- Belogurov, G. A., R. A. Mooney, V. Svetlov, R. Landick & I. Artsimovitch, (2009) Functional specialization of transcription elongation factors. *Embo J* **28**: 112-122.
- Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill & D. A. Hopwood, (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141-147.
- Bibb, M. J., V. Molle & M. J. Buttner, (2000)  $\zeta$ BldN, an Extracytoplasmic Function RNA Polymerase Sigma Factor Required for Aerial Mycelium Formation in *Streptomyces coelicolor* A3(2). *Journal of Bacteriology* **182**: 4606-4616.
- Blaby-Haas, C. E., R. Furman, D. A. Rodionov, I. Artsimovitch & V. de Crécy-Lagard, (2011) Role of a Zn-independent DksA in Zn homeostasis and stringent response. *Molecular Microbiology* **79**: 700-715.
- Blatter, E. E., W. Ross, H. Tang, R. L. Gourse & R. H. Ebricht, (1994) Domain organization of RNA polymerase  $\alpha$  subunit: C-terminal 85 amino acids

- constitute a domain capable of dimerization and DNA binding. *Cell* **78**: 889-896.
- Breitling, R., P. Armengaud, A. Amtmann & P. Herzyk, (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Letters* **573**: 83-92.
- Busby, S. & R. H. Ebright, (1999) Transcription activation by catabolite activator protein (CAP). *Journal of Molecular Biology* **293**: 199-213.
- Buttner, M. J., K. F. Chater & M. J. Bibb, (1990) Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3(2). *J Bacteriol* **172**: 3367-3378.
- Campbell, E. A. & S. A. Darst, (2000) The anti-sigma factor SpoIIAB forms a 2:1 complex with sigma(F), contacting multiple conserved regions of the sigma factor. *J Mol Biol* **300**: 17-28.
- Campbell, E. A., R. Greenwell, J. R. Anthony, S. Wang, L. Lim, K. Das, H. J. Sofia, T. J. Donohue & S. A. Darst, (2007) A Conserved Structural Module Regulates Transcriptional Responses to Diverse Stress Signals in Bacteria. *Molecular Cell* **27**: 793-805.
- Campbell, E. A., N. Korzheva, A. Mustaev, K. Murakami, S. Nair, A. Goldfarb & S. A. Darst, (2001) Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* **104**: 901-912.
- Campbell, E. A., O. Muzzin, M. Chlenov, J. L. Sun, C. A. Olson, O. Weinman, M. L. Trester-Zedlitz & S. A. Darst, (2002) Structure of the Bacterial RNA Polymerase Promoter Specificity  $\sigma$  Subunit. *Molecular Cell* **9**: 527-539.
- Campbell, E. A., L. F. Westblade & S. A. Darst, (2008) Regulation of bacterial RNA polymerase [sigma] factor activity: a structural perspective. *Current Opinion in Microbiology* **11**: 121-127.
- Canals, A. & M. Coll, (2009) Cloning, expression, purification and crystallization of the Rho transcription termination factor from *Thermotoga maritima*. *Protein Expr Purif* **65**: 174-178.
- Cavanagh, A. T., P. Chandrangsu & K. M. Wassarman, (2010) 6S RNA regulation of relA alters ppGpp levels in early stationary phase. *Microbiology* **156**: 3791-3800.
- Cavanagh, A. T., A. D. Klocko, X. Liu & K. M. Wassarman, (2008) Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of sigma70. *Mol Microbiol* **67**: 1242-1256.
- Chadsey, M. S. & K. T. Hughes, (2001) A multipartite interaction between *Salmonella* transcription factor sigma28 and its anti-sigma factor FlgM: implications for sigma28 holoenzyme destabilization through stepwise binding. *J Mol Biol* **306**: 915-929.
- Chadsey, M. S., J. E. Karlinsey & K. T. Hughes, (1998) The flagellar anti-sigma factor FlgM actively dissociates *Salmonella typhimurium* sigma28 RNA polymerase holoenzyme. *Genes Dev* **12**: 3123-3136.

- Chandrangsu, P., J. J. Lemke & R. L. Gourse, (2011) The *dksA* promoter is negatively feedback regulated by DksA and ppGpp. *Mol Microbiol* **80**: 1337-1348.
- Chater, K. F., (2001) Regulation of sporulation in *Streptomyces coelicolor* A3(2): a checkpoint multiplex? *Current Opinion in Microbiology* **4**: 667-673.
- Chater, K. F., C. J. Bruton, K. A. Plaskitt, M. J. Buttner, C. Méndez & J. D. Helmann, (1989) The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility  $\sigma$  factor of *B. subtilis*. *Cell* **59**: 133-143.
- Cho, Y. H., E. J. Lee, B. E. Ahn & J. H. Roe, (2001) SigB, an RNA polymerase sigma factor required for osmoprotection and proper differentiation of *Streptomyces coelicolor*. *Mol Microbiol* **42**: 205-214.
- Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead & B. G. Barrell, (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.
- Cole, S. T., K. Eiglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward & B. G. Barrell, (2001) Massive gene decay in the leprosy bacillus. *Nature* **409**: 1007-1011.
- Copeland, A., J. Sikorski, A. Lapidus, M. Nolan, T. G. Del Rio, S. Lucas, F. Chen, H. Tice, S. Pitluck, J. F. Cheng, R. Pukall, O. Chertkov, T. Brettin, C. Han, J. C. Detter, C. Kuske, D. Bruce, L. Goodwin, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Chen, K. Palaniappan, P. Chain, M. Rohde, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides & H. P. Klenk, (2009) Complete genome sequence of *Atopobium parvulum* type strain (IPP 1246). *Stand Genomic Sci* **1**: 166-173.
- Crack, J. C., C. D. den Hengst, P. Jakimowicz, S. Subramanian, M. K. Johnson, M. J. Buttner, A. J. Thomson & N. E. Le Brun, (2009) Characterization of [4Fe-4S]-Containing and Cluster-Free Forms of *Streptomyces* WhiD. *Biochemistry-U* **48**: 12252-12264.
- Dalton, K. A., A. Thibessard, J. I. B. Hunter & G. H. Kelemen, (2007) A novel compartment, the 'subapical stem' of the aerial hyphae, is the location of a sigN-dependent, developmentally distinct transcription in *Streptomyces coelicolor*. *Molecular Microbiology* **64**: 719-737.

- Darst, S. A., E. W. Kubalek & R. D. Kornberg, (1989) Three-dimensional structure of *Escherichia coli* RNA polymerase holoenzyme determined by electron crystallography. *Nature* **340**: 730-732.
- Davies, K. M., A. J. Dedman, S. Van Horck & P. J. Lewis, (2005) The NusA:RNA polymerase ratio is increased at sites of rRNA synthesis in *Bacillus subtilis*. *Molecular Microbiology* **57**: 366-379.
- Dey, A., A. K. Verma & D. Chatterji, (2010) Role of an RNA polymerase interacting protein, MsRbpA, from *Mycobacterium smegmatis* in phenotypic tolerance to rifampicin. *Microbiology* **156**: 873-883.
- Dombroski, A. J., (1996) Sigma factors: purification and DNA binding. *Methods Enzymol* **273**: 134-144.
- Dove, S. L., S. A. Darst & A. Hochschild, (2003) Region 4 of sigma as a target for transcription regulation. *Mol Microbiol* **48**: 863-874.
- Dove, S. L., F. W. Huang & A. Hochschild, (2000) Mechanism for a transcriptional activator that works at the isomerization step. *Proceedings of the National Academy of Sciences* **97**: 13215-13220.
- England, P., L. F. Westblade, G. Karimova, V. Robbe-Saule, F. Norel & A. Kolb, (2008) Binding of the unorthodox transcription activator, Crl, to the components of the transcription machinery. *J Biol Chem* **283**: 33455-33464.
- Feklistov, A. & S. A. Darst, (2011) Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell* **147**: 1257-1269.
- Feklistov, A., V. Mekler, Q. Jiang, L. F. Westblade, H. Irschik, R. Jansen, A. Mustaev, S. A. Darst & R. H. Ebright, (2008) Rifamycins do not function by allosteric modulation of binding of Mg<sup>2+</sup> to the RNA polymerase active center. *Proceedings of the National Academy of Sciences* **105**: 14820-14825.
- Fisher, M. T., (2006) Proline to the rescue. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 13265-13266.
- Flardh, K. & M. J. Buttner, (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol* **7**: 36-49.
- Flatten, I., Morigen & K. Skarstad, (2009) DnaA protein interacts with RNA polymerase and partially protects it from the effect of rifampicin. *Mol Microbiol* **71**: 1018-1030.
- Forti, F., V. Mauri, G. Deho & D. Ghisotti, (2011) Isolation of conditional expression mutants in *Mycobacterium tuberculosis* by transposon mutagenesis. *Tuberculosis (Edinb)* **91**: 569-578.
- Fujii, T., H. C. Gramajo, E. Takano & M. J. Bibb, (1996) redD and actII-ORF4, pathway-specific regulatory genes for antibiotic production in *Streptomyces coelicolor* A3(2), are transcribed in vitro by an RNA polymerase holoenzyme containing sigma hrdD. *J Bacteriol* **178**: 3402-3405.
- Furman, R., A. Sevostyanova & I. Artsimovitch, (2011) Transcription initiation factor DksA has diverse effects on RNA chain elongation. *Nucleic Acids Research*.
- Gaal, T., M. J. Mandel, T. J. Silhavy & R. L. Gourse, (2006) Crl facilitates RNA polymerase holoenzyme formation. *J Bacteriol* **188**: 7966-7970.

- Gao, B., R. Paramanathan & R. S. Gupta, (2006) Signature proteins that are distinctive characteristics of Actinobacteria and their subgroups. *Antonie van Leeuwenhoek* **90**: 69-91.
- Ghosh, T., D. Bose & X. Zhang, (2010) Mechanisms for activating bacterial RNA polymerase. *FEMS Microbiol Rev* **34**: 611-627.
- Goker, M., B. Held, S. Lucas, M. Nolan, M. Yasawong, T. Glavina Del Rio, H. Tice, J. F. Cheng, D. Bruce, J. C. Detter, R. Tapia, C. Han, L. Goodwin, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, M. Rohde, J. Sikorski, R. Pukall, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk & A. Lapidus, (2010) Complete genome sequence of *Olsenella uli* type strain (VPI D76D-27C). *Stand Genomic Sci* **3**: 76-84.
- Hagelueken, G., L. Wiehlmann, T. M. Adams, H. Kolmar, D. W. Heinz, B. TÅ¼mmler & W.-D. Schubert, (2007) Crystal structure of the electron transfer complex rubredoxinâ€“rubredoxin reductase of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences* **104**: 12276-12281.
- Haldenwang, W. G., N. Lang & R. Losick, (1981) A sporulation-induced sigma-like regulatory protein from *b. subtilis*. *Cell* **23**: 615-624.
- Haugen, S. P., M. B. Berkmen, W. Ross, T. Gaal, C. Ward & R. L. Gourse, (2006) rRNA Promoter Regulation by Nonoptimal Binding of  $\sigma$  Region 1.2: An Additional Recognition Element for RNA Polymerase. *Cell* **125**: 1069-1082.
- Haugen, S. P., W. Ross, M. Manrique & R. L. Gourse, (2008) Fine structure of the promoter- $\sigma$  region 1.2 interaction. *Proceedings of the National Academy of Sciences* **105**: 3292-3297.
- Hawley, D. K. & W. R. McClure, (1982) Mechanism of activation of transcription initiation from the  $\lambda$ PRM promoter. *Journal of Molecular Biology* **157**: 493-525.
- Helmann, J. D., (1991) Alternative sigma factors and the regulation of flagellar gene expression. *Molecular Microbiology* **5**: 2875-2882.
- Hesketh, A., W. Chen, J. Ryding, S. Chang & M. Bibb, (2007) The global role of ppGpp synthesis in morphological differentiation and antibiotic production in *Streptomyces coelicolor* A3(2). *Genome Biology* **8**: R161.
- Hilbert, D. W. & P. J. Piggot, (2004) Compartmentalization of Gene Expression during *Bacillus subtilis* Spore Formation. *Microbiology and Molecular Biology Reviews* **68**: 234-262.
- Hofmann, N., R. Wurm & R. Wagner, (2011) The *E. coli* anti-sigma factor Rsd: studies on the specificity and regulation of its expression. *PLoS ONE* **6**: e19235.
- Hopwood, D., (2007) *Streptomyces in Nature and Medicine: The antibiotic Makers*, p. 250. Oxford University Press, New York.
- Hsu, L. M., (2009) Monitoring abortive initiation. *Methods* **47**: 25-36.
- Hu, Y. & A. R. M. Coates, (1999) Transcription of Two Sigma 70 Homologue Genes, sigA and sigB, in Stationary-Phase *Mycobacterium tuberculosis*. *Journal of Bacteriology* **181**: 469-476.

- Hu, Y., Z. Morichaud, S. Chen, J.-P. Leonetti & K. Brodolin, (2012) Mycobacterium tuberculosis RbpA protein is a new type of transcriptional activator that stabilizes the  $\sigma$ A-containing RNA polymerase holoenzyme. *Nucleic Acids Research*.
- Huang, J., J. Shi, V. Molle, B. Sohlberg, D. Weaver, M. J. Bibb, N. Karoonuthaisiri, C. J. Lih, C. M. Kao, M. J. Buttner & S. N. Cohen, (2005) Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Mol Microbiol* **58**: 1276-1287.
- Hughes, K. T., K. L. Gillen, M. J. Semon & J. E. Karlinsey, (1993) Sensing structural intermediates in bacterial flagellar assembly by export of a negative regulator. *Science* **262**: 1277-1280.
- Igarashi, K. & A. Ishihama, (1991) Bipartite functional map of the *E. coli* RNA polymerase  $\alpha$  subunit: Involvement of the C-terminal region in transcription activation by cAMP-CRP. *Cell* **65**: 1015-1022.
- Ignatova, Z. & L. M. Gierasch, (2006) Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 13357-13361.
- Ilag, L. L., L. F. Westblade, C. Deshayes, A. Kolb, S. J. W. Busby & C. V. Robinson, (2004) Mass Spectrometry of *Escherichia coli* RNA Polymerase: Interactions of the Core Enzyme with  $\sigma$ 70 and Rsd Protein. *Structure* **12**: 269-275.
- Jia Y, T.G Dewey, IN Shindyalov, PE Bourne (2004) A new scoring function and associated statistical significance for structure alignment by CE. *J Comput Biol*. 11(5):787-99.
- Jin, D. J., C. Cagliero & Y. N. Zhou, (2011a) Growth rate regulation in *Escherichia coli*. *FEMS Microbiol Rev* **36**: 269-287.
- Jin, D. J. & C. A. Gross, (1988) Mapping and sequencing of mutations in the *Escherichia coli* *rpoB* gene that lead to rifampicin resistance. *Journal of Molecular Biology* **202**: 45-58.
- Jin, D. J., Y. N. Zhou, G. Shaw & X. Ji, (2011b) Structure and function of RapA: a bacterial Swi2/Snf2 protein required for RNA polymerase recycling in transcription. *Biochim Biophys Acta* **1809**: 470-475.
- Jishage, M. & A. Ishihama, (1995) Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of sigma 70 and sigma 38. *Journal of Bacteriology* **177**: 6832-6835.
- Jishage, M. & A. Ishihama, (1998) A stationary phase protein in *Escherichia coli* with binding activity to the major sigma subunit of RNA polymerase. *Proc Natl Acad Sci U S A* **95**: 4953-4958.
- Jishage, M. & A. Ishihama, (1999) Transcriptional organization and in vivo role of the *Escherichia coli* *rsd* gene, encoding the regulator of RNA polymerase sigma D. *J Bacteriol* **181**: 3768-3776.
- Kallifidas, D., D. Thomas, P. Doughty & M. S. B. Paget, (2010) The  $\sigma$ R regulon of *Streptomyces coelicolor* A3(2) reveals a key role in protein quality control during disulphide stress. *Microbiology* **156**: 1661-1672.



- Kalyani, B. S., G. Muteeb, M. Z. Qayyum & R. Sen, (2011) Interaction with the nascent RNA is a prerequisite for the recruitment of Rho to the transcription elongation complex in vitro. *J Mol Biol* **413**: 548-560.
- Kang, J.-G., M.-Y. Hahn, A. Ishihama & J.-H. Roe, (1997) Identification of sigma factors for growth phase-related promoter selectivity of RNA polymerases from *Streptomyces coelicolor* A3(2). *Nucleic Acids Research* **25**: 2566-2573.
- Kang, J. G., M. S. Paget, Y. J. Seok, M. Y. Hahn, J. B. Bae, J. S. Hahn, C. Kleanthous, M. J. Buttner & J. H. Roe, (1999) RsrA, an anti-sigma factor regulated by redox change. *Embo J* **18**: 4292-4298.
- Kapanidis, A. N., E. Margeat, S. O. Ho, E. Kortkhonjia, S. Weiss & R. H. Ebright, (2006) Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism. *Science* **314**: 1144-1147.
- Karimova, G., J. Pidoux, A. Ullmann & D. Ladant, (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences* **95**: 5752-5756.
- Kelemen, G. H., G. L. Brown, J. Kormanec, L. Potůčková, K. F. Chater & M. J. Buttner, (1996) The positions of the sigma-factor genes, *whiG* and *sigF*, in the hierarchy controlling the development of spore chains in the aerial hyphae of *Streptomyces coelicolor* A3(2). *Molecular Microbiology* **21**: 593-603.
- Kelemen, G. H., K. A. Plaskitt, C. G. Lewis, K. C. Findlay & M. J. Buttner, (1995) Deletion of DNA lying close to the *glkA* locus induces ectopic sporulation in *Streptomyces coelicolor* A3(2). *Molecular Microbiology* **17**: 221-230.
- Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater & D. A. Hopwood, (2000) *Practical Streptomyces Genetics*. The John Innes Foundation., Norwich.
- Kim, E. S., J. Y. Song, D. W. Kim, K. F. Chater & K. J. Lee, (2008) A Possible Extended Family of Regulators of Sigma Factor Activity in *Streptomyces coelicolor*. *Journal of Bacteriology* **190**: 7559-7566.
- Kingsford, C. L., K. Ayanbule & S. L. Salzberg, (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* **8**: R22.
- Klocko, A. D. & K. M. Wassarman, (2009) 6S RNA binding to Esigma(70) requires a positively charged surface of sigma(70) region 4.2. *Mol Microbiol* **73**: 152-164.
- Kresge, N., R. D. Simoni & R. L. Hill, (2004) Selman Waksman: the Father of Antibiotics. *J. Biol. Chem.* **279**: e7-.
- Kulbachinskiy, A. & A. Mustaev, (2006) Region 3.2 of the  $\sigma$  Subunit Contributes to the Binding of the 3'-Initiating Nucleotide in the RNA Polymerase Active Center and Facilitates Promoter Clearance during Initiation. *Journal of Biological Chemistry* **281**: 18273-18276.
- Kulp, A. & M. J. Kuehn, (2011) Recognition of  $\beta$ -Strand Motifs by RseB Is Required for  $\sigma^E$  Activity in *Escherichia coli*. *Journal of Bacteriology* **193**: 6179-6186.
- Kumar, A., H. S. Williamson, N. Fujita, A. Ishihama & R. S. Hayward, (1995) A partially functional 245-amino-acid internal deletion derivative of *Escherichia coli* sigma 70. *Journal of Bacteriology* **177**: 5193-5196.

- Kusuya, Y., K. Kurokawa, S. Ishikawa, N. Ogasawara & T. Oshima, (2011) Transcription factor GreA contributes to resolving promoter-proximal pausing of RNA polymerase in *Bacillus subtilis* cells. *J Bacteriol* **193**: 3090-3099.
- Lakey, J. H., E. J. A. Lea, B. A. M. Rudd, H. M. Wright & D. A. Hopwood, (1983) A New Channel-forming Antibiotic from *Streptomyces coelicolor* A3(2) Which Requires Calcium for its Activity. *Journal of General Microbiology* **129**: 3565-3573.
- Lee, J.-H., C. W. Lennon, W. Ross & R. L. Gourse, (2012) Role of the Coiled-Coil Tip of *Escherichia coli* DksA in Promoter Control. *Journal of Molecular Biology*.
- Leibman, M. & A. Hochschild, (2007) A [sigma]-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. *EMBO J* **26**: 1579-1590.
- Lemke, J. J., P. Sanchez-Vazquez, H. L. Burgos, G. Hedberg, W. Ross & R. L. Gourse, (2011) Direct regulation of *Escherichia coli* ribosomal protein promoters by the transcription factors ppGpp and DksA. *Proc Natl Acad Sci U S A* **108**: 5712-5717.
- Li, M., H. Moyle & M. Susskind, (1994) Target of the transcriptional activation function of phage lambda cl protein. *Science* **263**: 75-77.
- Li, W., A. R. Bottrill, M. J. Bibb, M. J. Buttner, M. S. Paget & C. Kleanthous, (2003) The Role of zinc in the disulphide stress-regulated anti-sigma factor RsrA from *Streptomyces coelicolor*. *J Mol Biol* **333**: 461-472.
- Li, W., C. E. Stevenson, N. Burton, P. Jakimowicz, M. S. Paget, M. J. Buttner, D. M. Lawson & C. Kleanthous, (2002) Identification and structure of the anti-sigma factor-binding domain of the disulphide-stress regulated sigma factor sigma(R) from *Streptomyces coelicolor*. *J Mol Biol* **323**: 225-236.
- Loewen, P. C. & R. Hengge-Aronis, (1994) The Role of the Sigma Factor sigmas (KatF) in Bacterial Global Regulation. *Annual Review of Microbiology* **48**: 53-80.
- Lonetto, M. A., K. L. Brown, K. E. Rudd & M. J. Buttner, (1994) Analysis of the *Streptomyces coelicolor* sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci U S A* **91**: 7573-7577.
- MacNeil, D. J., K. M. Gewain, C. L. Ruby, G. Dezeny, P. H. Gibbons & T. MacNeil, (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**: 61-68.
- Makarova, K. S., V. A. Ponomarev & E. V. Koonin, (2001) Two C or not two C: recurrent disruption of Zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. *Genome Biol* **2**: RESEARCH 0033.
- Malhotra, A., E. Severinova & S. A. Darst, (1996) Crystal structure of a sigma(70) subunit fragment from *E-coli* RNA polymerase. *Cell* **87**: 127-136.
- Malshetty, V., K. Kurthkoti, A. China, B. Mallick, S. Yamunadevi, P. B. Sang, N. Srinivasan, V. Nagaraja & U. Varshney, (2010) Novel insertion and deletion

- mutants of RpoB that render *Mycobacterium smegmatis* RNA polymerase resistant to rifampicin-mediated inhibition of transcription. *Microbiology* **156**: 1565-1573.
- Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. T. Cole & R. Brosch, (2004) Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* **150**: 483-496.
- Mavrommatis, K., R. Pukall, C. Rohde, F. Chen, D. Sims, T. Brettin, C. Kuske, J. C. Detter, C. Han, A. Lapidus, A. Copeland, T. Glavina Del Rio, M. Nolan, S. Lucas, H. Tice, J. F. Cheng, D. Bruce, L. Goodwin, S. Pitluck, G. Ovchinnikova, A. Pati, N. Ivanova, A. Chen, K. Palaniappan, P. Chain, P. D'Haeseleer, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, M. Rohde, H. P. Klenk & N. C. Kyrpides, (2009) Complete genome sequence of *Cryptobacterium curtum* type strain (12-3). *Stand Genomic Sci* **1**: 93-100.
- Miller, A., D. Wood, R. H. Ebright & L. B. Rothman-Denes, (1997) RNA polymerase beta' subunit: a target of DNA binding-independent activation. *Science (New York, N.Y.)* **275**: 1655-1657.
- Mitchell, J. E., T. Oshima, S. E. Piper, C. L. Webster, L. F. Westblade, G. Karimova, D. Ladant, A. Kolb, J. L. Hobman, S. J. Busby & D. J. Lee, (2007) The *Escherichia coli* regulator of sigma 70 protein, Rsd, can up-regulate some stress-dependent promoters by sequestering sigma 70. *J Bacteriol* **189**: 3489-3495.
- Monteil, V., A. Kolb, J. D'Alayer, P. Beguin & F. Norel, (2010a) Identification of Conserved Amino Acid Residues of the *Salmonella* {sigma}S Chaperone Crl Involved in Crl-{sigma}S Interactions. *J. Bacteriol.* **192**: 1075-1087.
- Monteil, V., A. Kolb, C. Mayer, S. Hoos, P. England & F. Norel, (2010b) Crl binds to domain 2 of {sigma}S and confers a competitive advantage to a natural rpoS mutant of *Salmonella enterica* serovar Typhi. *J. Bacteriol.*: JB.00801-00810.
- Mukherjee, R. & D. Chatterji, (2008) Stationary phase induced alterations in mycobacterial RNA polymerase assembly: A cue to its phenotypic resistance towards rifampicin. *Biochemical and Biophysical Research Communications* **369**: 899-904.
- Mukhopadhyay, J., K. Das, S. Ismail, D. Koppstein, M. Jang, B. Hudson, S. Sarafianos, S. Tuske, J. Patel, R. Jansen, H. Irschik, E. Arnold & R. H. Ebright, (2008) The RNA Polymerase "Switch Region" Is a Target for Inhibitors. *Cell* **135**: 295-307.
- Mukhopadhyay, J., E. Sineva, J. Knight, R. M. Levy & R. H. Ebright, (2004) Antibacterial Peptide Microcin J25 Inhibits Transcription by Binding within and Obstructing the RNA Polymerase Secondary Channel. *Molecular Cell* **14**: 739-751.
- Murakami, K. S. & S. A. Darst, (2003) Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol* **13**: 31-39.

- Murakami, K. S., S. Masuda, E. A. Campbell, O. Muzzin & S. A. Darst, (2002a) Structural Basis of Transcription Initiation: An RNA Polymerase Holoenzyme-DNA Complex. *Science* **296**: 1285-1290.
- Murakami, K. S., S. Masuda & S. A. Darst, (2002b) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**: 1280-1284.
- Newell, K. V., D. P. Thomas, D. Brekasis & M. S. Paget, (2006) The RNA polymerase-binding protein RbpA confers basal levels of rifampicin resistance on *Streptomyces coelicolor*. *Mol Microbiol* **60**: 687-696.
- Newton, G. L. & R. C. Fahey, (2008) Regulation of mycothiol metabolism by sigma(R) and the thiol redox sensor anti-sigma factor RsrA. *Mol Microbiol* **68**: 805-809.
- Nickels, B. E., S. L. Dove, K. S. Murakami, S. A. Darst & A. Hochschild, (2002) Protein-Protein and Protein-DNA Interactions of  $\sigma 70$  Region 4 Involved in Transcription Activation by  $\lambda$ cl. *Journal of Molecular Biology* **324**: 17-34.
- Niu, W., Y. Kim, G. Tau, T. Heyduk & R. H. Ebright, (1996) Transcription Activation at Class II CAP-Dependent Promoters: Two Interactions between CAP and RNA Polymerase. *Cell* **87**: 1123-1134.
- O'Connor, T. J., P. Kanellis & J. R. Nodwell, (2002) The ramC gene is required for morphogenesis in *Streptomyces coelicolor* and expressed in a cell type-specific manner under the direct control of RamR. *Molecular Microbiology* **45**: 45-57.
- Ohnishi, K., K. Kutsukake, H. Suzuki & T. Lino, (1992) A novel transcriptional regulation mechanism in the flagellar regulon of *Salmonella typhimurium*: an antisigma factor inhibits the activity of the flagellum-specific sigma factor, sigma F. *Mol Microbiol* **6**: 3149-3157.
- Owen, G. A., B. Pascoe, D. Kallifidas & M. S. Paget, (2007) Zinc-responsive regulation of alternative ribosomal protein genes in *Streptomyces coelicolor* involves zur and sigmaR. *J Bacteriol* **189**: 4078-4086.
- Paget, M. S., J. B. Bae, M. Y. Hahn, W. Li, C. Kleanthous, J. H. Roe & M. J. Buttner, (2001a) Mutational analysis of RsrA, a zinc-binding anti-sigma factor with a thiol-disulphide redox switch. *Mol Microbiol* **39**: 1036-1047.
- Paget, M. S., L. Chamberlin, A. Atrih, S. J. Foster & M. J. Buttner, (1999) Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol* **181**: 204-211.
- Paget, M. S. & J. D. Helmann, (2003) The sigma70 family of sigma factors. *Genome Biol* **4**: 203.
- Paget, M. S., J. G. Kang, J. H. Roe & M. J. Buttner, (1998) sigmaR, an RNA polymerase sigma factor that modulates expression of the thioredoxin system in response to oxidative stress in *Streptomyces coelicolor* A3(2). *Embo J* **17**: 5776-5782.
- Paget, M. S., V. Molle, G. Cohen, Y. Aharonowitz & M. J. Buttner, (2001b) Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Mol Microbiol* **42**: 1007-1020.

- Paget, M. S. B., G. Hintermann & C. P. Smith, (1994) Construction and application of streptomycete promoter probe vectors which employ the *Streptomyces glaucescens* tyrosinase-encoding gene as reporter. *Gene* **146**: 105-110.
- Paget, M. S. B., H.-J. Hong, M. J. Bibb & M. J. Buttner, (2002) The ECF sigma factors of *Streptomyces coelicolor* A3(2). In: SGM symposium 61: Signals, switches, regulons and cascades: control of bacterial gene expression. D. A. Hodgson & C. M. Thomas (eds). Cambridge: Cambridge University Press, pp.
- Panina, E. M., A. A. Mironov & M. S. Gelfand, (2003) Comparative genomics of bacterial zinc regulons: Enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proceedings of the National Academy of Sciences* **100**: 9912-9917.
- Patikoglou, G. A., L. F. Westblade, E. A. Campbell, V. Lamour, W. J. Lane & S. A. Darst, (2007) Crystal structure of the *Escherichia coli* regulator of sigma70, Rsd, in complex with sigma70 domain 4. *J Mol Biol* **372**: 649-659.
- Paul, B. J., M. M. Barker, W. Ross, D. A. Schneider, C. Webb, J. W. Foster & R. L. Gourse, (2004) DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* **118**: 311-322.
- Paul, B. J., M. B. Berkmen & R. L. Gourse, (2005) DksA potentiates direct activation of amino acid promoters by ppGpp. *Proc Natl Acad Sci U S A* **102**: 7823-7828.
- Perederina, A., V. Svetlov, M. N. Vassilyeva, T. H. Tahirov, S. Yokoyama, I. Artsimovitch & D. G. Vassilyev, (2004) Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. *Cell* **118**: 297-309.
- Pukall, R., A. Lapidus, M. Nolan, A. Copeland, T. Glavina Del Rio, S. Lucas, F. Chen, H. Tice, J. F. Cheng, O. Chertkov, D. Bruce, L. Goodwin, C. Kuske, T. Brettin, J. C. Detter, C. Han, S. Pitluck, A. Pati, K. Mavrommatis, N. Ivanova, G. Ovchinnikova, A. Chen, K. Palaniappan, S. Schneider, M. Rohde, P. Chain, P. D'Haeseleer, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, N. C. Kyrpides, H. P. Klenk & P. Hugenholtz, (2009) Complete genome sequence of *Slackia heliotrinireducens* type strain (RHS 1). *Stand Genomic Sci* **1**: 234-241.
- Raoult, D., H. Ogata, S. Audic, C. Robert, K. Suhre, M. Drancourt & J. M. Claverie, (2003) *Tropheryma whipplei* Twist: a human pathogenic Actinobacteria with a reduced genome. *Genome Res* **13**: 1800-1809.
- Rhayat, L., S. Duperrier, R. Carballido-López, O. Pellegrini & P. Stragier, (2009) Genetic Dissection of an Inhibitor of the Sporulation Sigma Factor [sigma]G. *Journal of Molecular Biology* **390**: 835-844.
- Rhodium, V. A. & S. J. W. Busby, (1998) Positive activation of gene expression. *Current Opinion in Microbiology* **1**: 152-159.
- Rhodium, V. A. & S. J. W. Busby, (2000) Interactions between activating region 3 of the *Escherichia coli* cyclic AMP receptor protein and region 4 of the RNA polymerase  $\sigma$ 70 subunit: application of suppression genetics. *Journal of Molecular Biology* **299**: 311-324.

- Riccardi, G., A. Milano, M. R. Pasca & D. H. Nies, (2008) Genomic analysis of zinc homeostasis in *Mycobacterium tuberculosis*. *FEMS Microbiol Lett* **287**: 1-7.
- Rigali, S., F. Titgemeyer, S. Barends, S. Mulder, A. W. Thomae, D. A. Hopwood & G. P. van Wezel, (2008) Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Rep* **9**: 670-675.
- Ring, B. Z., W. S. Yarnell & J. W. Roberts, (1996) Function of *E. coli* RNA Polymerase  $\sigma$  Factor-  $\sigma 70$  in Promoter-Proximal Pausing. *Cell* **86**: 485-493.
- Rodrigue, S., R. Provvedi, P. E. Jacques, L. Gaudreau & R. Manganelli, (2006) The sigma factors of *Mycobacterium tuberculosis*. *FEMS Microbiol Rev* **30**: 926-941.
- Ross, W., K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov & R. Gourse, (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**: 1407-1413.
- Rudd, B. A. M. & D. A. Hopwood, (1980) A Pigmented Mycelial Antibiotic in *Streptomyces coelicolor*: Control by a Chromosomal Gene Cluster. *Journal of General Microbiology* **119**: 333-340.
- Rutherford, S. T., C. L. Villers, J. H. Lee, W. Ross & R. L. Gourse, (2009) Allosteric control of *Escherichia coli* rRNA promoter complexes by DksA. *Genes Dev* **23**: 236-248.
- Ryu, Yong-Gu, E.S. Kim, D.W. Kim, S.K. Kim, K.J. Lee (2007) Differential Stringent Responses of *Streptomyces coelicolor* M600 to Starvation of Specific Nutrients. *Journal of Microbiology & Biotechnology* **17**(2):305-312.
- Sachdeva, P., R. Misra, A. K. Tyagi & Y. Singh, (2010) The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. *FEBS Journal* **277**: 605-626.
- Sassetti, C. M., D. H. Boyd & E. J. Rubin, (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**: 77-84.
- Saunders, E., R. Pukall, B. Abt, A. Lapidus, T. Glavina Del Rio, A. Copeland, H. Tice, J. F. Cheng, S. Lucas, F. Chen, M. Nolan, D. Bruce, L. Goodwin, S. Pitluck, N. Ivanova, K. Mavromatis, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, P. Chain, L. Meincke, D. Sims, T. Brettin, J. C. Detter, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk & C. Han, (2009) Complete genome sequence of *Eggerthella lenta* type strain (IPP VPI 0255). *Stand Genomic Sci* **1**: 174-182.
- Serrano, M., G. Real, J. Santos, J. Carneiro, C. P. Moran, Jr. & A. O. Henriques, (2011) A Negative Feedback Loop That Limits the Ectopic Activation of a Cell Type-Specific Sporulation Sigma Factor of *Bacillus subtilis*. *PLoS Genet* **7**: e1002220.
- Sheeler, N. L., S. V. MacMillan & J. R. Nodwell, (2005) Biochemical Activities of the *absA* Two-Component System of *Streptomyces coelicolor*. *Journal of Bacteriology* **187**: 687-696.

- Shiina, T., K. Tanaka & H. Takahashi, (1991) Sequence of hrdB, an essential gene encoding sigma-like transcription factor of *Streptomyces coelicolor* A3(2): homology to principal sigma factors. *Gene* **107**: 145-148.
- Shindyalov, I. N. & P. E. Bourne, (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering* **11**: 739-747.
- Sivashanmugam, A., V. Murray, C. Cui, Y. Zhang, J. Wang & Q. Li, (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science* **18**: 936-948.
- Spagnolo, L., I. Toro, M. D'Orazio, P. O'Neill, J. Z. Pedersen, O. Carugo, G. Rotilio, A. Battistoni & K. DjinoVIC-Carugo, (2004) Unique features of the sodC-encoded superoxide dismutase from *Mycobacterium tuberculosis*, a fully functional copper-containing enzyme lacking zinc in the active site. *J Biol Chem* **279**: 33447-33455.
- Stallings, C. L. & M. S. Glickman, (2011) CarD: A new RNA polymerase modulator in mycobacteria. *Transcription* **2**: 15-18.
- Stallings, C. L., N. C. Stephanou, L. Chu, A. Hochschild, B. E. Nickels & M. S. Glickman, (2009) CarD is an essential regulator of rRNA transcription required for *Mycobacterium tuberculosis* persistence. *Cell* **138**: 146-159.
- Staroń, A., H. J. Sofia, S. Dietrich, L. E. Ulrich, H. Liesegang & T. Mascher, (2009) The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF)  $\sigma$  factor protein family. *Molecular Microbiology* **74**: 557-581.
- Stepanova, E., J. Lee, M. Ozerova, E. Semenova, K. Datsenko, B. L. Wanner, K. Severinov & S. Borukhov, (2007) Analysis of Promoter Targets for *Escherichia coli* Transcription Elongation Factor GreA In Vivo and In Vitro. *J. Bacteriol.* **189**: 8772-8785.
- Studier, F. W., (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**: 207-234.
- Studier, F. W. & B. A. Moffatt, (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**: 113-130.
- Sun, J., A. Hesketh & M. Bibb, (2001) Functional analysis of relA and rshA, two relA/spoT homologues of *Streptomyces coelicolor* A3(2). *J Bacteriol* **183**: 3488-3498.
- Szalewska-Palasz, A., A. Wegrzyn, A. Blaszcak, K. Taylor & G. Wegrzyn, (1998) DnaA-stimulated transcriptional activation of ori $\lambda$ : *Escherichia coli* RNA polymerase  $\beta$  subunit as a transcriptional activator contact site. *Proceedings of the National Academy of Sciences* **95**: 4241-4246.
- Tan, H., H. Yang, Y. Tian, W. Wu, C. A. Whatling, L. C. Chamberlin, M. J. Buttner, J. Nodwell & K. F. Chater, (1998) The *Streptomyces coelicolor* sporulation-specific sigma WhiG form of RNA polymerase transcribes a gene encoding a ProX-like protein that is dispensable for sporulation. *Gene* **212**: 137-146.

- Tanaka, K., T. Shiina & H. Takahashi, (1991) Nucleotide sequence of genes *hrdA*, *hrdC*, and *hrdD* from *Streptomyces coelicolor* A3(2) having similarity to *rpoD* genes. *Mol Gen Genet* **229**: 334-340.
- Tehranchi, A. K., M. D. Blankschien, Y. Zhang, J. A. Halliday, A. Srivatsan, J. Peng, C. Herman & J. D. Wang, (2010) The Transcription Factor DksA Prevents Conflicts between DNA Replication and Transcription Machinery. *Cell* **141**: 595-605.
- Trotochaud, A. E. & K. M. Wassarman, (2004) 6S RNA function enhances long-term cell survival. *J Bacteriol* **186**: 4978-4985.
- Tupin, A., M. Gualtieri, F. Roquet-Banères, Z. Morichaud, K. Brodolin & J.-P. Leonetti, (2010) Resistance to rifampicin: at the crossroads between ecological, genomic and medical concerns. *International Journal of Antimicrobial Agents* **35**: 519-523.
- Tuske, S., S. G. Sarafianos, X. Wang, B. Hudson, E. Sineva, J. Mukhopadhyay, J. J. Birktoft, O. Leroy, S. Ismail, A. D. Clark Jr, C. Dharia, A. Napoli, O. Laptenko, J. Lee, S. Borukhov, R. H. Ebright & E. Arnold, (2005) Inhibition of Bacterial RNA Polymerase by Streptolydigin: Stabilization of a Straight-Bridge-Helix Active-Center Conformation. *Cell* **122**: 541-552.
- Typas, A., C. Barembruch, A. Possling & R. Hengge, (2007) Stationary phase reorganisation of the *Escherichia coli* transcription machinery by Crl protein, a fine-tuner of sigma activity and levels. *Embo J* **26**: 1569-1578.
- Uguru, G. C., K. E. Stephens, J. A. Stead, J. E. Towle, S. Baumberg & K. J. McDowall, (2005) Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. *Molecular Microbiology* **58**: 131-150.
- van Keulen, G., J. Alderson, J. White & R. G. Sawers, (2007) The obligate aerobic actinomycete *Streptomyces coelicolor* A3(2) survives extended periods of anaerobic stress. *Environ Microbiol* **9**: 3143-3149.
- van Wezel, G. P. & K. J. McDowall, (2011) The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Natural Product Reports* **28**.
- Vassilyev, D. G., S. Sekine, O. Laptenko, J. Lee, M. N. Vassilyeva, S. Borukhov & S. Yokoyama, (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**: 712-719.
- Villain-Guillot, P., L. Bastide, M. Gualtieri & J.-P. Leonetti, (2007) Progress in targeting bacterial transcription. *Drug Discovery Today* **12**: 200-208.
- Viollier, P. H., G. H. Kelemen, G. E. Dale, K. T. Nguyen, M. J. Buttner & C. J. Thompson, (2003) Specialized osmotic stress response systems involve multiple SigB-like sigma factors in *Streptomyces coelicolor*. *Mol Microbiol* **47**: 699-714.
- Vrentas, C. E., T. Gaal, M. B. Berkmen, S. T. Rutherford, S. P. Haugen, D. G. Vassilyev, W. Ross & R. L. Gourse, (2008) Still looking for the magic spot: the crystallographically defined binding site for ppGpp on RNA polymerase is unlikely to be responsible for rRNA transcription regulation. *J Mol Biol* **377**: 551-564.



- Wassarman, K. M. & R. M. Saecker, (2006) Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science* **314**: 1601-1603.
- Wassarman, K. M. & G. Storz, (2000) 6S RNA regulates E. coli RNA polymerase activity. *Cell* **101**: 613-623.
- Wayne, L. G. & L. G. Hayes, (1996) An in vitro model for sequential study of shutdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. *Infection and Immunity* **64**: 2062-2069.
- Wayne, L. G. & C. D. Sohaskey, (2001) Nonreplicating persistence of *Mycobacterium tuberculosis*1. *Annual Review of Microbiology* **55**: 139-163.
- Weber, H., T. Polen, J. Heuveling, V. F. Wendisch & R. Hengge, (2005) Genome-Wide Analysis of the General Stress Response Network in Escherichia coli:  $\sigma$ S-Dependent Genes, Promoters, and Sigma Factor Selectivity. *Journal of Bacteriology* **187**: 1591-1603.
- Westpheling, J. & M. Brawner, (1989) Two transcribing activities are involved in expression of the Streptomyces galactose operon. *Journal of Bacteriology* **171**: 1355-1361.
- Wilson, C. & A. J. Dombroski, (1997) Region 1 of sigma(70) is required for efficient isomerization and initiation of transcription by Escherichia coli RNA polymerase. *Journal of Molecular Biology* **267**: 60-74.
- Wollmann, P. & K. Zeth, (2007) The Structure of RseB: A Sensor in Periplasmic Stress Response of E. coli. *Journal of Molecular Biology* **372**: 927-941.
- Wright, L. F. & D. A. Hopwood, (1976a) Actinorhodin is a Chromosomally-determined Antibiotic in Streptomyces coelicolor A3(2). *Journal of General Microbiology* **96**: 289-297.
- Wright, L. F. & D. A. Hopwood, (1976b) Identification of the Antibiotic Determined by the SCP1 Plasmid of Streptomyces coelicolor A3 (2). *Journal of General Microbiology* **95**: 96-106.
- Yang, X. & P. J. Lewis, (2010) The interaction between bacterial transcription factors and RNA polymerase during the transition from initiation to elongation. *Transcription* **1**: 66-69.
- Yang, X., S. Molimau, G. P. Doherty, E. B. Johnston, J. Marles-Wright, R. Rothnagel, B. Hankamer, R. J. Lewis & P. J. Lewis, (2009) The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO Rep* **10**: 997-1002.
- Yanisch-Perron, C., J. Vieira & J. Messing, (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* **33**: 103-119.
- Young, B. A., L. C. Anthony, T. M. Gruber, T. M. Arthur, E. Heyduk, C. Z. Lu, M. M. Sharp, T. Heyduk, R. R. Burgess & C. A. Gross, (2001) A Coiled-Coil from the RNA Polymerase  $\beta'$  Subunit Allosterically Induces Selective Nontemplate Strand Binding by  $\sigma$ 70. *Cell* **105**: 935-944.
- Young, B. A., T. M. Gruber & C. A. Gross, (2002) Views of transcription initiation. *Cell* **109**: 417-420.

- Yuan, A. H., B. D. Gregory, J. S. Sharp, K. D. McCleary, S. L. Dove & A. Hochschild, (2008) Rsd family proteins make simultaneous interactions with regions 2 and 4 of the primary sigma factor. *Mol Microbiol* **70**: 1136-1151.
- Yuan, A. H., B. E. Nickels & A. Hochschild, (2009) The bacteriophage T4 AsiA protein contacts the  $\beta$ -flap domain of RNA polymerase. *Proceedings of the National Academy of Sciences* **106**: 6597-6602.
- Zdanowski, K., P. Doughty, P. Jakimowicz, L. O'Hara, M. J. Buttner, M. S. Paget & C. Kleanthous, (2006) Assignment of the zinc ligands in RsrA, a redox-sensing ZAS protein from *Streptomyces coelicolor*. *Biochemistry-Us* **45**: 8294-8300.
- Zenkin, N., A. Kulbachinskiy, Y. Yuzenkova, A. Mustaev, I. Bass, K. Severinov & K. Brodolin, (2007) Region 1.2 of the RNA polymerase [sigma] subunit controls recognition of the -10 promoter element. *EMBO J* **26**: 955-964.
- Zhang, G., E. A. Campbell, L. Minakhin, C. Richter, K. Severinov & S. A. Darst, (1999) Crystal Structure of *Thermus aquaticus* Core RNA Polymerase at 3.3 Å Resolution. *Cell* **98**: 811-824.