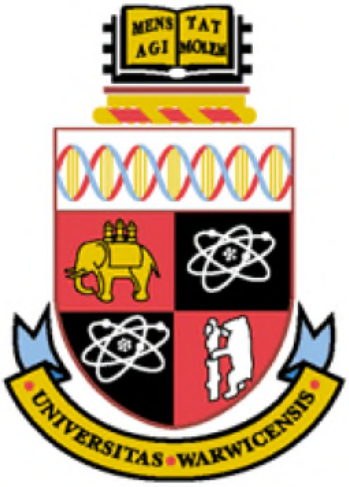


Discovery and Biosynthesis of Novel Natural Products from *Streptomyces venezuelae*

John David Sidda

Supervisor: Dr. Christophe Corre

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry

Department of Chemistry, University of Warwick

August 2015

Contents

Declaration viii

Acknowledgements ix

List of Figures xi

List of Schemes xxv

List of Tables xxvii

Abbreviations xxviii

Abstract xxxi

1. Introduction 1

1.1 *Streptomyces* – the natural product makers 1

1.2 Strategies for natural product discovery 5

1.2.1 Phenotype-guided discovery 5

1.2.2 Pleiotropic regulation of secondary metabolism 6

1.3 Genome mining approaches to natural product discovery 12

1.3.1 Sequencing projects have revealed the biosynthetic potential of *Streptomyces* 12

1.3.2 Bioinformatics can be used to predict natural product scaffolds 15

1.3.3 Heterologous expression 17

1.3.4 Gene knockouts 18

1.3.5 Genetic inactivation of repressors/overexpression of activators 19

1.4 TetR transcriptional regulators 21

1.4.1 Functions of TetR family regulators 21

1.4.2 Ligands for TetRs 24

1.4.3 ArpA family transcriptional repressors 25

1.4.4 Ligands for ArpA-like repressors 27

1.5 Biosynthesis of *Streptomyces* signalling molecules 32

1.5.1 The key step in GBL biosynthesis 32

1.5.2 Reduction of 6-keto-type GBLs to 6-hydroxy-type GBLs 37

1.5.3 AHFCA biosynthesis 39

1.6 Ligand-protein and protein-DNA interactions of ArpA-like repressors 43

1.6.1 Determination of ligand-repressor specificity 43

1.6.2 A potential role for ArpA ligands in interspecies signalling 46

1.6.3 Repressor-DNA interactions 47

1.6.4 Phenotypes of *arpA* deletion mutants 49

1.6.5 Regulatory cascade controlling methylenomycin biosynthesis 50

1.7 *mmfLHP*-like biosynthetic gene clusters 53

1.7.1 Identification of other *mmfLHP* clusters 53

1.7.2 Previous work on *gbnABC* system in *S. venezuelae* 55

1.8 Aims of project 58

1.9 Outline of thesis 59

2. Identification and characterisation of the novel natural products gaburedins A-F from *Streptomyces venezuelae* 60

2.1 Growth conditions 61

2.1.1 Production of metabolites **80**-**85** in *S. venezuelae gbnR::apra* mutant 61

2.1.2 Culture conditions for maximising production of *S. venezuelae gbnR::apra* metabolites **80**-**85** 63

2.2 Structure elucidation 64

2.2.1 Liquid chromatography-mass spectrometry (LC-MS) analysis 64

2.2.2 Characterisation of metabolite **80** by nuclear magnetic resonance (NMR) spectroscopy 67

2.2.3 Synthesis of authentic standard of gaburedin A 69

2.2.4 Feeding experiments to overproduce gaburedins **80**-**85** 72

2.2.5 Chiral HPLC analysis of L- and D-gaburedin A 78

2.3 Discussion 81

3. Towards gaburedin biosynthesis 84

3.1 Bioinformatic analyses of GbnA, GbnB, GbnC 85

3.1.1 BLAST analysis 85

3.1.2 *gbnA* encodes for a PLP-binding putative decarboxylase 87

3.1.3 *gbnB* encodes for a putative AMP-ligase 89

3.1.4 *gbnC* encodes for a transport protein 91

3.2 GbnB is essential for gaburedin biosynthesis 91

3.2.1 Characterisation of the *gbnB/gbnR* double mutant strain 91

3.2.2 Chemotyping of the *gbnB/gbnR* double mutant strain by LC-MS analyses 93

3.3 Proposed biosynthetic route to gaburedins 94

3.3.1 Feeding experiments with putative precursor molecules 94

3.3.2 Proposed biosynthetic scheme 97

3.3.3 Origin of the ureidyl carbon 100

3.3.4 Origin of the conserved GABA moiety 102

3.4 Probing substrate diversity of GbnA and GbnB 105

3.4.1 Substrate specificity of GbnA 105

3.4.2 Probing GbnB substrate specificity 114

3.5 Summary 120

4. Exploring diversity of gaburedin aminoacyl moieties 122

4.1 *In vivo* investigations of aminoacyl moiety tolerance 123

4.1.1 Incorporation experiments with small amino acid side chains 123

4.1.2 Incorporation experiments with large amino acid side chains 127

4.2 Incorporation experiments with amino acid analogues 131

4.2.1 Incorporation of amines 131

4.2.2 Incorporation experiments with aminoalcohols and aminoacyl esters 134

4.2.3 Incorporation of a dipeptide 137

4.3 Discussion of gaburedin biosynthesis 137

4.3.1 Effect of altering SMMS composition on production of gaburedins A-F 137

4.3.2 The acetylation observation 141

5. Reconstitution of *gbnAB* system from *Streptomyces venezuelae* 143

5.1 Cloning *gbnAB (sven_4179 and sven_4180)* from *S. venezuelae* 144

5.2 Cloning of synthetic *S. venezuelae gbnB* gene and orthologues 148

5.2.1 Synthetic genes for heterologous expression in *E. coli* 148

5.2.2 Metabolic profiling of engineered *E. coli* strains carrying *gbnB* and orthologues 150

5.2.3 Overproduction of recombinant enzymes in *E. coli* BL21star 151

5.3 Alternative cloning strategies for reconstitution of *gbnABC* in *E. coli* 153

5.3.1 Sub-cloning synthetic *gbnA* into pJDS06 and synthetic *gbnB* into pJDS05 153

5.3.2 Cloning with pET-Duet and pACYC-Duet vectors 156

5.3.3 Metabolic profiling of *E.coli* carrying pET/pACYC-Duet derived plasmids159

5.4 – Summary of molecular biology work 161

6. Regulation of gaburedin biosynthesis in *S. venezuelae* 163

6.1 Identification and characterisation of *Streptomyces venezuelae* signalling molecules 166

6.1.1 Metabolic profiling of *S. venezuelae* mutants 166

6.1.2 Comparison of putative signalling molecules with *Streptomyces* M1152 167

6.1.3 Characterisation of novel SCBs in *S. coelicolor* M1152 169

6.1.4 AHFCA signalling molecules in *Streptomyces venezuelae* 175

6.1.5 Deductions about SCB/AHFCA biosynthesis in *S. venezuelae* 181

6.2 Production of *S. venezuelae* signalling molecules in *E. coli* 183

6.2.1 Overexpression of *sgnL* in *E. coli* 183

6.2.2 Implications for AHFCA/GBL biosynthesis 186

6.3 Comparison of SgnR and MmfR ligand binding sites 188

6.4 Biological role of *S. venezuelae* AHFCA signalling molecules 191

6.4.1 AHFCAs induce gaburedin production in an *S. venezuelae sgnL-P* mutant 191

6.4.2 Effect of AHFCAs on *S. venezuelae gbnR* mutant and wild-type strains 193

7. Summary and future work 201

7.1 Conclusions 201

7.1.1 Deletion of *S. venezuelae gbnR* leads to overproduction of novel natural products 201

7.1.2 Gaburedin biosynthesis proceeds via an intermediate derived from glutamic acid 201

7.1.3 GbnABC systems exist across a range of species 202

7.1.4 Heterologous expression of *gbnB* in *E. coli* is not sufficient to produce gaburedins 202

7.1.5 Gaburedin biosynthesis is controlled by AHFCA signalling molecules 203

7.2 Outlook 204

7.2.1 Deletion of *arpA*-like repressors is a valuable tool for natural product discovery 204

7.2.2 GbnB-type enzymes may have potential for producing drug-like molecules 204

7.2.3 Some signalling molecules produced by *S. venezuelae gbnR* mutants are also produced by other *Streptomyces* 205

7.2.4 MmfR-MmfLHP-MmyR-like systems exist in a range of *Streptomyces* 205

7.3 Future work 207

7.3.1 Biological role of gaburedins 207

7.3.2 Characterisation of SCB signalling molecules in *S. venezuelae* 208

7.3.3 AHFCA signalling in *S. venezuelae* 208

7.3.4 Deletion of *mmyR*-like repressors in other *Streptomyces* species 209

8. Experimental 210

8.1 Analytical Chemistry 210

8.1.1 Mass spectrometry and NMR spectroscopy 210

8.1.2 Extraction protocol 210

8.1.3 Liquid-chromatography-mass spectrometry (LC-MS) 210

8.1.4 Reverse phase HPLC 212

8.1.5 Chiral HPLC analysis 213

8.2 Synthetic Chemistry 213

8.2.1 Preparation of **90** 213

8.2.1 Preparation of **91** 215

8.3 Microbiological methods 216

8.3.1 Bacterial strains 216

8.3.2 Culture media 216

8.3.3 Modified culture media 217

8.3.4 Culture media enriched with stable isotope precursors 217

8.3.5 Feeding *Streptomyces coelicolor* M1152 with stable isotope precursors 218

8.3.6 Preparation of bacterial spores 218

8.3.7 Preparation of *S. venezuelae* Genomic DNA 219

8.3.8 Preparation of SV4 F01 Cosmid DNA 219

8.3.9 Chemical transformations 220

8.5.10 Restriction digests 220

8.3.11 Polymerase Chain Reaction (PCR) conditions 221

8.3.12 Agarose Gel Electrophoresis 222

8.3.13 Protein production in *E. coli* BL21star 222

8.3.14 Preparation of cell lysates 222

8.3.15 Sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE) 223

8.4 Biological assays 223

8.4.1 Induction assay to screen for gaburedin production by non-producing strain 223

8.4.2 Induction assay with purified AHFCA compounds 224

8.4.3 Antibiotic assay 225

8.4.4 *In vivo* assay for gaburedin or gaburedin analogue production by *E. coli*

BL21star/pJDS02 and *E. coli* BL21star/pJDS04 225

8.4.5 *In vivo* assay for gaburedin or gaburedin analogue production by *E. coli* BL21star/pJDS10 225

8.5 Molecular biology 226

8.5.1 Construction of plasmids pJDS02, pJDS04, pJDS05, pJDS06 and pJDS08 226

8.5.2 Cloning of *SgbnB* into pJDS05 and *SgbnA* into pJDS06 228

8.5.3 Construction of plasmids pJDS09-pJDS11 from Duet vectors 229

9. References 233

Appendices 244

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree apart from the background material in Appendix 1 which was previously submitted for the degree of MChem. Chemical Biology (2011).

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

Work in Chapter 2 on chiral HPLC was performed by Dr. Lijiang Song and NMR data analysis of compound **80** was completed in collaboration with Prof. Greg Challis and Dr. Lijiang Song

Work in Chapter 6 on the feeding experiments with *Streptomyces coelicolor* M1152 was performed in collaboration with Dr. Christophe Corre, and the EMSAs were performed by Shanshan Zhou

Parts of this thesis have been published by the author:

'Discovery of a family of γ-aminobutyrate ureas via rational derepression of a silent bacterial gene cluster' J. D. Sidda, L. Song, V. Poon, M. Al-Bassam, O. Lazos, M. J. Buttner, G. L. Challis and C. Corre *Chemical Science* 2014, **5**, 86-89

'Gamma-butyrolactone and furan signalling systems in *Streptomyces*' J. D. Sidda and C. Corre *Methods in Enzymology* 2012, **517**, 71-87

John David Sidda

Acknowledgements

Firstly I would like to express my gratitude to my supervisor, Dr. Christophe Corre for all of his enthusiasm, patience and support over the last 3 ½ years and for giving me the opportunity to develop such an interesting and diverse project. I also greatly appreciate his sense of humour and the many interesting discussions we have had over coffee – they have definitely made the last 3 years easier!

Secondly, I would like to thank Prof. Greg Challis and Dr. Manuela Tosin for all the useful discussions and advice, guidance and enthusiasm throughout this project. For their technical advice and helpful comments in the lab, I wish to thank Sarah, Orestis, Yuki, Dean, Lauren, Lona, Samantha, James, Judith, Elena, Jenner, Dani and all of the other people who have made working in the Chemical Biology Research Facility such a rewarding experience.

I am especially grateful to Lijiang Song, who assisted with the structure elucidation of gaburedin A and for his time spent pouring over NMR data with me, for running the chiral HPLC samples and for allowing me use the MaXis. David Fox's advice with planning the synthesis of the authentic standard of gaburedin A was also a great help. My thanks are extended to Mark Buttner and Mahmoud Al-Bassam who kindly donated some of the strains I have worked on. I also wish to thank Orestis and Christophe for making other strains and Shanshan for undertaking the *in vitro* work discussed in Chapter 6. I also wish to thank Phil Aston and Ivan Prokes for all of their training and technical support with the mass spec and NMR aspects of the project.

I would also like to thank Kathryn and Vincent for being my travel buddies, and Sophie for keeping me sane in the lab for the final six months. Particularly I would like to

thank Anne Smith for all of her help with the biological aspects of the work (and for countless hours spent with the autoclaves) and David Withall, who was always available to assist with any chemistry (mainly HPLC) problems.

I also wish to thank my parents, Gillian and David and my sister Rachel, for all of their support throughout my PhD studies. I would also like extend my greatest thanks to Leo Bowsher, Matt Lougher, Peter Harrison and Jon Peters for all of their help and their friendship throughout my time at Warwick. Finally, I wish to thank Kerrie Smith – your support means more than I can tell you. I would not have been able to enjoy as much, or achieve so much, in the last three years without you.

List of Figures

Figure 1.1 – A: *Streptomyces* life cycle schematic (Flärdh and Buttner) and B: *Streptomyces venezuelae* ISP5230 grown on mannitol/yeast extract medium, MYM (image obtained during the present study)

Figure 1.2 – *Streptomyces* metabolites of historical and industrial significance

Figure 1.3 – Anticancer (doxorubicin and bleomycin) and immunosuppressant (rapamycin) drugs derived from *Streptomyces* discovered by screening programs

Figure 1.4 – Some examples of *Streptomyces* metabolites produced in response to altered culture conditions. For example, jadomycin B and validamycin A are each produced by their host organisms under conditions of heat and ethanol stress. Methylenomycin is produced in acidic pH. Naphthomycin K and chaxalactin were discovered by application of the OSMAC approach

Figure 1.5 – Examples of known elicitors of secondary metabolism in *Streptomyces*. The siderophore desferrioxamine has been implicated in interspecies signalling. The role of A-factor in antibiotic biosynthesis has been extensively studied and will be discussed later. Lincomycin, triclosan and ARC2 are proposed to affect secondary metabolism by exerting stress upon the target bacteria

Figure 1.6 – Strategies for phenotype-based natural product discovery

Figure 1.7 – Genome mining strategies for natural product discovery. Bioinformatics can be used to predict natural product structures assembled by gene clusters, or by specific enzymes (A and B). Genes may be expressed in a heterologous host (C) or knocked out (D) in order to deduce their roles in natural product biosynthesis. Alternatively, if activators/repressors can be identified, they may also be overexpressed/deleted to increase expression of the biosynthetic cluster of interest (E and F)

Figure 1.8 – Examples of *Streptomyces* metabolites, the structures of which were predicted by bioinformatics comparisons of their biosynthetic gene clusters with previously characterised ones

Figure 1.9 – Examples of *S. coelicolor* metabolites discovered by heterologous expression of genes (AHFCAs and epi-isozizaene) and by gene knockouts (germicidin B)

Figure 1.10 – Examples of Actinomycete metabolites discovered by overexpression of a gene encoding for an activator (stambomycin B **28**) and by deletion of a genes encoding for a transcriptional repressors (kitasetaline **29** and coelimycin P1 **30**)

Figure 1.11 – Mode of action of TetR. TetR binds to the *tetA-tetR* promoter, repressing transcription of *tetA* (A). When tetracycline is present inside the cell, it binds to the ligand binding domain of TetR (B). This leads to dissociation of TetR from its DNA target (C) allowing *tetA* to be transcribed (D); TetA then exports tetracycline out of the cell

Figure 1.12 – Examples of structural diversity of ligands known to interact with TetR proteins

Figure 1.13 – Mode of action of ArpA and the regulatory cascade activating streptomycin biosynthesis upon A-factor binding to ArpA. ArpA prevents transcription of the activator *adpA* (pathway A), until A-factor is detected, allowing for transcription of *adpA* and subsequent activation of streptomycin biosynthesis (pathway B)

Figure 1.14 – Structures of known *Streptomyces* signalling molecules produced by AfsA or homologous proteins. The absolute stereochemistry of Gräfe factors and factor I have not been reported

Figure 1.15 – Stereochemistry of γ-butyrolactones varies at carbons 2, 3 and 6. The stereochemistry is 2*R*, 3*R* for all the examples discussed in the text (A-factor, SCBs and VBs)

Figure 1.16 – Examples of other butyrolactones isolated from *Streptomyces* species. Stereochemistry at position 3 is *R* in all cases where the structure have been elucidated fully, and others have derivatised alkyl chain substituents

Figure 1.17 – A-factor and A-factor analogues produced by *E. coli* when expressing *afsA* as proposed by Kato and co-workers

Figure 1.18 – Examples of genetic arrangement *arpA-afsA* like homologues in different *Streptomyces* signalling molecule systems. The diversity of structures produced by each system arise from the action of the biosynthetic enzymes shown in yellow/blue

Figure 1.19 –Example of electrophoretic mobility shift assay (EMSA)

Figure 1.20 – Examples of ARE consensus sequences from operator regions of transcriptional activators in gene clusters from different *Streptomyces* species. Conserved nucleotides are highlighted in yellow

Figure 1.21 – Signalling cascade involved in methylenomycin biosynthesis. ARE sequences are shown in pink. Stimulatory events are denoted by blocked arrowheads, inhibitory events are denoted by circular arrowheads

Figure 1.22 – Analogous *mmfLHP* systems identified in *S. avermitilis* and *S. venezuelae*. The *mmfLHP* system in *S. venezuelae* is located adjacent to the cryptic *gbnABC* cluster. AREs are highlighted in pink, and *S. coelicolor* and *S. venezuelae* AREs compared with the general consensus proposed by Folcher and co-workers are shown are shown. W = A/T, Y = C/T

Figure 1.23 –Secondary metabolite gene clusters identified in *S. venezuelae* by antiSMASH (October 2014)

Figure 1.24 – Proposed regulatory mechanism involved in regulation of the cryptic *gbn* cluster

Figure 2.1 – Panel A: genetic arrangement of the *gbn/sgn* cluster in *S. venezuelae*. Panel B: Base peak chromatograms of metabolites extracted at pH 3 from the *Streptomyces venezuelae* wild type (red trace) and *gbnR* mutant (blue trace) after being grown on SMMS for 3 days

Figure 2.2 – Extracted ion chromatograms for *m/z* = 247.0, 261.0, 278.9, 292.9 and 294.9 for metabolites **80-85** extracted from the *Streptomyces venezuelae* wild type (green trace), *gbnR* mutant (blue trace) and *sgnL-P* mutant (purple trace) at pH 3 after being grown on SMMS for 3 days

Figure 2.3 – Extracted ion chromatograms *m/z* = 247.0, 261.0, 278.9, 292.9 and 294.9 for metabolites **80-85** extracted from the *Streptomyces venezuelae gbnR* at pH 3 after being grown on SMMS for 24, 48 and 72 h

Figure 2.4 – UHR-LC-MS trace of metabolites extracted from *S. venezuelae gbnR::apra* mutant. Extracted ion chromatogram corresponds to *m/z* = 295.1, 261.1, 247.1, 279.1, 293.1 for metabolites **80-85**. The mass spectrum of compound **80** is also shown, with molecular formulae for molecular and fragment ions

Figure 2.5 – HPLC traces from purification of compound **80** from 300 μL injection of crude *S. venezuelae gbnR::apra* extract. The peak collected in red during the first round (left panel) was collected and re-injected (right panel)

Figure 2.6 – Proton NMR spectrum and assignments of gaburedin A (**80**) in d_6-DMSO (700 MHz). 2D correlations are also shown (COSY correlations in bold bonds, HMBC correlations denoted by red arrows)

Figure 2.7 – Extracted ion chromatogram for *m/z* = 323.0 for the intermediate **90** and extracted ion chromatograms *m/z* = 294.9, corresponding to the authentic standard **91** and proposed structure of gaburedin A **80** in metabolites extracted from *S. venezuelae gbnR* mutant after 3 days (upper panel). Mass spectra of the **90**, **91** and **80** are also shown (lower panel)

Figure 2.8 – 1H NMR spectra in d_6-DMSO of **80** isolated from *S. venezuelae* grown for 3 days on SMMS (upper spectrum) and authentic standard of **91** (50-0 MHz, lower spectrum). 2D COSY and HMBC correlations for **91** are also shown

Figure 2.9 – Extracted ion chromatograms for *m/z* = 247.0 for gaburedin C **83** for metabolites extracted from LC-MS analysis of the metabolites extracted at pH 3 from the *S. venezuelae gbnR::apra* strain when grown on SMMS for 3 days with different concentrations of L-valine added to the media

Figure 2.10 – Extracted ion chromatograms for *m/z* = 247.0 for metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS enriched with 20 mM L-valine and 20 mM d_8-DL-valine (upper panels) and mass spectra of the unlabelled gaburedin C (lower panels)

Figure 2.11 – Extracted ion chromatograms for *m/z* = 260.9 for metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS and SMMS enriched with 20 mM L-leucine and 20 mM isoleucine. Culture extracts were diluted by factor of 10 prior to LC-MS analysis

Figure 2.12 – Extracted ion chromatograms for *m/z* = 271.0, corresponding to d_{10}-gaburedin B (**81**) and d_{10}-gaburedin C (**82**) extracted from the *S. venezuelae gbnR::apra* strain after growing on SMMS for 3 days upon addition of 10 mM d_{10}-leucine (black trace) 10 mM d_{10}-isoleucine (red trace). Extracted ion chromatogram for *m/z* = 261.0, corresponding to unlabelled

gaburedins B and C extracted from *S. venezuelae gbnR::apra* strain after growing on SMMS for 3 days (blue trace). Mass spectra of the labelled gaburedins B and C are also shown

Figure 2.13 **–** Extracted ion chromatograms for *m/z* = 247.0, 261.0, 279.0, 293.0 and 295.0 for gaburedins A-F in metabolites extracted from *S. venezuelae gbnR::apra* grown on SMMS and SMMS enriched with 5 mM and 10 mM *N*-acetyl-L-cysteine. The mass spectrum and proposed structure for gaburedin F derived from *N*-acetyl-L-cysteine (**85**) are also shown

Figure 2.14 **–** Extracted ion chromatograms for *m/z* = 247.0 corresponding to metabolites **92** and **93** extracted from the *S. venezuelae gbnR* mutant when grown on SMMS enriched with 20 mM L-valine and 20 mM D-valine (blue traces) and extracted ion chromatograms for *m/z* = 295.0 for metabolites **94** and **95** extracted from the *S. venezuelae gbnR* mutant when grown on SMMS enriched with 20 mM L-phenylalanine and 20 mM D-phenylalanine (red traces)

Figure 2.15 **–** Extracted ion chromatograms from chiral HPLC-MS analysis of metabolites extracted from the *S. venezuelae gbnR::apr* mutant grown on 20 mM L- or D-phenylalanine for *m/z* = 295, corresponding to L- and D- enantiomers of gaburedin A, (panels **a** and **b**, respectively). Coinjection of extracts **a** and **b** (panel **c**) confirmed that the gaburedins **94** and **95** produced upon supplementation with L- or D-phenylalanine are different; as gaburedin **94** (panel **a**) and its enantiomer **95** (panel **b**) have different retention times

Figure 2.16 **–** Proposed structures of gaburedins A-F (**80**-**85**)

Figure 2.17 **–** Examples of natural products and synthetic candidate drug molecules containing ureido bridges

Figure 3.1 – General structures of gaburedins **1**-**6** characterised in Chapter 2. Chapter 3 will focus on the origin of the GABA-like moiety (blue) and ureido bridge. Chapter 4 will focus on the diversity of substrates with different amino acid side chains (red) and different groups in place of the aminoacyl carboxylic acid group (green) which can be incorporated into gaburedins

Figure 3.2– *gbnABC* clusters identified by BLAST searches

Figure 3.3 – A) Conservation of amino acid residues involved in PLP-binding domains in GbnA and orthologues B) C-terminal motifs of GbnA and orthologues, illustrating the IXYPPGXP consensus

Figure 3.4 – Structures of L-glutamic acid, L-ornithine, L-lysine and L-arginine. Conserved structural motif is highlight in red

Figure 3.5 – Sequence alignment of GbnB and orthologues, illustrating the acyl-activating motif TSGTSXXPK identified during conserved domain (CD) searches

Figure 3.6 – Agarose gel of PCR products used to verify the correct construction of the *S. venezuelae gbnB/gbnR* mutant strain. The templates used were genomic *gbnB::apra, gbnR::scar* mutant (B, R) and cosmids SV4 F01 containing the wild-type *gbn-sgn* gene clusters (WT) and SV4 F01 *gbnR::scar* (R). Different screening primers were used to amplify the genes shown

Figure 3.7 – Extracted ion chromatograms *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 of gaburedins A-F (**80**-**85**) extracted at pH 3 from the *Streptomyces venezuelae* wild-type (red trace), *gbnR* mutant (blue trace) and *gbnB/gbnR* mutant (green trace) after being grown on SMMS for 3 days

Figure 3.8 – Extracted ion chromatograms for *m/z* = 292.9 (blue) for gaburedin F (**85**) , *m/z* = 277.0 (green) for new gaburedin **106** and *m/z* = 233.0 (orange) for new gaburedin **107** for metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS enriched with 10 mM L-glutamate, 10 mM GABA and unmodified SMMS. Mass spectra of the new gaburedin **106** containing L-glutamate and gaburedin **107** containing GABA are shown

Figure 3.9 – Mass spectra of gaburedins **80**, **106** and **107** from metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS and SMMS enriched with 10 mM [1-^{13}C]-L-glutamic acid illustrating the only labelled gaburedin is gaburedin **106**, the product of the reaction of [1-^{13}C]-L-glutamic acid with the proposed biosynthetic intermediate **109/111**

Figure 3.10 – Mass spectra of gaburedin A (**80**) in metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS (top panel) and SMMS enriched with 10 mM Na$H^{13}CO_3$

(bottom panel), illustrating that the ^{13}C label from $NaH^{13}CO_3$ is not retained in gaburedin biosynthesis

Figure 3.11 – Biosynthetic route to gaburedins, showing the proposed incorporation of ^{13}C labels upon feeding the *gbnR::apra* strain with [U-^{13}C]-L-glutamic acid (top panel). Mass spectra of gaburedins A **80**, **106** and **107** from metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS and SMMS enriched with 10 mM [U-^{13}C]-L-glutamic acid (bottom panel), showing retention of ^{13}C labels in gaburedins identified in this culture extract

Figure 3.12 – Extracted ion chromatograms for *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue) for gaburedins A-F (**80**-**85**) extracted from the *S. venezuelae gbnR* mutant when grown on SMMS enriched with 1 mM L-arginine, 1 mM lysine and 1 mM L-ornithine. The structures and extracted ion chromatograms for the $[M+H]^+$ species of gaburedins **127**-**129** expected if these amino acids attack intermediate **109** are also shown

Figure 3.13 – Extracted ion chromatograms for *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue) for gaburedins A-F (**80**-**85**) and *m/z* = 244.0 (red trace) for new gaburedin-like compound **130** extracted from the *S. venezuelae gbnR* mutant when grown on SMMS supplemented with 20 mM L-ornithine (top panel) and mass spectrum and proposed structure of new gaburedin **130**

Figure 3.14 – Extracted ion chromatogram for *m/z* = 247.0 corresponding to gaburedin D (**83**) and new gaburedin **138** corresponding to incorporation of L-norvaline from the metabolites extracted from the *gbnR* mutant after growing on SMMS + 1 mM L-norvaline for 3 days

Figure 3.15 – Extracted ion chromatograms for *m/z* = 276.1 of gaburedin **145** which would be derived from L-glutamine (red trace), *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 of the gaburedins A-F (**80**-**85**, blue trace) in metabolites extracted from the *Streptomyces venezuelae gbnR* strain after being grown on SMMS containing 20 mM L-glutamine for 3 days

Figure 3.16 – Extracted ion chromatograms for metabolites extracted from the *S. venezuelae gbnR* mutant after growing on SMMS + 10 mM L-aspartate for 3 days. Gaburedins A-F (**80**-**85**) *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) and *m/z* = 263.0 (green trace) for new

gaburedin **152** that would be expected for incorporation of L-aspartate into gaburedins via attack of intermediate **109**

Figure 3.17 **–** Extracted ion chromatograms for gaburedins **1-6** *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) *m/z* = 203.0 (black trace) and *m/z* = 247.0 (pink trace) for new gaburedins **156** and **164** observed in the metabolites extracted from the *gbnR* mutant after growing on SMMS + 1 mM *n*-butylamine and SMMS + 10 mM methyl 4-aminobutyrate for 3 days

Figure 4.1 – General structure of gaburedins identified thus far. Chapter 4 will focus on the diversity of substrates with different amino acid side chains (red) and different groups in place of the aminoacyl carboxylic acid group (green) which can be incorporated into gaburedins

Figure 4.2 **–** Molar concentrations of proteinogenic amino acids in SMMS

Figure 4.3 **–** Extracted ion chromatograms for gaburedins A-F (**80-85)** *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue trace) and new gaburedin **165** *m/z* = 219.0 (green trace) in metabolites extracted at pH 3 from the *Streptomyces venezuelae gbnR* strain after being grown on SMMS + 20 mM L-alanine for 3 days. The mass spectrum of new gaburedin **165** is also shown

Figure 4.4 **–** Extracted ion chromatograms for gaburedins A-F (**80-85**) *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) and new gaburedins (green traces) of metabolites extracted at pH 3 from the *Streptomyces venezuelae gbnR* strain after being grown on SMMS + 20 mM L-proline, glycine, L-serine and L-threonine for 3 days

Figure 4.5 **–** Extracted ion chromatograms for gaburedins A-F (**80-85**) *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) and *m/z* = 251.0 for the expected new gaburedin **170** that would be expected from incorporation of L-cysteine (green traces) in metabolites extracted at pH 3 from the *Streptomyces venezuelae gbnR* strain after being grown on SMMS and SMMS supplemented with L-cysteine

Figure 4.6 **–** Extracted ion chromatograms for *m/z* = 262.0 of gaburedin **171** (green trace) and *m/z* = 277.0 of the gaburedin **106** from incorporation of L-glutamate (orange trace) in metabolites extracted at pH 3 from the *Streptomyces venezuelae gbnR* strain after being grown

on SMMS containing 20 mM L-asparagine and SMMS + 20 mM L-glutamate (top panels), and mass spectra of gaburedins **171** and **106**

Figure 4.7 **–** Extracted ion chromatograms for gaburedins A-F (**80-85**) *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) and new gaburedin **173** with *m/z* = 311.0 (red trace) expected for incorporation of L-tyrosine, and UHR mass spectrum of this new gaburedin **173**

Figure 4.8 **–** Mass spectra new gaburedins **156**, **175**, **176** and **103** containing *n*-butylamine, isobutylamine, isopentylamine and cyclohexylamine displaying fragments ions at *m/z* = 104.0 – corresponding to GABA – that are common to all four amine-derived gaburedins

Figure 4.9 **–** Mass spectra of new gaburedins containing isobutylamine, isopentylamine and cyclohexylamine (panels a, c and e) and the species corresponding to $[M+14]^+$ ions for additional metabolites present in the extracts upon feeding with these precursor molecules (panels b, d and f)

Figure 4.10 **–** Structures of amino acid analogues fed to the *gbnR::apra* mutant to probe diversity of aminoacyl compounds with different C-terminal groups which can be incorporated into gaburedins

Figure 4.11 **–** Extracted ion chromatograms for *m/z* = 281.0 (top panel) and 233.0 (bottom panel) and mass spectra of the new gaburedins **183** and **184** that are produced by the *gbnR* mutant when grown on SMMS + 1 mM L-phenylalaninol (top panel) and 1 mM L-valinol (bottom panel)

Figure 4.12 **–** Extracted ion chromatograms for gaburedin A *m/z* = 294.9 (red trace) and *m/z* = 309.0 (black trace) expected for incorporation of L-phenylalanine-*O*-Me ester in the metabolites extracted after growing the *gbnR* mutant on SMMS + 10 mM L-phenylalanine-*O*-methyl ester. Mass spectra of gaburedins **80**, **185** and **186** are also shown

Figure 4.13 **–** Extracted ion chromatograms gaburedins A-F (**80**-**85**) *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) and *m/z* = 289.9 (red traces) for L-ala-ala incorporation into a new gaburedin **187** and *m/z* = 218.8 (black traces) for gaburedin **165** arising from incorporation of L-alanine. Mass spectra of **165** and **187** also shown, with the molecular and fragment ions for **187** highlighted in red

***Figure 4.14* –** Library of gaburedins produced during this study

Figure 4.15 – Schematic to illustrate the substrate tolerance of GbnAB leading to conserved GABA moiety (blue) and diversity of different amino acids and amino acid analogues that can be incorporated into gaburedins *in vivo* (orange, red and green)

***Figure 4.16* –** Schematic to show incorporation of *N*-acetyl-L-cysteine into the proposed structure for gaburedin F (**85**) either directly or via the cysteine derived gaburedin **170**. Both routes would require very poor nucleophiles attacking either the intermediate **109**, or acetylation occurring at the poorly nucleophilic ureido nitrogen of **170**. The possible structure of gaburedin F if derived from *S*-acetylcysteine is highlighted

Figure 5.1 – *gbnABC* clusters from *Streptomyces venezuelae*, *Salmonella enterica* and *Streptococcus mutans* GS-5. Only the metabolic products of the *gbnABC* system in *S. venezuelae* have been reported so far

Figure 5.2 – Schematic to show the steps in creating pET151-derived plasmids using the *gbnA* and *gbnB* genes from *S. venezuelae*. Primers were designed to include a 5'-CACC overhang to allow topoisomerase-directed cloning of *gbnA* and *gbnB* into pET151

Figure 5.3 – 1% agarose gel of PCR products obtained for DNA regions amplified using control screening primers F/F' specific for *gbnR*, and TOPO cloning primers C/C' specific for *gbnA*. The bands for the PCR products obtained with primers C/C' were shorter than expected. Colours have been inverted for clarity

Figure 5.4 – Diagram to illustrate TOPO cloning of synthetic *S. venezuelae gbnB, Salmonella enterica seag_b2133* and *Streptococcus mutans* GS-5 *smugs5_01685* into pET151. Restriction digests using the enzymes *Nde*I and *Pst*I were used to confirm the nature of plasmids pJDS06 and pJDS02 and pJDS04, respectively. These plasmids were also analysed by PCR using primer pairs J/J', K/K' and L/L' and the inserted genes were sequenced using the T7 primer pair

Figure 5.5 – Metabolites extracted from culture supernatants of *E. coli*/pJDS02, pJDS04 and pJDS06 grown in LB supplemented with 10 mM GABA. Base peak chromatograms (red traces) and xtracted ion chromatograms for gaburedins A-F (**80-85**) (*m/z* = 247.0, 261.0, 279.0,

292.9 and 294.9) are also shown (black traces). The large peak at t = 14.0 min has *m/z* = 261.0, but is unrelated to gaburedins

Figure 5.6 – 10% SDS-PAGE gels to show comparison of soluble and insoluble proteins extracted from *E. coli*/pJDS02 carrying synthetic *seag_gbnB*. The only lane with an intense band for a protein of the expected mass of 60.2 kDa is the lane loaded with the insoluble protein fraction

Figure 5.7 – Schematic of the strategy used to sub-clone synthetic *gbnA* into pJDS06 carrying synthetic *gbnB* and sub-cloning of synthetic *gbnB* into pJDS05 carrying synthetic *gbnA*

Figure 5.8 – Agarose gel of *Nde*I restriction digests of plasmids after ligations to introduce *gbnB* into pJDS05. Lanes giving examples of the different outcomes of the ligations are annotated

Figure 5.9 – Vectors pET-Duet and pACYC-Duet used for cloning of synthetic *S. venezuelae gbnA*, *gbnB* and *gbnC*

Figure 5.10 – Cloning strategy used to generate plasmids pJDS09-pJDS11 containing synthetic *S. venezuelae gbnA*, *gbnB* and *gbnC*. Step (i) – amplification of each gene of interest by PCR; step (ii) – digestion using specifically selected restriction enzymes to leave appropriate overhangs; step (iii) – ligation of DNA fragments with complementary overhangs

Figure 5.11 – Base peak chromatograms (pink traces) and extracted ion chromatograms for gaburedins A-F (**80-85**) *m/z* = 247.0, 261.0, 279.0, 292.9, 294.9 (blue traces) of metabolites extracted from culture supernatants of *E. coli* BL21star/pJDS10 grown without IPTG and with 1 mM IPTG in LB with and without addition of 1 mM L-glutamic acid

Figure 6.1 – Schematic to demonstrate phenotype of *S. coelicolor mmyR* deletion mutant and proposed phenotype of *S. venezuelae gbnR* deletion mutant. Repressor binding sites (AREs) shown in pink. Sizes of genes not to scale

Figure 6.2 – UHR-MS data for SCB1 (**36**, top panel) and AHFCA5 (**53**, bottom panel) illustrating the typical mass spectra of 6-hydroxy GBLs and AHFCAs observed in electrospray MS in positive ion mode

Figure 6.3 **–** Extracted ion chromatograms for $[M+Na]^+$ ions of 6-hydroxy type GBLs *m/z* = 239.0, 253.0, 267.0 and 281.0 observed in metabolites extracted from *S. venezuelae gbnR* mutants. Mass spectra of these GBL-like molecules are also shown with the molecular formulae generated for the $[M+Na]^+$, $[M+H]^+$ and $[M-H_2O+H]^+$ ions

Figure 6.4 **–** Top: genetic organisation of the *scb-cpk* gene cluster as modified in *S. coelicolor* M1152. The gene encoding for the transcriptional repressor protein *scbR2* has been removed. Bottom: Extracted ion chromatograms for $[M+Na]^+$ ions *m/z* = 239.0, 253.0, 267.0 and 281.0 of 6-hydroxy type butyrolactones in the metabolites extracted from *S. coelicolor* M1152 and *S. venezuelae gbnB/R*. Compound denoted by * has *m/z* = 267.0 but is not related to GBLs

Figure 6.5 **–** UHR-MS of **197**-**199** extracted from *S. coelicolor* M1152 grown on different AlaMM media. From top to bottom: **197** without any labelling, **197** labelled upon feeding with d_8-DL-valine, **198** without any labelling, **198** labelled upon feeding with d_7-butyric acid, **199** without any labelling, and **199** labelled upon feeding with d_{10}-isoleucine

Figure 6.6 **–** UHR-MS of metabolites **193**, **194**, **195**, **196** and **200** extracted from *S. coelicolor* M1152 grown on different AlaMM media. **193** was labelled with 7 deuteriums from incorporation of d_7-butyric acid; **194** was labelled with 9 deuteriums from incorporation of d_{10}-isoleucine; **195** was labelled with 9 labels from incorporation of d_{10}-leucine; **196** was labelled with 5 labels from incorporation of d_5-propionic acid; and **200** was labelled with 7 deuterium labels from incorporation of d_7-butyric acid

Figure 6.7 **–** Extracted ion chromatograms for $[M-H_2O+H]^+$ ions of AHFCA species present in the metabolites extracted from the *S. venezuelae gbnB/gbnR*, *gbnR*, *sgnL* and wild type. Chromatograms correspond to *m/z* = 195.0 (blue traces), 209.0 (black traces) and 223.0 (red traces). Peaks denoted by * are unrelated to AHFCA compounds. EICs for *m/z* = 167.0, 181.0 and 237.0 have been omitted for clarity

Figure 6.8 **–** Mass spectra of the AHFCAs 5-8 (**201**-**204**) identified in the *S. venezuelae gbnR* deletion mutants when grown on modified SMMS. AHFCA5 is labelled with 7 deuteriums from d_7-butyric acid; AHFCA6 is labelled with 9 deuteriums from d_{10}-leucine; AHFCA7 is

labelled with 5 labels from d_5-propionic acid; and AHFCA8 is labelled with 7 deuteriums from d8-valine. $[M+Na]^+$ ions are highlighted in red

Figure 6.9 **–** AHFCAs produced by *S. venezuelae gbnR* deletion mutants as demonstrated by incorporation of the specific precursor molecules shown, and UHR-MS spectra of AHFCA8 (**204**) unlabelled (top panel) and AHFCA8 labelled with 7 deuteriums from incorporation of d_8-valine (bottom panel) and molecular formulae generated for the $[M-H_2O+H]^+$ and $[M+Na]^+$ and ions are shown

Figure 6.10 **–** Mass spectra and proposed structures of **205** produced upon feeding the *gbnB/R* strain with d_{10}-isoleucine (top panel) and **206** produced upon feeding the *gbnB/R* strain with d_{10}-leucine (bottom panel)

Figure 6.11 **–** Extracted ion chromatograms for $[M-H_2O+H]^+$ ions of AHFCA species present in the metabolites extracted from the *S. venezuelae gbnB/gbnR* with no AHFCA added and with AHFCAs 5, 6 and 7 co-injected. Chromatograms correspond to *m/z* = 195.0 (blue traces), 209.0 (black traces) and 223.0 (red traces)

Figure 6.12 **–** Extracted ion chromatograms *m/z* = 167.0, 195.0 and 223.0 for $[M-H_2O+H]^+$ ions of **212-214** present in the metabolites extracted from the *E. coli* BL21star/pJDS8 strain and *E. coli* BL21star/pJDS8 strain induced with IPTG. A synthetic standard of AHFCA2 (**32**) was run in parallel. Mass spectra of AHFCA **212-214** are also shown

Figure 6.13 **–** Clustal Omega alignments of SgnR (SVEN_4182), MmfR and homologues in other *Streptomyces* species. The amino acid residues identified as interacting with the AHFCA ligand are highlighted in yellow. The key tyrosine residues identified are highlighted in cyan

Figure 6.14 **–** a) Active site residues of MmfR interacting with AHFCA2 from solved crystal structure b) Active site residues of SgnR proposed to interact with AHFCA2 from SgnR structure generated using Phyre2 c) overlaid active sites residues of MmfR (orange) and SgnR (cyan) showing that the active site residues are predicted to be conserved between MmfR and SgnR. Images generated using PyMOL with views restricted to 6Å around the AHFCA2 ligand

Figure 6.15 **–** Extracted ion chromatograms *m/z* = 292.9, 279.0, 247.0, 261.0, 249.9 for gaburedins A-F (**80**-**85**) in metabolites extracted from *S. venezuelae sgnL-P::apra* strain upon addition of 10 μM, 100 μM and 400 μM AHFCA5

Figure 6.16 **–** Extracted ion chromatograms *m/z* = 239.0, 253.0, 267.0, 281.0 for SCBs in metabolites extracted from *S. venezuelae sgnL-P::apra* strain upon addition of 10 μM, 100 μM and 400 μM AHFCA5, with a positive control experiment adding 400 μM AHFCA5 to the *gbnR* mutant. The compound denoted by the * has *m/z* = 267 and is unrelated to SCBs

Figure 6.17 **–** Extracted ion chromatograms *m/z* = 292.9, 279.0, 247.0, 261.0, 249.9 for gaburedins A-F (**80**-**85**) in metabolites extracted from *S. venezuelae gbnR::apra* strain upon addition of different AHFCAs to the growth media

Figure 6.18 **–** Extracted ion chromatograms *m/z* = 292.9, 279.0, 247.0, 261.0, 249.9 for gaburedins A-F (**80**-**85**) in metabolites extracted from *S. venezuelae sgnL-P::apra* strain upon addition of different AHFCAs to the growth media

Figure 6.19 **–** Extracted ion chromatograms *m/z* = 292.9, 279.0, 247.0, 261.0, 249.9 for gaburedins A-F (**80**-**85**) in metabolites extracted from *S. venezuelae* wild-type strain upon addition of different AHFCAs to the growth media

Figure 6.20 **–** Proposed signalling cascade involved in gaburedin biosynthesis. AREs are shown in pink. Inhibitory events are denoted by circular arrowheads

Figure 6.21 **–** Gel shift assay with SgnR protein (200 nM) and 28-bp sequence which includes the *gbnA* ARE sequence. Concentrations of AHFCA6 added to each sample shown above each lane. Similar experiments with the *sgnR-sgnL* and *gbnR-sgnH* promoters have also been performed (S. Zhou)

Figure 7.1 **–** *mmfR-mmfLHP-mmyR* gene clusters in other *Streptomyces* species shown with 10 kb upstream/downstream of the regulatory mini-cluster. AREs identified are shown, as are transcriptional activators identified in the adjacent clusters

List of Schemes

Scheme 1.1 – Specific incorporation of ^{13}C labels from acetate, isovalerate and glycerol precursor molecules into virginiae butenolide-A (VB-A)

Scheme 1.2 – Biosynthetic route to A-factor as proposed by Kato and co-workers

Scheme 1.3 – General scheme of proposed steps in GBL biosynthesis, showing a conserved pathway to 6-keto type GBLs, where the stereochemistry of carbon 3 is determined by BprA (or homologous proteins), then the absolute configuration of carbons 2 and 6 are determined by either SCO6264/JadW3 or BarS1 enzymes

Scheme 1.4 – Proposed biosynthetic route to AHFCAs isolated from *Streptomyces coelicolor* W74, showing the proposed origin isobutyryl-CoA being incorporated into AHFCA3

Scheme 1.5 – General structures of *Streptomyces* signalling molecules which are proposed to share an intermediate butenolide phosphate. Labels of enzymes catalysing each step are shown, although the specificity of each type (AfsA, MmfL, ScbA, for example) is presumably determined by the length of the alkyl chain in the β-ketothioester precursors

Scheme 2.1 – Synthetic route to gaburedin **1** from *mono*-methyl glutaric acid

Scheme 3.1 – Schematic to show incorporation of L-glutamate and GABA into gaburedins. Decarboxylation of L-glutamate by GbnA produces GABA, which is incorporated into gaburedins

Scheme 3.2 – Proposed biosynthetic route to gaburedins. L-glutamate is decarboxylated to give GABA, which may proceed by one of two possible intermediates **109** or **111**, to furnish gaburedins

Scheme 3.3 – Proposed biosynthetic origin of ureido-linked dipeptides in syringolin biosynthesis as proposed by Imker and co-workers

Scheme 3.4 – Biosynthetic route to gaburedin analogues that would be produced if GbnA decarboxylated L-ornithine, L-lysine or L-arginine, and if the resulting carbamoyl amines are activated by GbnB

Scheme 3.5 **–** Biosynthetic route to gaburedin analogues that would be produced if GbnA decarboxylated L-norvaline, L-glutamine or L-aspartate and the resulting carbamoyl amines are activated by GbnB

Scheme 3.6 **–** Biosynthetic route to gaburedin analogues that would be produced if GbnA decarboxylated any of the putative precursors (left hand column) and the resulting amines attack **109** to furnish gaburedin analogues

Scheme 3.7 **–** Biosynthetic route to gaburedin analogues that would be produced if carbamoylated *n*-butylamine or methyl 4-aminobutyrate were activated by GbnB (top panels) and structures of gaburedins containing *n*-butylamine or methyl 4-aminobutyrate via reaction with intermediate **109** (bottom panel)

Scheme 6.1 – GBL and AHFCA biosynthesis proceed via a shared butenolide phosphate intermediate. The downstream enzymes (for example MmfH, MmfP) determine the nature of the final signalling molecules produced

Scheme 6.2 **–** Biosynthetic route to SCBs which would give rise to $[M+Na]^+$ ions with *m/z* = 239.0, 253.0, 267.0 and 281.0. The number of chain extensions in fatty acid biosynthesis is also shown

Scheme 6.3 **–** Proposed structures of the GBLs and AHFCAs produced by *S. venezuelae gbnR* deletion mutants. The same series of butenolide phosphate intermediates **207**-**211** are proposed to be utilized in GBL and AHFCA biosynthesis

List of Tables

Table 1.1 – Examples of biosynthetic potential of some sequenced *Streptomyces*

Table 1.2 – Examples of ArpA-like proteins and their cognate ligands and antibiotic biosynthetic pathways they regulate

Table 1.3 – Effect of deletion of genes encoding for ArpA-like repressors on antibiotic production in host strains

Table 2.1 – UHR-MS assignments of molecular and fragment ions for metabolites **80-85**

Table 3.1 – Sequence homology of *S. venezuelae* GbnA, GbnB and GbnC and orthologous proteins present in other systems as revealed by BLAST searches. Numbers denoted in each column indicate number of amino acids, % identity, % similarity compared with *S. venezuelae* GbnA, GbnB and GbnC

Table 4.1 – Gaburedins identified and overproduced throughout this study

Table 6.1 – UHR-MS of SCBs1-8 identified in culture extract of *S. coelicolor* M1152 grown on AlaMM media supplemented with different stable isotope precursor molecules. Full UHR-MS characterisation shown in Appendix 5

Table 6.2 – UHR-MS of $[M+Na]^+$ ions of AHFCAs 5-10 identified in culture extract of *S. venezuelae gbnB/gbnR* grown on SMMS supplemented with different stable isotope precursor molecules

Table 8.1– Primers used for cloning *gbnA* and *gbnB* from *Streptomyces venezuelae* into pET151 for protein overproduction. Restriction sites are highlighted in red, and 5'-CACC overhangs are underlined

Table 8.2 – pET-151-derived plasmids constructed during this study

Table 8.3 – Conditions for ligations of *Sac*I digested products and plasmids

Table 8.4 – Primers used for construction and sequencing of plasmids pJDS09-pJDS11

Table 8.5 – Duet-derived co-expression vectors constructed during this study

Abbreviations

ACP – acyl carrier protein

ARC – antibiotic remodelling compound

AdpA/*adpA* – A-factor-responsive transcriptional activator protein/*gene*

AfsA/*afsA* – A-factor synthase protein/*gene*

AHFCA – 2-alkyl-4-hydroxymethylfuran-3-carboxylic acid

AMP – adenosine monophosphate

antiSMASH – antibiotics and secondary metabolite analysis shell

ARE – autoregulatory response elements

ArpA/*arpA* – A-factor receptor A protein/*gene*

ATP – adenosine triphosphate

BLAST – basic local alignment search tool

BPC – base peak chromatogram

BprA – butenolide phosphate reductase A

CD – circular dichroism

CoA – coenzyme A

COSY – Correlation Spectroscopy

CprB/*cprB* – *coelicolor* repressor protein/*gene* B

DAD – diode array detector

DHAP – dihydroxyacetone phosphate

DNA – deoxyribonucleic acid

DPPA – diphenylphosphoryl azide

EIC – extracted ion chromatogram

EMSA – electrophoretic mobility shift assay

ESI-MS – electrospray ionisation mass spectrometry

GABA – γ-aminobutyrate

GBL – γ-butyrolactone

HCT – high-capacity trap

HMBC – heteronuclear multiple bond connectivity

HPLC – high performance liquid chromatography

HSQC – heteronuclear single quantum coherence

ITPG – isopropyl β-D-1-thiogalactopyranoside

IR – infrared

LC-MS – liquid chromatography-mass spectrometry

MEME – multiple Em for elucidation of motifs

MIC – minimum inhibitory concentration

MMF – methylenomycin furan

MmfR/*mmfR* – methylenomycin furan repressor protein/*gene*

MmyR/*mmyR* – methylenomycin repressor protein/*gene*

MYM – mannitol-yeast medium

m/z – mass/charge ratio

NAD(P)H – nicotinamide adenine dinucleotide (phosphate)

NCBI – National Center for Biotechnology Information, USA

NMR – nuclear magnetic resonance

NRPS – non-ribosomal peptide synthetase

OSMAC – one strain, many compounds

PCR – polymerase chain reaction

PENDANT – polarisation enhancement during attached nucleus testing

pfam – protein family

Phyre2 – protein homology / analogy recognition engine

PKS – polyketide synthetase

PLP – pyridoxal 5'-phosphate

RiPP – ribosomally synthesised, post-translationally-modified peptides

RNA – ribonucleic acid

RT-PCR – reverse transcriptase-polymerase chain reaction

SARP – *Streptomyces* antibiotic regulatory protein

SCB – *Streptomyces coelicolor* butanolide

SCP1 – *Streptomyces coelicolor* plasmid 1

SFM – soya flour medium

SMMS – supplemented minimal medium solid

SRB – *Streptomyces rochei* butenolide

TetR/*tetR* – tetracycline resistance protein/*gene*

TLC – thin-layer chromatography

TOF – time-of-flight

t_R – retention time

UHR – ultra high resolution

UV – ultraviolet

VB – *virginiae* butanolides

XDR – extremely drug-resistant

Abstract

Streptomyces bacteria are prolific producers of secondary metabolites. These natural products find uses ranging from antibiotics and antifungals to immunomodulatory agents and pesticides. Biosynthesis of secondary metabolites is often tightly controlled; therefore novel strategies for activation of cryptic secondary metabolites are required.

The metabolites of the *Streptomyces venezuelae* wild type and a mutant strain lacking the ArpA-like transcriptional repressor GbnR have been compared, leading to the identification of the gaburedins – novel, ureido-linked dipeptides – in the *gbnR* mutant. A combination of *in vivo* precursor-directed studies has led to a proposed biosynthetic route to gaburedins. Metabolic profiling of a range of other mutant strains has identified *gbnB* as an essential gene in the *gbnABC* gene cluster that is responsible for gaburedin biosynthesis. Synthesis of an authentic standard of gaburedin A combined with chiral HPLC analysis of culture extracts has allowed the proposed structure of gaburedins to be confirmed. Bioinformatic analyses of the *gbnABC* cluster responsible for gaburedin biosynthesis have revealed orthologous systems in a range of other bacterial genera and efforts to reconstitute the *S. venezuelae gbnABC* pathway in *E. coli* have begun.

The regulatory mechanism controlling gaburedin biosynthesis in *S. venezuelae* has also been investigated, leading to the discovery of new AHFCA signalling molecules which have been shown to induce gaburedin biosynthesis in a mutant strain unable to produce AHFCAs. This work demonstrates for the first time that AHFCA signalling is involved in regulation of other natural products as well as the methylenomycin cluster from which AHFCAs were first identified.

Gaburedins represent the first example by which deletion of an *arpA*-like regulatory gene has been used as a strategy for de-repression of a biosynthetic pathway for cryptic natural product biosynthesis. The current project demonstrates that rational deletion of proposed regulatory genes is a powerful approach to natural product discovery.

“From so simple a beginning, endless forms most beautiful and most wonderful have been – and are being – evolved.”

– Charles Darwin

1. Introduction

1.1 *Streptomyces* – the natural product makers

Streptomyces are Gram-positive, soil-dwelling bacteria and belong to the large phylum of Actinobacteria, characterised by the high G+C content in their DNA.[1] *Streptomyces* undergo a complex life cycle in which a single spore germinates, growing into a branched, vegetative mycelium. Nutrient depletion is accompanied by development of aerial hyphae, leading ultimately to sporulation (Figure 1.1).[2-4] It has been widely reported that sporulation coincides with biosynthesis of antibiotics such as actinorhodin and streptorubin B in *Streptomyces coelicolor* A3(2) and production of 'spore pigments' for example the green pigment in the case of *Streptomyces venezuelae* (Figure 1.1).[5-6]

Streptomyces bacteria produce a vast array of compounds useful in medicine. Indeed, over 70% of clinically useful antibiotics derive from compounds first isolated from

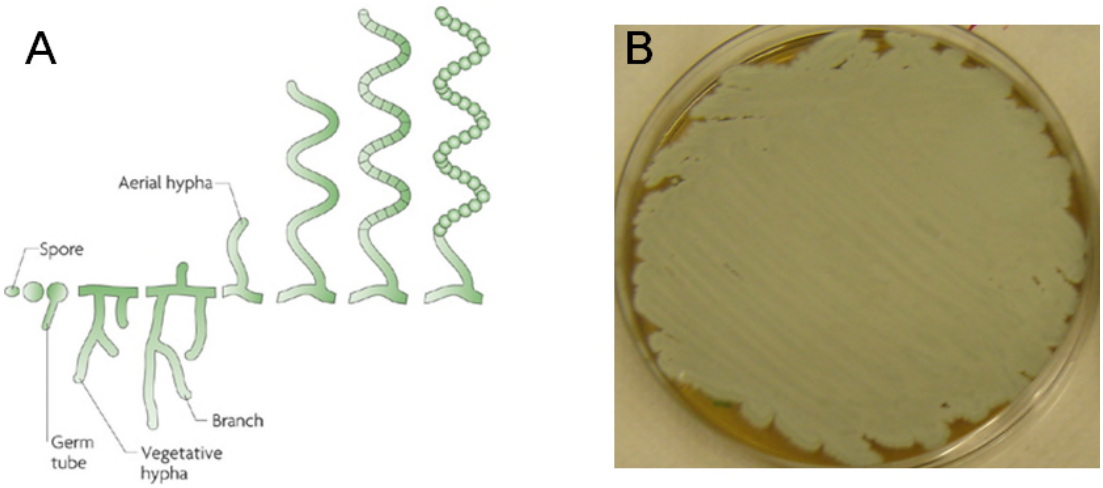


Figure 1.1 – A: *Streptomyces* life cycle schematic (Flärdh and Buttner)[4] and B: *Streptomyces venezuelae* ISP5230 grown on mannitol/yeast extract medium, MYM (image obtained during the present study)

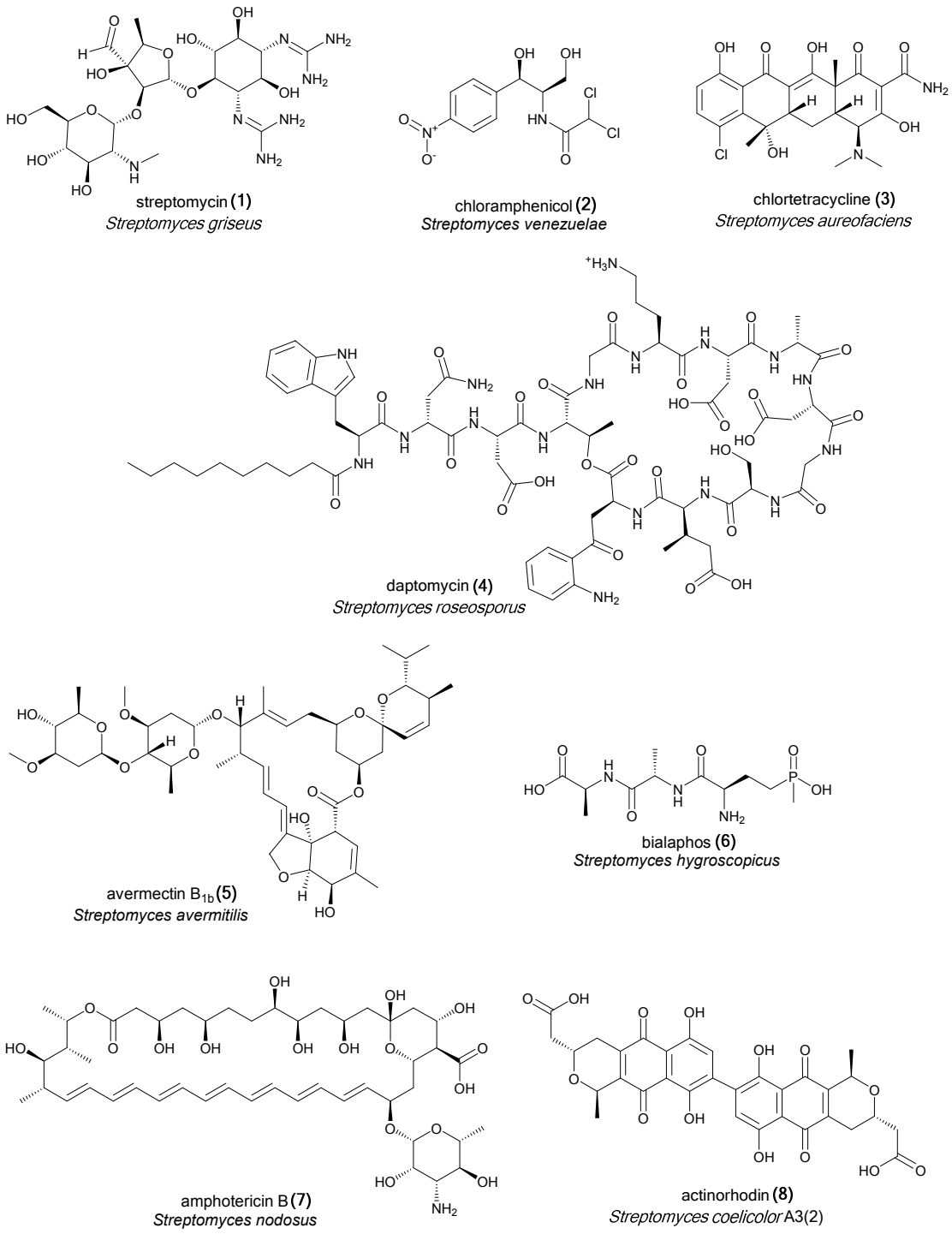


Figure 1.2 – *Streptomyces* metabolites of historical and industrial significance

Streptomyces species.[7,8] Since Alexander Fleming's discovery of the *Penicillium rubens* antibiotic natural product penicillin in 1928, the capability of microorganisms to inhibit the growth of other microbes has been rigorously explored.[9] The potential of *Streptomyces* to produce antibiotics was first realised as early as the 1940s, when

Streptomyces were first cultivated for this purpose. Discovered in the laboratory of Selman Waksman in 1943, the first natural product isolated from a *Streptomyces* species was the aminoglycoside antibiotic streptomycin, produced by *Streptomyces griseus*. Streptomycin quickly found industrial use as an agent to treat tuberculosis.[10-12] Since the isolation of streptomycin, many other antimicrobial natural products have been isolated from *Streptomyces* species, including tetracyclines, chloramphenicol and daptomycin (Figure 1.2).[13-15] It is estimated approximately two thirds of clinically useful antibiotics can be attributed to *Streptomyces*.[7,8,16]

Many of these natural products produced by *Streptomyces* are only produced by a small range of bacterial species and are typically not essential for growth. These so called 'secondary' metabolites have garnered *Streptomyces* bacteria a great deal of industrial importance. Examples include *S. avermitilis* (producer of the anthelmintic agent avermectin) and *S. hygroscopicus* (producer of the herbicide bialaphos and the immunosuppressant rapamycin), which have been in use for decades.[17-19]

Despite the success of strategies for natural product discovery that will be discussed over the following sections, there is still a great need for natural product discovery. Indeed, resistance to every single class of clinically used antibiotic has been reported within years of being introduced.[20,21] Methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant *Enterococci* and extremely drug resistant (XDR) strains of *Mycobacterium tuberculosis* are particularly well-known examples. It is estimated that there are as many as 360,000 cases of XDR tuberculosis every year and approximately 53% of clinical strains of *Escherichia coli* (*E. coli*) are resistant to the first-line treatment ampicillin.[22,23]

This rise in antibiotic resistance has been contrasted by a dramatic shift in the investment of big pharmaceutical companies away from antimicrobial research. Indeed, during the 1990s, many companies scaled back development of new antibiotic compounds, citing rediscovery of previously known natural products, flouting of intellectual property laws and the immense cost of clinical trials compared to the return on investment.[24,25] Furthermore, there has been increased focus on more profitable drugs to treat ailments such as cancer, depression and cardiovascular diseases. These types of drugs are generally prescribed for periods of months or years (compared with the relatively short-term timescales of most antibiotic regimens), therefore are more profitable.

One example of antibiotic screening programs giving rise to low numbers of drug leads are the *in vitro* target-based assays employed by GlaxoSmithKline in the late 1990s/early 2000s, which involved genetic manipulation and screening of culture extracts of over 300 mutated strains. This generated only five new leads after seven years of research[20,26]. In the last few years however, antimicrobial development has begun to return to development of narrow-spectrum antimicrobial agents and *in vivo* assays, with the ultimate aim of keeping our most effective, broad-spectrum antimicrobials reserved for only the most life-threatening cases.[26,27]

Natural products are an invaluable source of compounds in the whole field of medicine, and medicinal chemists are always in search of new compounds which are more efficacious and exhibit fewer side effects.[16,28,29] These characteristics are required for development of all classes of medicines and are not limited to antibiotics. Indeed, *Streptomyces* bacteria have made significant contributions to our arsenal of available medicines, as discussed in the next section.[2,3,28,29]

1.2 Strategies for natural product discovery

1.2.1 Phenotype-guided discovery

Many *Streptomyces* secondary metabolites have been discovered as a result of phenotypic-directed assays, such as the aforementioned antimicrobials streptomycin, chloramphenicol and daptomycin. Indeed, the biosynthetic potential of Streptomycetes has been exploited in screening programs for discovery of other medicinal natural products. Examples include anticancer agents such as the anthracycline doxorubicin isolated from *Streptomyces peucetius* and the glycopeptide bleomycin isolated from *Streptomyces verticillus*.[30,31] Immunomodulatory agents such as rapamycin, isolated from *S. hygroscopicus* and antifungal compounds such as the polyene amphotericin B, isolated from *S. nodosus*, provide further examples of *Streptomyces* natural products discovered by phenotypic screens.[32-34] Example of the diversity of such structures are shown in Figures 1.2 and 1.3.

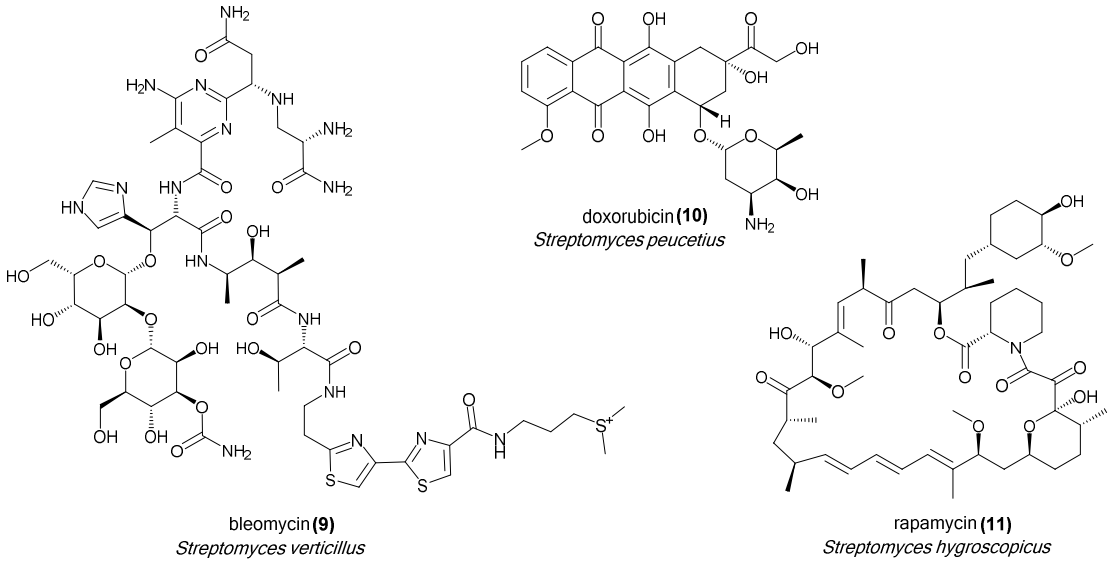


Figure 1.3 – Anticancer (doxorubicin and bleomycin) and immunosuppressant (rapamycin) drugs derived from *Streptomyces* discovered by screening programs

The natural product(s) responsible for the desired phenotype can then be isolated by bioassay-guided fractionation in which the compounds produced by the organism(s) of interest are separated by chromatography.[35,36] Fractions which maintain activity of the culture extract are separated further to purify and identify the compound of interest. Analytical chemistry techniques such as liquid chromatography-mass spectrometry (LC-MS) offer an efficient way of comparing the metabolites produced by different bacterial strains.[35,36]

This provides information on the different chemical phenotype – or 'chemotype' of bacterial strains to identify natural products before establishing their physiological function.[35-38] If the natural products identified appear to be novel, then culture conditions can be scaled up for isolation and more rigorous characterisation of the natural product of interest, followed by activity-based assays. LC-MS analysis is an extremely powerful technique as it allows the metabolites of different strains to be observed either by UV detection or by mass spectrometry detection.

1.2.2 Pleiotropic regulation of secondary metabolism

Traditional approaches of growing strains on media containing different carbon and nitrogen sources may simply alter the metabolic flux in primary metabolism, therefore affect the intracellular availability of different precursors which may be siphoned off for incorporation into secondary metabolites. Indeed, *S. venezuelae* ISP5230 is known to produce chloramphenicol when grown at 28 °C in a glucose-isoleucine medium and to produce jadomycin B when grown at 37 °C in media that contains galactose as the primary carbon source (Figure 1.4).[13,39] Furthermore, the reduced yield of actinorhodin produced by *S. coelicolor* A3(2) when grown in high phosphate concentrations is also well characterised.[40,41]

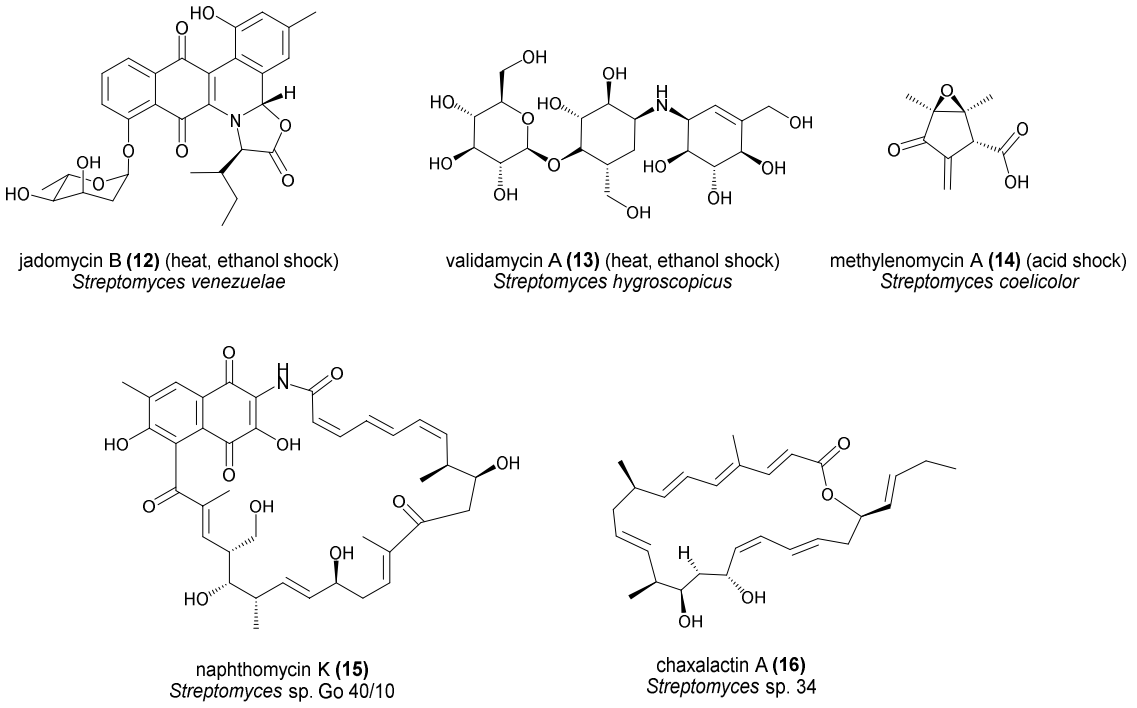


Figure 1.4 – Some examples of *Streptomyces* metabolites produced in response to altered culture conditions. For example, jadomycin B and validamycin A are each produced by their host organisms under conditions of heat and ethanol stress. Methylenomycin is produced in acidic pH. Naphthomycin K and chaxalactin were discovered by application of the OSMAC approach

The idea that a single bacterial strain is able to produce a host of different secondary metabolites when grown in different conditions has been exploited for decades and has relatively recently been referred to as 'one strain, many compounds' – OSMAC.[42-44] This principle has been applied to *Streptomyces*, leading to the observation that over 50% of metabolites extracted from *Streptomyces* sp. Gö 40/10 grown in different conditions proved to be novel compounds. These new metabolites have structures varying from butyrolactones, to glycosylated macrolides and naphthalene ansamycins.[45] More recently, a similar approach has been applied to the extremophile *Streptomyces* sp. C34, resulting in the discovery of a series of new polyene macrolactones, chaxalactins A-C.[46]

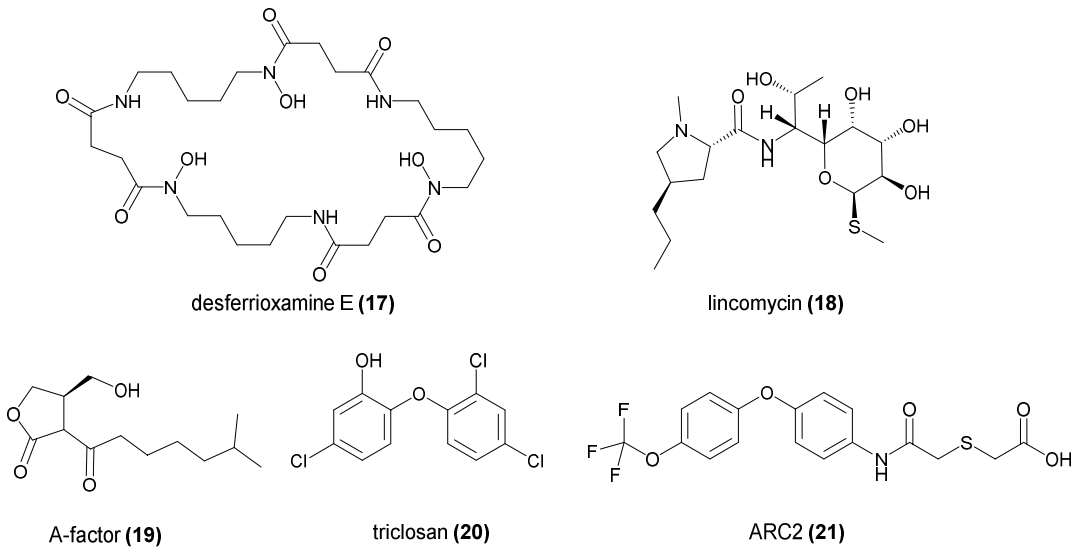


Figure 1.5 – Examples of known elicitors of secondary metabolism in *Streptomyces*. The siderophore desferrioxamine has been implicated in interspecies signalling.[52] The role of A-factor in antibiotic biosynthesis has been extensively studied and will be discussed later. Lincomycin, triclosan and ARC2 are proposed to affect secondary metabolism by exerting stress upon the target bacteria[49]

Stress responses to environmental factors such as temperature and pH often result in changes in secondary metabolism. Examples include production of the antifungal pesticide validamycin A by *S. hygroscopicus*, which is dramatically increased by heat/ethanol shock and methylenomycin production by *S. coelicolor* when grown in acidic conditions.[44,47,48] However, the mechanisms involved in responses to heat shock and pH are poorly understood.[34,47]

The ability of *Streptomyces* to produce different metabolites in response to stress can also be exploited by addition of antimicrobial compounds at sub-lethal concentrations.[49] Nodwell and co-workers screened a library of over 100 compounds for enhanced secondary metabolite production profiles in *S. coelicolor*, leading to the observation that Triclosan (which is a known inhibitor of FabI in fatty acid biosynthesis, Figure 1.5) could change the metabolic profile of *S. coelicolor* when added to cultures at concentrations lower than their minimum inhibitory concentrations (MICs). This led to the identification of antibiotic remodelling compounds (ARCs)

which consistently increased actinorhodin production in *S. coelicolor*. One such compound, ARC2, has been shown to alter the metabolite profiles of sixty strains of *Streptomyces* when added to the culture medium.[49,50] Imai and co-workers have also very recently reported that addition of sub-lethal concentrations of lincomycin to *S. coelicolor* results in overproduction of actinorhodin. Furthermore, addition of lincomycin to *Streptomyces lividans* at approximately one third of its MIC led to overproduction of novel derivatives of the calcium dependent antibiotics.[51]

Co-culturing of *Streptomyces* species has also been used to alter the metabolic profiles of some strains. For example, Ueda and co-workers have recently shown when *Streptomyces griseus* and *Streptomyces tanashiensis* are co-cultured, antibiotic production in *S. tanashiensis* is increased. This was attributed to secretion of desferrioxamine E by *S. griseus*.[52] This method of co-culturing had previously been used to establish a precedent for interspecies signalling being involved in antibiotic biosynthesis and it was suggested that this is a widespread phenomenon.[53]

Furthermore, Marmann and co-workers have recently reported from a study of co-cultivating several soil and marine fungi with *Streptomyces* species, that the soil bacterium *Tsukamurella pulmonis* can induce production of a red pigment in *S. lividans* and can also induce production of the novel antibiotic alchivemycin A in *Streptomyces endus*.[54] Interestingly, there were also several compounds reported in this study that were produced by fungi belonging to the genera *Aspergillus* and *Fusarium* only when co-cultivated with *S. lividans* or *S. peucetius*, but not by exenic cultures. Therefore, *Streptomyces* can both alter the secondary metabolism of other organisms, as well as having their own secondary metabolism altered when co-cultured.

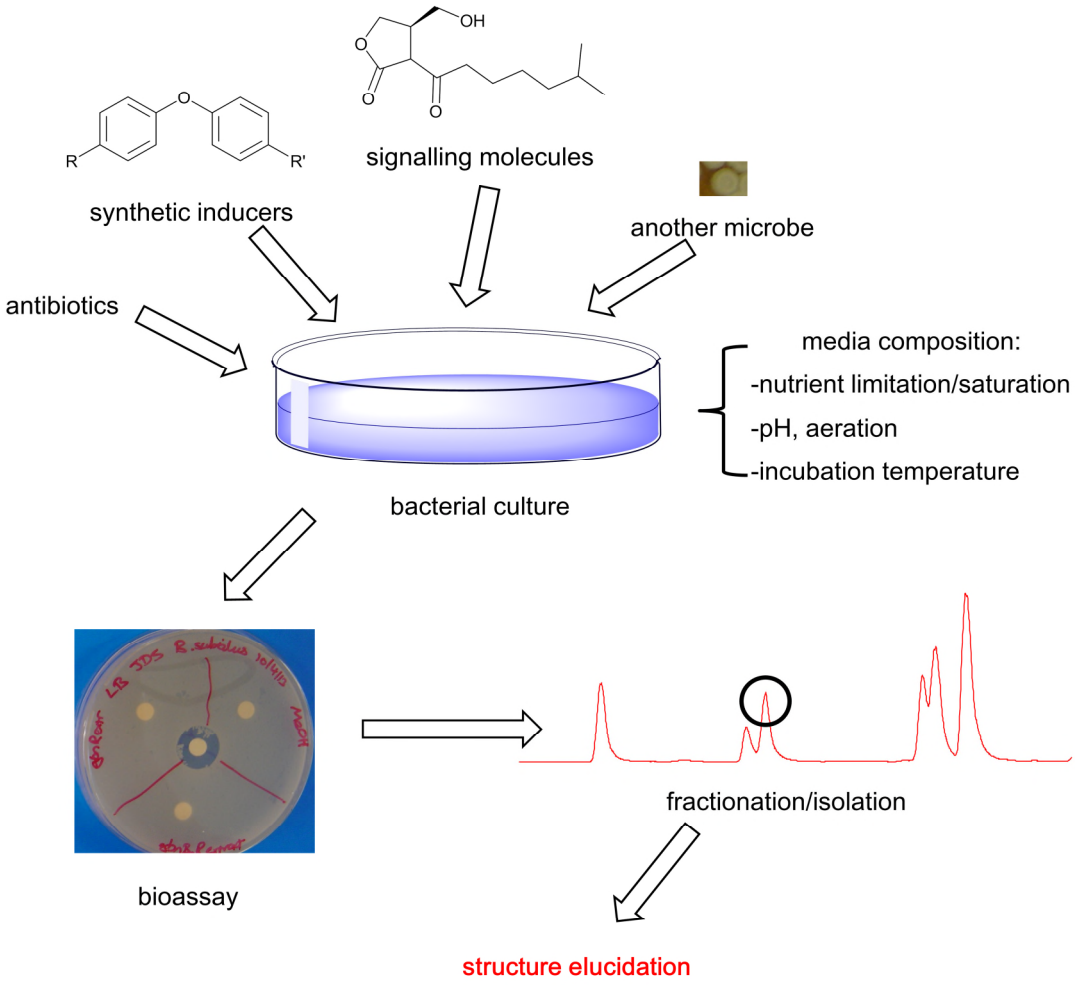


Figure 1.6 – Strategies for phenotype-based natural product discovery

Co-cultivation is therefore a valuable method for natural product discovery. However, in many cases it may not be clear whether the induction of natural product biosynthesis in heterologous strains is due to a stress response (due to change in pH of culture media or nutrient starvation) or whether the response is caused by secretion of a specific elicitor molecule by a particular strain.[52,53] For example, some *Streptomyces* species are known produce signalling molecules such as autoregulatory factor (A-factor, Figure 1.5)[55], which will be discussed in Section 1.4. Therefore, the molecular mechanisms involved in interspecies induction of natural product biosynthesis can be complex. Co-cultivation has also resulted in the observation that many previously characterised natural products were also overproduced by some of the test strains.[52-54]

In summary, the methods for natural product discovery discussed thus far all involve 'black box' approaches. Whilst some are phenotypically-guided in order to search for novel compounds with antibiotic activity, for example, other strategies exploit bacterial responses to particular extracellular conditions such as nutrient availability, stress or chemical signals (Figure 1.6). More rational strategies are therefore required to unlock the biosynthetic potential of *Streptomyces* (and other bacteria) in order to discover novel natural products. Secondary metabolite biosynthesis is tightly regulated and poor yields of natural products are often observed regardless of the nature of the culture medium.[16,34,56,57]

One caveat of these phenotype-based methods discussed is the large amount of chemical redundancy inherent in this approach; the same classes of compounds (or the exact same compounds) are rediscovered time and again. This has in part been attributed to cultivation conditions in the laboratory being far removed from the environment in which bacteria exist in nature – *Streptomyces* for example inhabit soil and marine ecosystems, with complex biotic and abiotic factors that are difficult to reproduce in the laboratory.[28,29,56,58]

In order to address this problem – and therefore gain access to previously uncultivatable bacteria, Ling and co-workers have recently reported how emerging technologies such as the iChip can be utilised to culture new strains of bacteria. This is achieved by inoculating a chip with a suspension of cells and growing them in their native environment, separated by a semi-permeable membrane.[58] Their report of the new antimicrobial NRPS teixobactin earlier this year received a significant amount of media attention.[58] Whilst the iChip represents a significant advance in the biochemical arsenal available to researchers (it is estimated that only 1 % of microorganisms can be

grown in a laboratory environment)[58,59] this still represents a 'black box' approach to natural product discovery.

1.3 Genome mining approaches to natural product discovery

1.3.1 Sequencing projects have revealed the biosynthetic potential of *Streptomyces*

One of the primary reasons for *Streptomyces* – and indeed other microorganisms – being unable to synthesise the full range of secondary metabolites they have the genetic capability to produce in a laboratory environment, is that secondary metabolite pathways are very tightly regulated.[56,57] In an evolutionary context, this represents a way in which bacteria can preserve energy by only producing proteins that they need in response to specific extracellular stimuli.[16,56,59]

The genes encoding for biosynthetic enzymes involved in the synthesis of a specific natural product are often clustered together in bacteria. Regulatory elements and genes encoding for self-resistance mechanisms are also often located within the same cluster as the biosynthetic (structural) genes.[60,61] Upon sequencing of the *S. coelicolor* A3(2) genome in 2002, it was revealed that the *S. coelicolor* chromosome is 8.7 Mbp in size and contains 25 gene clusters involved in secondary metabolite biosynthesis. As *S. coelicolor* had been extensively studied for decades, this large number of secondary

Table 1.1 – Examples of biosynthetic potential of some sequenced *Streptomyces*

Organism	*Size / Mbp*	*Number of genes*	*Number of secondary metabolite clusters*	*Ref*
Streptomyces coelicolor A3(2)	8.7	7825	25	61
Streptomyces avermitilis	9.0	7583	34	62
Streptomyces scabies 87-22	10.1	8530	28	20
Streptomyces venezuelae	8.2	7238	31	63
Streptomyces griseus IFO 13350	8.5	7138	34	64

metabolite gene clusters was very surprising.[3,61] Since then, dozens of other *Streptomyces* genomes have also been sequenced (Table 1.1 includes examples from some industrially important microbes). It would appear that 20-30 secondary metabolite gene clusters per Streptomycete is the norm, demonstrating that the *Streptomyces* genus is still a largely untapped reservoir of natural products.[61-64]

Sequencing projects have identified many gene clusters in *Streptomyces* species that might not be expressed in laboratory conditions.[20,61-64] Indeed, many of these silent pathways are also 'cryptic'; they encode enzymes for biosynthesis of unknown natural products. Exploiting the vast quantity of new genomic data has reinvigorated natural product discovery by development of a new paradigm – 'genome mining'.[59,65-69]

Bioinformatics tools such as BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) can be used to find homologous gene and protein sequences in vast depositories of genomic information. Newer tools such as multi-gene BLAST and antiSMASH 3.0 (http://antismash.secondarymetabolites.org/) can now be used to identify whole biosynthetic gene clusters for a natural product of interest.[70,71] Whilst in most cases nucleotide or protein sequences alone cannot be used as a predictive tool to reveal the structure of the natural product encoded for by a whole gene cluster, these bioinformatics tools are invaluable to infer the function of individual biosynthetic enzymes and regulatory proteins.

In order to determine the nature of cryptic natural products, it is important to first establish that the cluster is expressed. This may be achieved by reverse transcription-polymerase chain reaction (RT-PCR), DNA microarrays or RNA-Seq analyses.[67,68] Assigning functions to a specific biosynthetic gene cluster based on bioinformatics

analyses alone is still extremely challenging in most cases. Therefore, genetic manipulation is required in order to link a specific pathway to a particular natural product.[65,66] Some genome mining methods for overproduction of natural products are detailed in the next section.

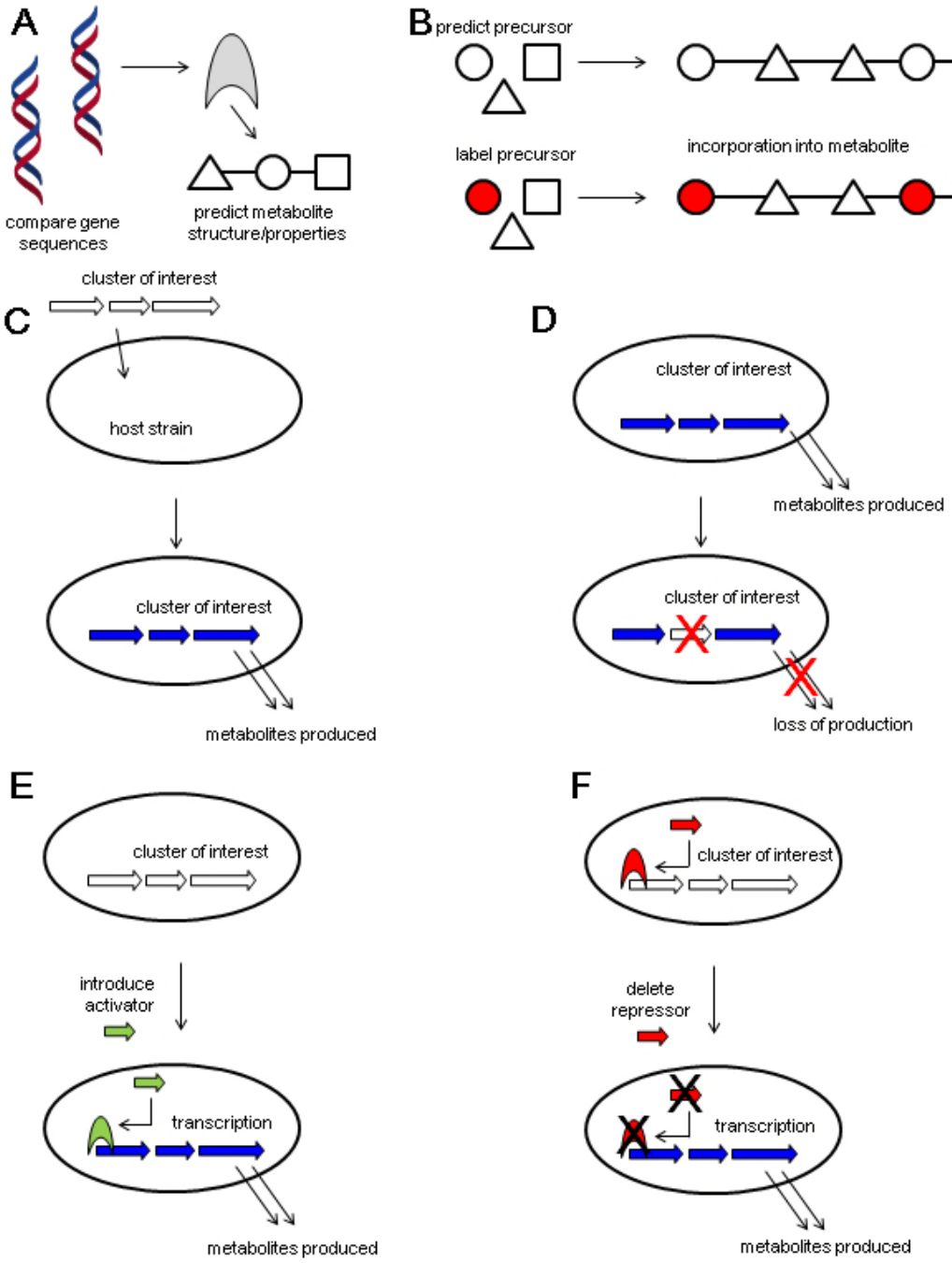


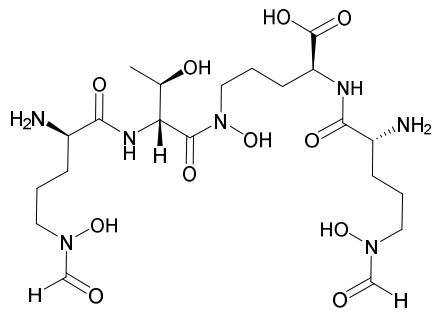
Figure 1.7 – Genome mining strategies for natural product discovery. Bioinformatics can be used to predict natural product structures assembled by gene clusters, or by specific enzymes (A and B). Genes may be expressed in a heterologous host (C) or knocked out (D) in order to deduce their roles in natural product biosynthesis. Alternatively, if activators/repressors can be identified, they may also be overexpressed/deleted to increase expression of the biosynthetic cluster of interest (E and F)

1.3.2 Bioinformatics can be used to predict natural product scaffolds

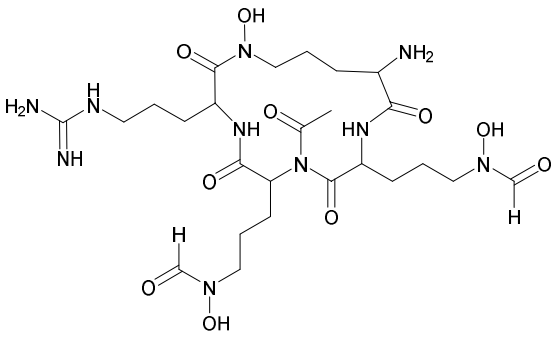
Highly conserved multienzyme complexes such as polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) have distinctive amino acid sequences. Indeed, the sequences of acetyl transferase (AT) domains within PKSs and adenylation (A) domains in NPRSs are conserved to the extent that substrate specificity can often be predicted from the amino acid/nucleotide sequence alone. Indeed, there are now tools such as NRPSpredictor (nrps.informatik.uni-tuebingen.de/) available to help with predicting substrate specificity of A domains.[72,73] Predictions of the substrate specificity of the A domains of the NRPS involved in biosynthesis siderophore coelichelin **22** allowed the structure to be predicted with a high level of accuracy; this information was used to guide the purification of this cryptic metabolite.[74] A similar approach (panel A in Figure 1.7) was used recently to identify a new siderophore natural product **23** produced by *S. peucetius* (Figure 1.8).[75]

Furthermore, sequencing data can also provide information on what tailoring reactions may occur during the synthesis of natural products synthesised by these enzyme complexes. Similarly, reactions involved in the synthesis of ribosomally synthesised, post-translationally-modified peptides (RiPPs) can be predicted. One recent example of a RiPP structure predicted using bioinformatic analysis is grisemycin **24** (Figure 1.8), produced by the *grn* gene cluster in *S. griseus* IFO 13350.[76]

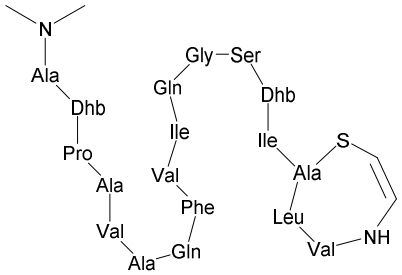
The *grn* cluster was compared to the cypemycin gene cluster in *Streptomyces* sp. OH-4156 (each gene in the *S. griseus* cluster was >60% identical to its homologue in the OH-4156 cluster).[76] However, such bioinformatic comparisons were only possible because a very similar biosynthetic system had been studied previously. If the gene cluster of interest has little or no homology to characterised metabolic pathways, then



coelichelin **(22)**
Streptomyces coelicolor



unnamed siderophore **(23)**
Streptomyces peucetius



grisemycin **(24)**
Streptomyces griseus

Figure 1.8 – Examples of *Streptomyces* metabolites, the structures of which were predicted by bioinformatics comparisons of their biosynthetic gene clusters with previously characterised ones[74,75]

other approaches may be necessary to link the specific genes involved to a natural product structure.

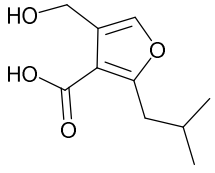
Bioinformatics tools such as the conserved domain search on NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) may also be used to determine the specific function of a particular enzyme, therefore allowing the substrate to be predicted.[77] When bioinformatic analyses allow partial structural predictions to be made, stable-isotope precursor molecules can be used to guide the purification of the compound of interest. This 'genomisotopic' methodology (panel B in Figure 1.7) has been applied by Gross and co-workers to discover orfamides A-C produced by *Pseudomonas fluorescens* Pf-5.[78]

1.3.3 Heterologous expression

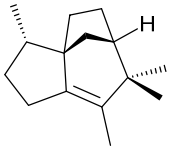
In order to identify and solve the structure of a cryptic natural product, it may be advantageous to express the gene cluster of interest in a heterologous host. It may be particularly difficult to detect the metabolic product of a particular pathway if there are many other metabolites produced by the host strain or if the gene cluster is poorly expressed.

Upon introduction of the gene cluster of interest into a heterologous host, metabolites produced by the wild type and mutant strains may then be compared to discover new natural products and link their biosynthesis to the gene cluster introduced into the host organism (panel C in Figure 1.7). The 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids (AHFCA) compounds such as **25** from *S. coelicolor* A3(2) provide an example of novel natural products discovered by this approach and will be discussed in more detail later.[79] A novel sesquiterpene, epi-isozizaene **26**, was shown to be the product of the cryptic *Streptomyces coelicolor* gene *sco_5222* when it was heterologously expressed in *E. coli*.[80] However, epi-isozizaene is not produced by wild-type *S. coelicolor*, which converts it into albaflavenol and albaflavenones.

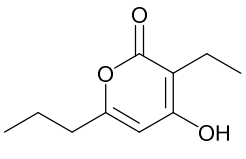
Indeed, there are now several 'superhost' strains of *S. coelicolor* available, such as M1146 and M1152. These hosts were constructed by the removal of large gene clusters



AHFCA1 **(25)**



epi-isozizaene **(26)**



germicidin C **(27)**

Figure 1.9 – Examples of *S. coelicolor* metabolites discovered by heterologous expression of genes (AHFCAs and epi-isozizaene) and by gene knockouts (germicidin B)[79,83]

encoding for known metabolites such as actinorhodin (*act*), streptorubin B (*red*), calcium-dependant antibiotic (*cda*) and coelimycin (*cpk*) clusters.[7,81] These gene deletions were made in order to increase metabolic flux by removing the sink for many secondary metabolite precursor molecules from primary metabolism and to provide a cleaner background in LC-MS analyses of natural products. *Streptomyces* M1152 also carries a mutation in the RNA polymerase B gene *rpoB* that favours biosynthesis of secondary metabolites.[81]

Heterologous expression of the aforementioned *S. griseus grm* gene cluster in the *Streptomyces coelicolor* host strain M1146 allowed the new peptide metabolite product grisemycin (Figure 1.8) to be characterised by comparison of the metabolite profiles of the parent and daughter strains.[7,76] However, if the biosynthetic gene cluster of interest is very large (for example, polyketide biosynthetic genes can exceed 100 kb in size), then heterologous expression is more difficult. One complementary approach is to target specific structural or regulatory genes in a particular cluster to identify a novel natural product.[66-68]

1.3.4 Gene knockouts

In order to assign functionality to a specific biosynthetic gene, the gene can be knocked out by replacement with a resistance marker (panel D in Figure 1.7).[82] For example, three new germicidins (as well as the two originally isolated from *Streptomyces viridochromogenes*) were identified as being present in the metabolite profile of *S. coelicolor* M145 but absent in a *sco_7221* deletion strain (Figure 1.9).[83]

To provide further evidence for a specific function of a given gene, the deletion mutants can be complemented either genetically (by re-introducing the deleted gene

into the genome) or by supplementing culture media with a putative or known biosynthetic intermediate. In the case of germicidins, the *sco_7221* gene was re-introduced into the *sco_7221* deletion mutants, restoring germicidin production in *S coelicolor*. The *sco_7221* gene was also heterologously expressed in *S. venezuelae* ISP5230, thus leading to production of germicidins by *S. venezuelae*.[83]

1.3.5 Genetic inactivation of repressors/overexpression of activators

As mentioned above, secondary metabolite gene clusters often include regulatory genes controlling the production of that metabolite.[56,57] For example, genes *actII-orf4*, *redD* and *mmyB* all encode for transcriptional activators are found within in the gene clusters for biosynthesis of actinorhodin, streptorubin B and methylenomycin, respectively, in *S. coelicolor*.[84,85] Therefore, specific regulatory elements in the biosynthetic pathway of interest may be targeted in order to turn on potentially silent pathways and to determine the nature of the corresponding cryptic natural product. This may be achieved by overexpression of transcriptional activators, or by deletion of transcriptional repressors (Figure 1.7 – panels E and F).[84,86]

The first example of this approach being used to discover novel metabolites in *Streptomyces* was overexpression of the *samR0484* gene encoding for a transcriptional activator in *Streptomyces ambofaciens*. Overexpression of *samR0484* was achieved by construction of a vector in which *samR0484* was placed under the control of the strong *ermE** promoter and subsequent transformation of *S. ambofaciens* using this vector. This led to the discovery of a series of new 51-membered macrolactones – the stambomycins (for example **28** in Figure 1.10).[86]

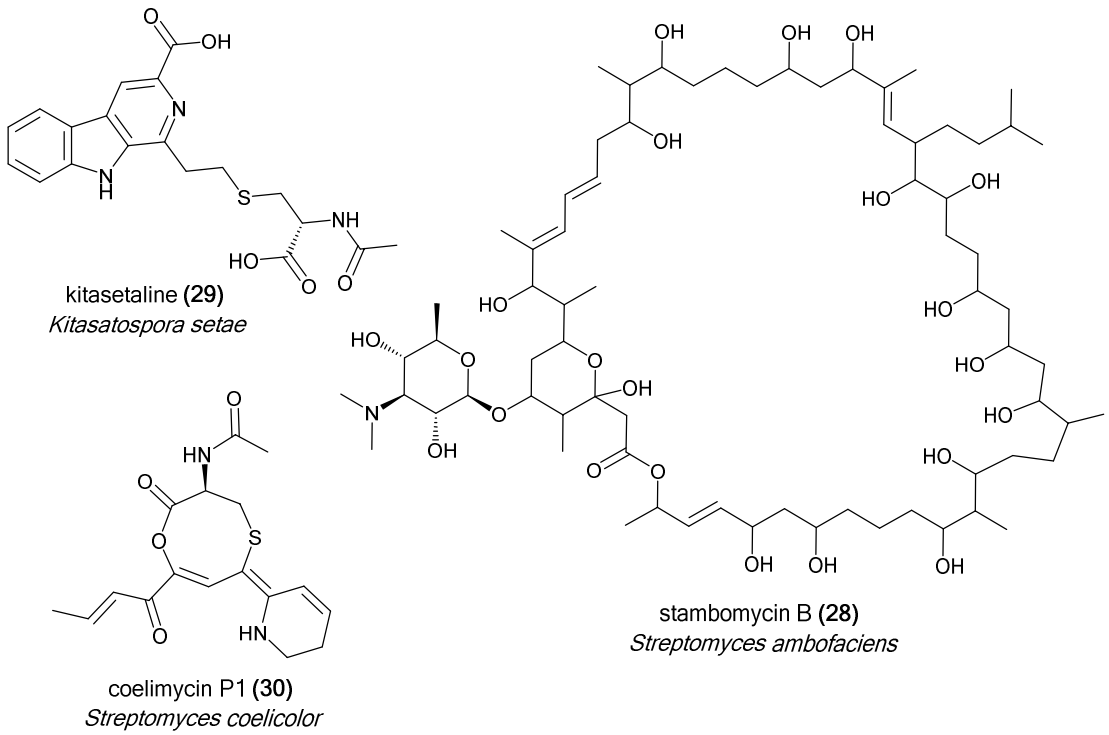


Figure 1.10 – Examples of Actinomycete metabolites discovered by overexpression of a gene encoding for an activator (stambomycin B **28**) and by deletion of a genes encoding for a transcriptional repressors (kitasetaline **29** and coelimycin P1 **30**)[87-89]

Furthermore, derepression of a biosynthetic pathway may be achieved by deletion of a gene encoding for a transcriptional repressor. For example, deletion of the *mmyR* gene leads to overproduction of methylenomycin in *S. coelicolor*.[85] MmyR is proposed to be a transcriptional repressor protein; when the *mmyR* gene was deleted, there was constitutive production of methylenomycin. However, there are very few examples of this approach being used to specifically 'wake up' silent pathways.[84]

One example from Aroonsri and co-workers demonstrates that deletion of the *ksbC* transcriptional regulator in the Actinomyctete *Kitasatospora setae* led to overproduction of the novel alkaloid ketasetaline **29** (Figure 1.10). The *ksbC* gene was identified as a pseudo γ-butyrolactone (GBL) receptor and subsequently deleted, leading to the discovery of kitasetaline. Even though it was initially predicted that KsbC would act as a repressor, it was found that KsbC also acts as an activator in the

biosynthesis of bafilomycin.[87] A further example is the yellow *S. coelicolor* pigment coelimycin **30**, which is overproduced by a strain in which the *scbR2* transcriptional repressor gene has been deleted (Figure 1.10).[88,89]

These examples provide precedent for overexpression of activators or deletion of transcriptional repressors as powerful strategies for natural product discovery, if the activators/repressors in a silent biosynthetic gene cluster can be identified and their potential functions predicted. Transcriptional activators and repressors may be pleiotropic (for example AdpA in *S. griseus*) or pathway specific (for example, MmyR in *S. coelicolor*) and may involve indirect regulatory effects.[85,90,91] In the case of pleiotropic regulators such as *adpA* or *ksbC*, the gene clusters regulated by a specific activator/repressor may be situated hundreds of kb away in the chromosome. Therefore, it is necessary to understand the molecular mechanisms by which particular transcriptional activators/repressors function in order to determine exactly which genes to overexpress/delete within a specific cluster to elicit overproduction of the cryptic metabolite.

1.4 TetR transcriptional regulators

1.4.1 Functions of TetR family regulators

There are many distinct classes of transcriptional regulator proteins in prokaryotes. Most are one-component systems (in which sensory and DNA-binding domains exist in a single polypeptide), for example, LysR, ArsR and TetR family transcriptional regulators.[92,93] Over twenty classes of one-component systems have been discovered in prokaryotes, as defined by the primary amino acid sequence of their DNA-binding domains. Members of each class typically act as either activators (such as the LuxR

family regulators), or as repressors (for example, the LacI and TetR families) or indeed act as both.[92] For example, the LysR family of transcriptional regulators can either promote or repress transcription.[92,93] In many cases, several classes of regulatory protein may exist in one biosynthetic gene cluster; hence deciphering regulatory mechanisms directing biosynthesis of a given natural product can be a very difficult task.[94,95]

For example, if a transcriptional regulator represses transcription of another repressor, which in turn represses expression of structural genes, then deletion of the first repressor will result in expression of the structural genes. In Jadomycin biosynthesis for example, JadR3 has been shown to repress transcription of the regulatory gene *jadR2*, which is also proposed to act as a repressor of jadomycin biosynthesis.[95] Hence, understanding of the regulatory mechanisms involved with specific pathways is necessary to deduce exactly which regulatory genes to delete in a given system.

Named after the tetracycline resistance protein TetR first discovered in *E. coli*, this family of transcriptional regulators mainly act as repressors and generally function as homodimers, with each subunit consisting of a ligand-binding domain and an N-terminal DNA-binding domain.[92,96] TetR family proteins are widespread in prokaryotes. For example, analysis of the *S. coelicolor* genome revealed it to contain genes encoding for 965 regulatory proteins, many of which are located in or adjacent to biosynthetic gene clusters.[61] Over 150 of these regulatory proteins belong to the TetR family.[92] As well as a clear role in controlling metabolite export, therefore contributing significantly to antibiotic self-resistance mechanisms, TetRs have been implicated in a number of other functions including regulation of amino acid, lipid and nitrogen

metabolism. Other functions of TetRs include roles as effector proteins in quorum sensing and regulators of secondary metabolism.[92]

TetR binds to the region in which its promoter and the promoter of the divergently adjacent *tetA* gene overlap (Figure 1.11). Upon binding tetracycline, TetR undergoes a conformational change, reducing the affinity of the DNA binding domain for its target sequence in the *tetA-tetR* intergenic region.[14,97] Therefore TetR dissociates from the DNA, allowing RNA polymerase to bind to the *tetA-tetR* promoter and therefore transcribe *tetA*. The translated TetA protein acts as export protein for tetracycline, thus conferring self-resistance to tetracycline when it reaches a high enough intracellular concentration to induce *tetA* expression (Figure 1.11).[98] This mechanism is important for self-resistance, as tetracycline will trigger transcription of *tetA* at sub-inhibitory concentrations, therefore allowing tetracycline to start being exported before it damages the host cell.[92]

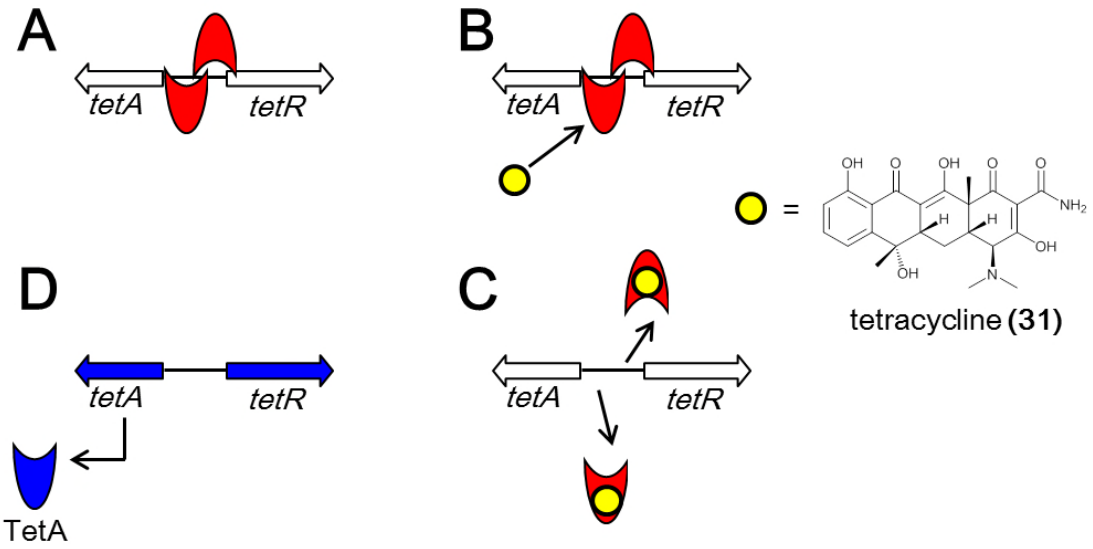


Figure 1.11 – Mode of action of TetR. TetR binds to the *tetA-tetR* promoter, repressing transcription of *tetA* (A). When tetracycline is present inside the cell, it binds to the ligand binding domain of TetR (B). This leads to dissociation of TetR from its DNA target (C) allowing *tetA* to be transcribed (D); TetA then exports tetracycline out of the cell[97,98]

1.4.2 Ligands for TetRs

Of the TetRs for which the structures have been solved, there are approximately 200 *apo* structures, 21 with ligands bound and 7 with DNA bound (as reported in 2013 by Cuthbertson and Nodwell).[92,99] The crystal structure of the *S. coelicolor* MmfR protein with its ligand AHFCA2 **32** was solved in our group in 2013 and will be discussed in Section 1.6.[100] Indeed, the known ligands for TetRs vary in structure from large amphipathic molecules such as stearoyl-CoA **33**, to spirotetronates such as kijanimicin, porphyrins such as heme **34** and small signalling molecules such as AHFCA2 (Figure 1.12).[92,99,101]

Of particular interest to secondary metabolite pathways are those TetRs which bind to diffusible, low molecular weight signalling molecules – 'bacterial hormones' – such as AHFCAs or γ-butyrolactones (GBLs). Repressors which recognise these types of molecules all belong to the TetR subfamily of ArpA-like transcriptional repressors and are often involved in regulation of antibiotic biosynthesis.[102-105]

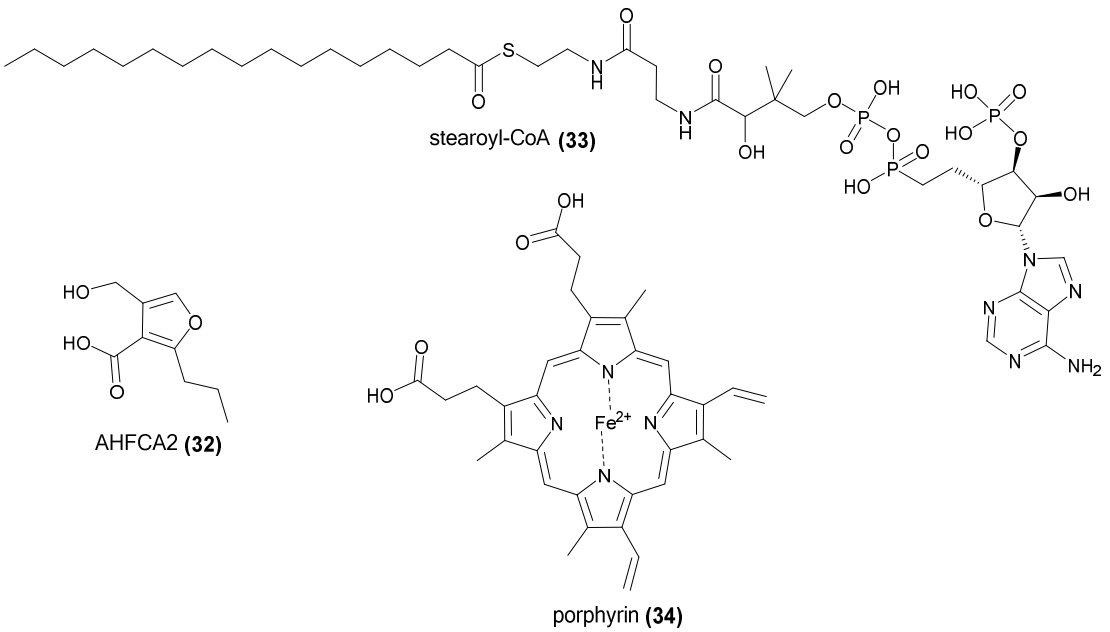


Figure 1.12 – Examples of structural diversity of ligands known to interact with TetR proteins[92,100]

1.4.3 ArpA family transcriptional repressors

The archetypal ArpA protein from *S. griseus* has been studied in detail.[90,102,106] ArpA acts as a negative regulator of expression of the *adpA* gene, binding to a specific DNA sequence – an autoregulatory response element (ARE) – situated in the *adpA* operator region. ArpA is responsive to autoregulatory factor (A-factor).[91,92] When the intracellular concentration of A-factor reaches a critical point, it binds to ArpA, causing a conformational change, therefore reducing the affinity of ArpA for the ARE sequence in the *adpA* operator. Therefore, RNA polymerase is able to bind the *adpA* promoter,

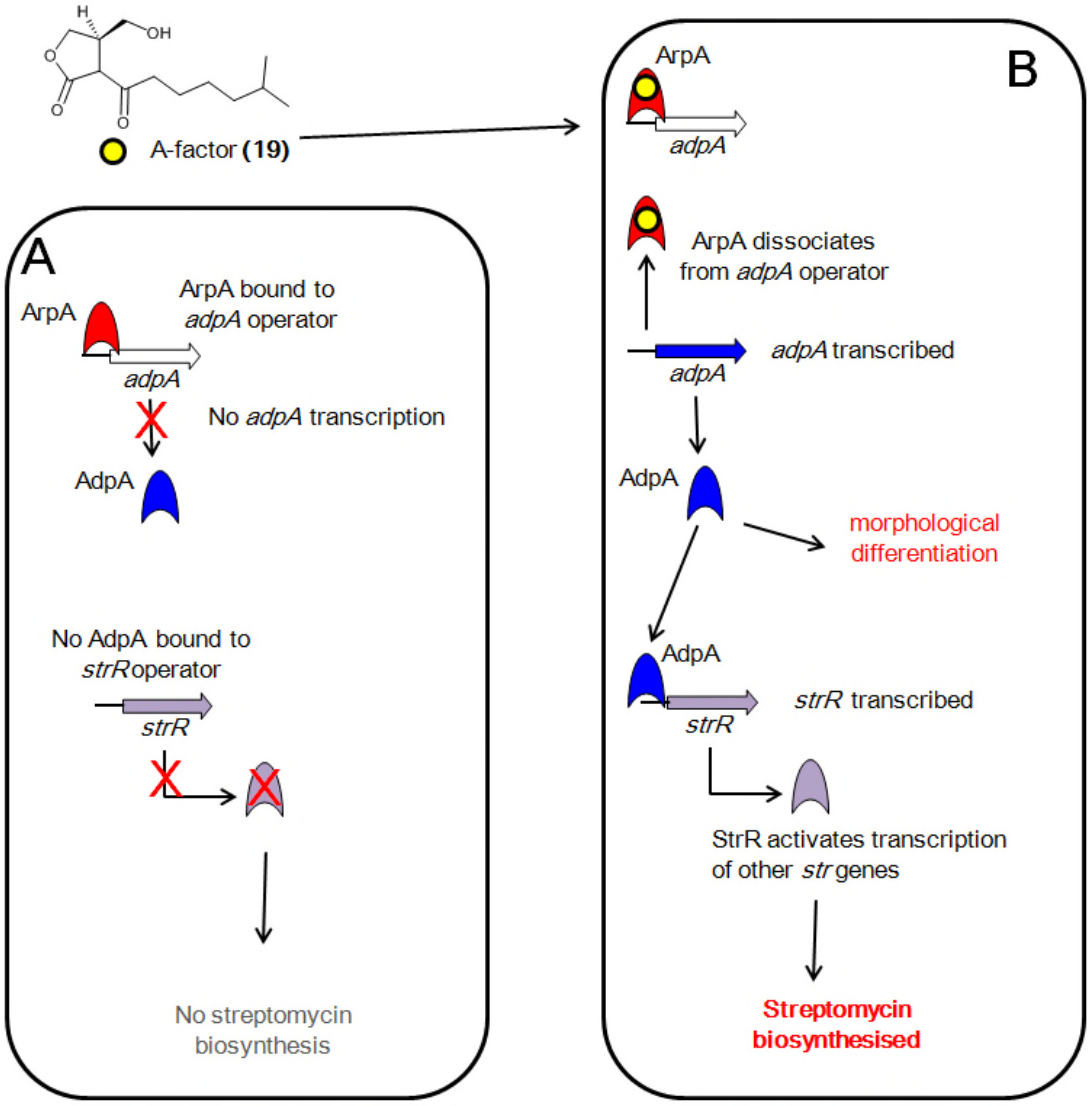


Figure 1.13 – Mode of action of ArpA and the regulatory cascade activating streptomycin biosynthesis upon A-factor binding to ArpA.[91] ArpA prevents transcription of the activator *adpA* (pathway A), until A-factor is detected, allowing for transcription of *adpA* and subsequent activation of streptomycin biosynthesis (pathway B)

allowing transcription and translation to occur.[91,106]

AdpA is a pleiotropic transcriptional activator implicated in the regulation of secondary metabolite production and morphological differentiation. It is believed to regulate transcription of as many as 80 different genes.[57,106] One of the targets of AdpA is the operator region of *strR*, a gene which encodes for the transcriptional activator StrR, which activates streptomycin biosynthesis (Figure 1.13).[90,91] AdpA also indirectly regulates expression of *griR*, a pathway-specific activator in biosynthesis of the yellow pigment grixazone.[103,107]

Interestingly, ArpA is distinct from other ArpA-like transcriptional regulators in the context of both its gene location and its function; it does not appear adjacent to a

Table 1.2 – Examples of ArpA-like proteins and their cognate ligands and antibiotic biosynthetic pathways they regulate

Strain	*Ligand biosynthetic enzyme*	*Ligand*	*Ligand acceptor(s)*	*Natural product biosynthesis regulated*	*Refs*
6-keto type GBLs:					
S. griseus	AfsA	A-factor	ArpA	streptomycin	55,102
6-hydroxy-type GBLs:					
S. coelicolor	ScbA	SCB 1-3	ScbR	act, red, coelimycin	88,110
S. venezuelae	JadW1	SCB3	JadR3	jadomycin B	95
S. lavendulae	FarX	IM-2	FarA	showdomycin	113,114
S. virginiae	BarX	VB A-E	BarA	virginiamycin	115,116
butenolides:					
S. avermitilis	Aco	avenolide	AvaR1	avermectin	117
S. rochei	SrrX	SRB1/2	SrrA	lankamycin	118
furans:					
S. coelicolor	MmfL	AHFCA 1-5	MmfR	methylenomycin	79,85
ligands unknown:					
S. ambofaciens	unknown	unknown	AlpZ	alpomycin	119
S.aureofaciens	SagA	unknown	SagR	auricin	120
S. acidiscabies	SabA	unknown	SabR	angucyclinone	121
S. fradiae	unknown	unknown	TylP	tylosin	122
S. pristinaespiralus	unknown	A-factor-like	SpbR	pristinamycim	94
K. setae	KsbS4	unknown	KsbA	bafilomycin	87

biosynthetic gene cluster which it regulates, and it appears to be an indirect pleiotropic regulator as it negatively affects transcription of the pleiotropic activator *adpA*. Whilst some other ArpA-like proteins are pathway-specific regulators of secondary metabolite biosynthesis, ArpA is also involved in morphological differentiation.[90,105] Some examples of ArpA-like transcriptional repressors and the antibiotic biosynthetic gene clusters they regulate are shown in Table 1.2. As mentioned in the previous section, targeted gene deletions of transcriptional regulators such as ArpA family regulators represent a valuable tool to access new natural product scaffolds from cryptic, silent gene clusters.

1.4.4 Ligands for ArpA-like repressors

Discovered in 1967, autoregulatory factor (A-factor **19**) was the first example of a γ-butyrolactone (GBL) to be isolated from *Streptomyces* species.[55] AfsA has been identified as the key biosynthetic enzyme required for A-factor production and is encoded for by the *afsA* gene.[108,109] Biosynthesis of A-factor is discussed in Section 1.5.

Concurrent research over the last 40 years has led to the isolation and characterisation of other γ-butyrolactones and GBL signalling systems in a range of *Streptomyces* species; factor I from *S. viridochromogenes*, three Gräfe factors isolated from *S. bikiniensis*, virginiae butanolides VB-A to E isolated from *S. virginiae*, factor IM-2 from *S. lavendulae* FRI-5 and *Streptomyces coelicolor* butyrolactones (SCBs) 1-3 from *S. coelicolor*.[112,113,115,123,124] Structures of known γ-butyrolactone and other inducer molecules proposed to interact with ArpA-like proteins are shown in Figure 1.14.

The functions of these GBLs and other *Streptomyces* signalling pathways have been the subject of several recent reviews[16,104,105] and the systems in which the functions of the

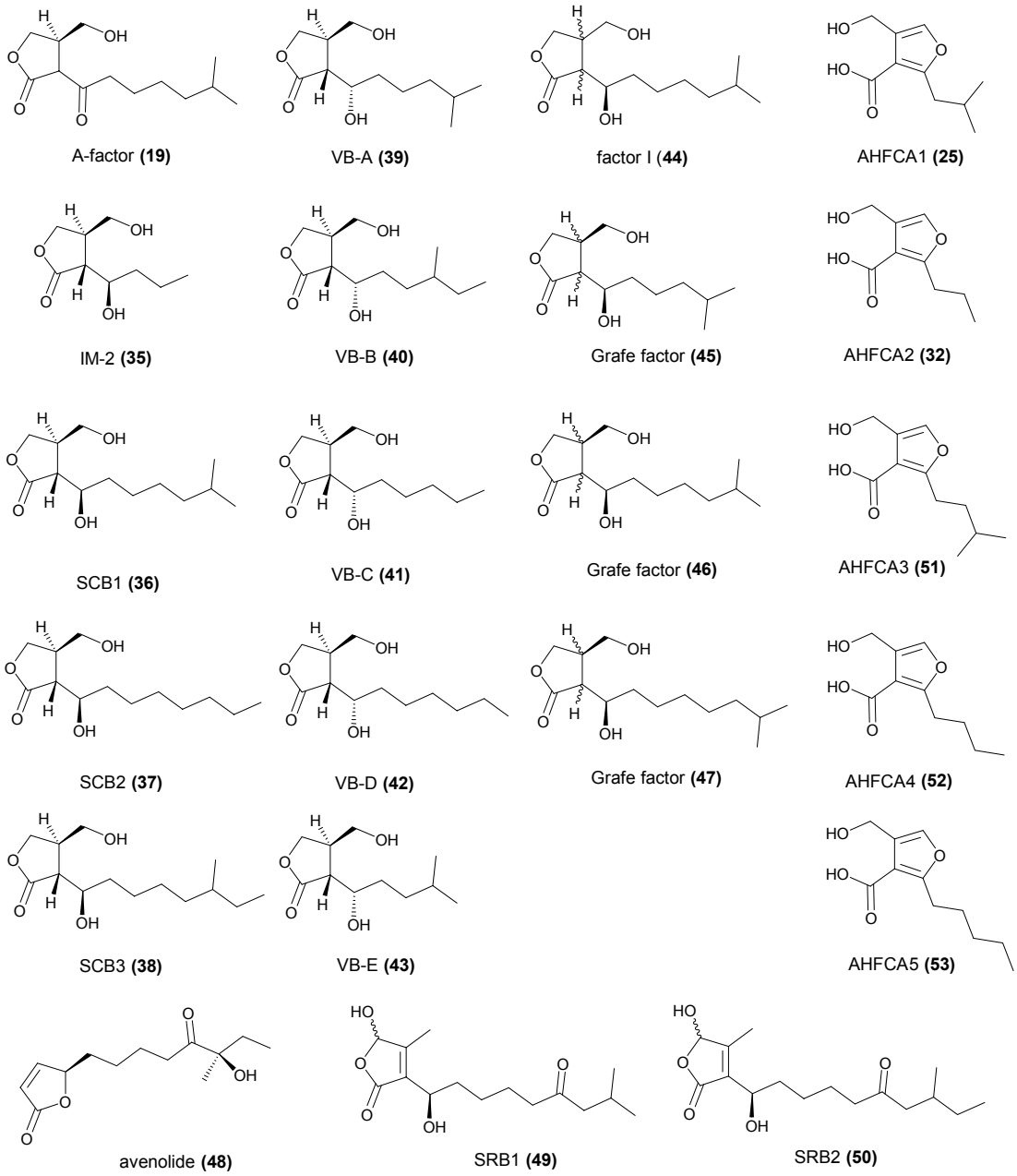


Figure 1.14 – Structures of known *Streptomyces* signalling molecules.[79,112-124] The absolute stereochemistry of Gräfe factors and factor I have not been reported

AfsA and ArpA homologues have been investigated are shown in Table 1.2. GBLs can be classified as '6-keto-type' – in which there is a keto group at carbon 6 (A factor is still the only example of this type of GBL) and '6-hydroxy-type' – in which the keto group has been converted to a hydroxyl group (Figure 1.15). The natural 6-hydroxy-type GBLs can be sub-divided into other classes based on the stereochemistry at carbons 2, 3 and 6 (Figure 1.15).

It has been shown that the absolute stereochemistry of carbons 2, 3 and 6 confers the specificity of each GBL receptor for its cognate ligand(s).[112,125,126] A-factor is the only natural example of the 6-keto type of GBL reported and readily undergoes epimerisation due to the relative acidity of the proton at carbon 2. The butyrolactones from the most studied systems, SCBs and VBs systems, all share 2*R*, 3*R* stereochemistry. In addition, the length of the alkyl chain also affects the ligand binding affinity of ligand receptor protein. For example IM-2 and SCBs 1-3 have the same (2*R*, 3*R*, 6*R*) configuration and vary only in the length of their alkyl chains. Their cognate receptor proteins, FarA for IM-2 and ScbR for SCBs, have different specificities – *in vitro* assays have shown FarA is most responsive to IM-2 type GBLs with alkyl chain lengths containing four or five carbons, and ScbR has been show *in vivo* to be most responsive to IM-2 type GBLs with eight or nine carbons in the alkyl chain.[112,125]

Other GBLs for example the aforementioned factor I and Gräfe factors have not been fully characterised. There are also examples of other γ-butyrolactones which have

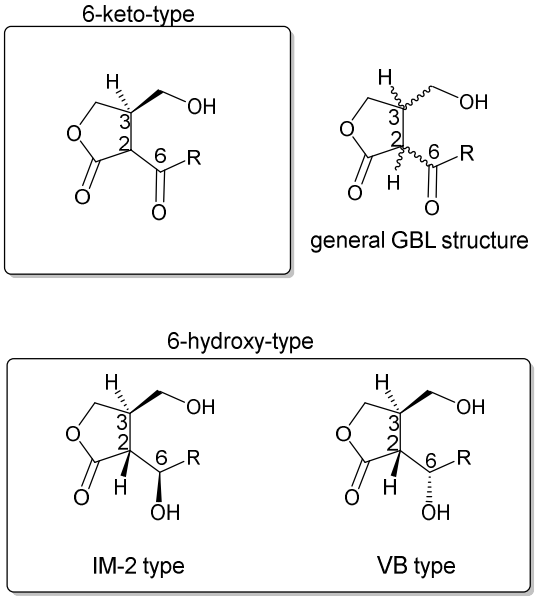


Figure 1.15 – Stereochemistry of γ-butyrolactones varies at carbons 2, 3 and 6. The stereochemistry is 2*R*, 3*R* for all the examples discussed in the text (A-factor, SCBs and VBs)

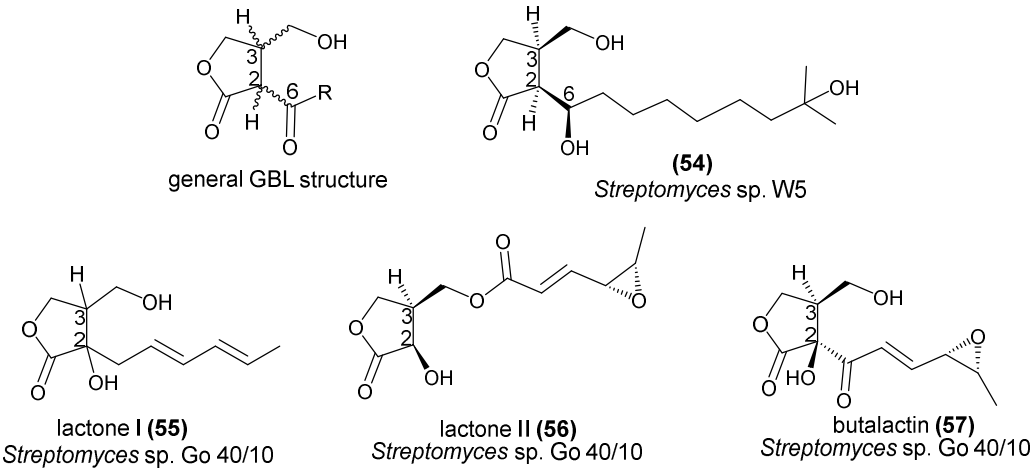


Figure 1.16 – Examples of other butyrolactones isolated from *Streptomyces* species[45,126] Stereochemistry at position 3 is *R* in all cases where the structure have been elucidated fully, and others have derivatised alkyl chain substituents

different stereochemistry (**54** has 2*S*, 3*R*, 6*R* stereochemistry) and derivatised hydroxymethyl or alkyl groups, such as the butyrolactones **54**-**57** isolated from *Streptomyces* sp. W5 and *Streptomyces* sp. Gö 40/10 (Figure 1.16).[45,126] The biosynthetic origin and functions of these GBLs have not been investigated, however.

Comparative genomic analysis has more recently led to the identification of many AfsA-like biosynthetic enzymes. ArpA-like putative GBL receptors and their proposed roles in the regulation of antibiotic biosynthesis and morphological differentiation have also been reviewed.[104,105,128,129] Some AfsA homologues and the ligands that they biosynthesise are shown in Table 1.2.

Furthermore, the ArpA-like repressor homologues are referred to in the literature as 'GBL receptors' and 'pseudo-GBL receptors', as defined by phylogenetic analyses.[92,128] This nomenclature is misleading, as the ligands for ArpA-like repressors are more structurally diverse than just different GBL structures (Figure 1.14 and Table 1.2).

Indeed, investigation of the role of the *aco* gene in regulation of avermectin biosynthesis in *S. avermitilis* has resulted in the discovery of avenolide (Figure 1.14) as a signalling molecule regulating that system. The butenolides SRB1 and SRB2 have also recently been reported to regulate lankamycin biosynthesis in *S. rochei*.[117,118] Investigation of another *afsA*-like signalling pathway involved in regulation of methylenomycin biosynthesis has led to the identification and structural characterisation the AHFCAs – as exemplified by the methylenomycin furans (MMFs), the structures of which are shown in Figure 1.14.[79,85]

If the nature of the signalling molecules interacting with specific ArpA-like repressors could be predicted, then adding that specific signalling molecule to the host strain might induce expression of biosynthetic pathways for cryptic natural products. This principle has been tested using an indirect kanamycin-activity reporter assay in the case of the SCBs, in which the ability of SCB analogues were used to de-repress ScbR protein from the *cpkO* promoter upstream of a kanamycin resistance gene.[112] AHFCAs have also been used to induce methylenomycin production in strains unable to produce them (as the *mmfLHP* genes had been deleted).[79,85]

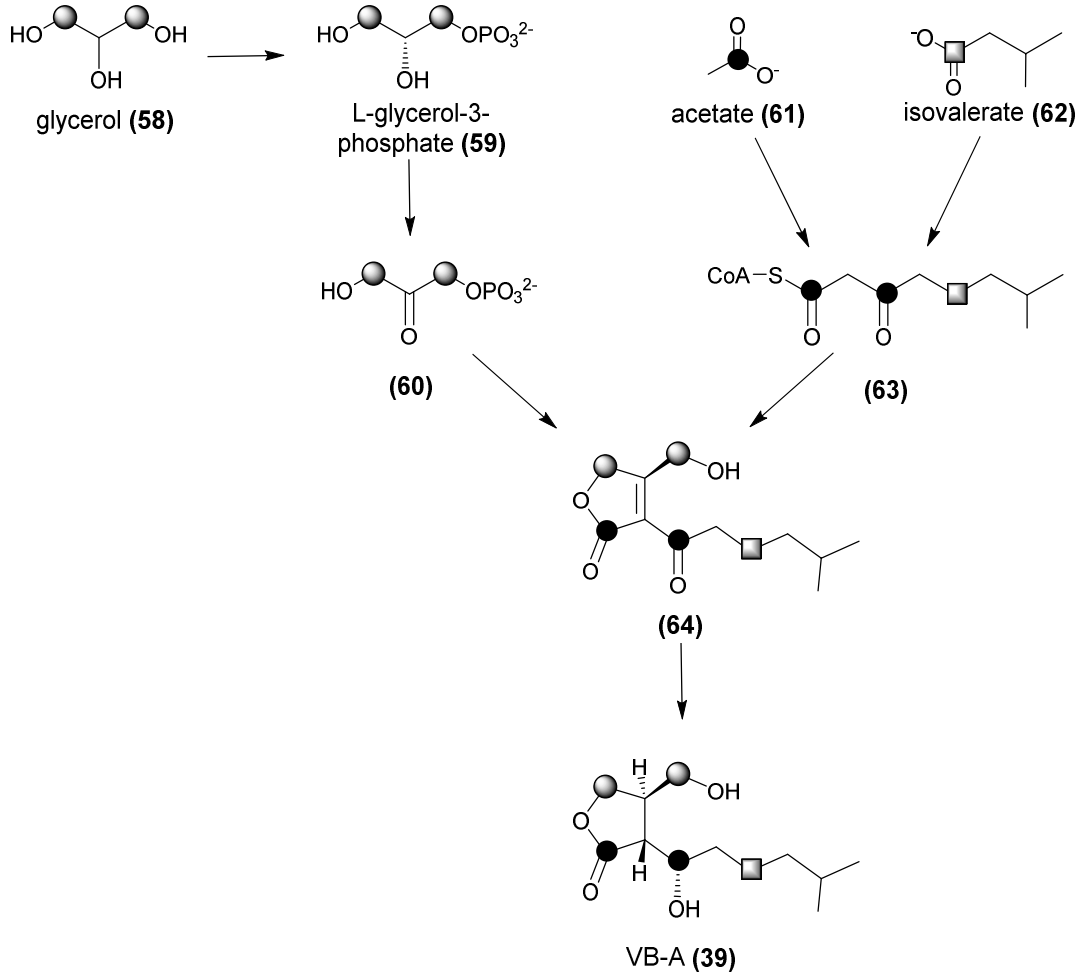
Many of the GBL or GBL-like signalling molecules proposed to be involved in regulation of antibiotic biosynthetic pathways remain unidentified due to poor production levels (see Table 1.2 for examples).[87,94,122] Indeed, the inductive effect of these signalling molecules is only observed at very low – and very narrow ranges – of concentrations (A-factor induces streptomycin biosynthesis in *S. griseus* at 0.5-250 nM concentrations and avenolide was also shown to induce avermectin production in an *aco* mutant of *S. avermitilis* at an extracellular concentration of 4-150 nM).[117,130] In order to understand the biochemical determinants linking a particular inducer molecule

to biosynthesis of a specific natural product, it is necessary to consider the substrate tolerance of the AfsA-like enzyme, the ligand binding properties of the ArpA-like receptor protein(s) and the DNA sequences to which the ArpA protein binds.

1.5 Biosynthesis of *Streptomyces* signalling molecules

1.5.1 The key step in GBL biosynthesis

Before the *afsA*-like enzyme BarX was identified, a biosynthetic route to virginiae butenolide A (VB-A) **39** was first proposed by Sakuda and co-workers and was verified by incorporation experiments using stable isotope precursors (**58**, **61** and **62** in Scheme

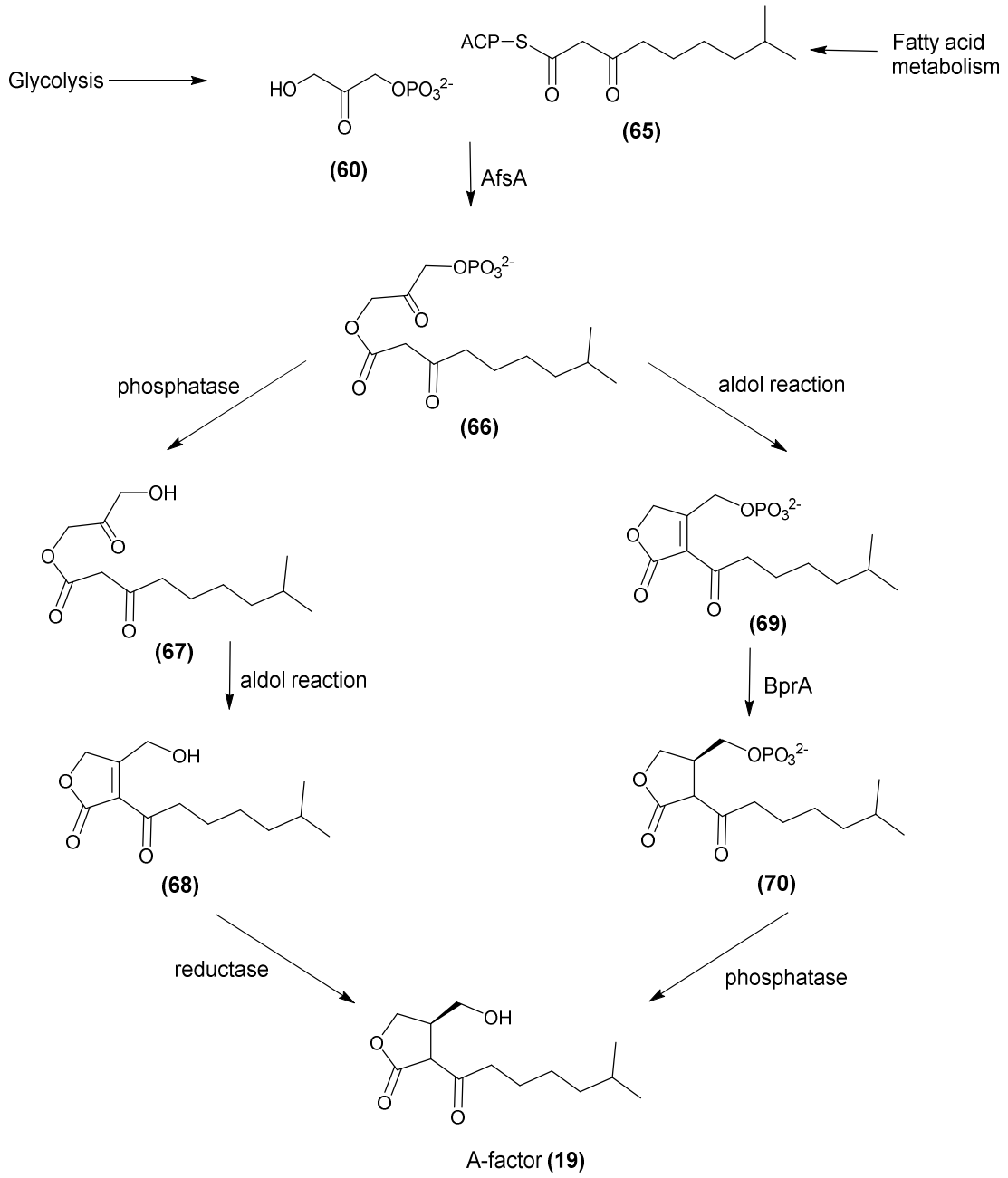


Scheme 1.1 – Specific incorporation of ^{13}C labels from acetate (filled circles), isovalerate (grey square) and glycerol (grey circles) precursor molecules into virginiae butenolide-A (VB-A)[131]

1.1).[131] Over the last decade, investigations into the biosynthesis of several GBLs and GBL-like molecules have revealed more details of the specific functions of AfsA and AfsA-like enzymes. In particular, AfsA has been shown to be the key enzyme required for A-factor biosynthesis by Kato and co-workers, ScbA has been shown to be required for SCB biosynthesis and MmfL was described as the key enzyme for MMF biosynthesis by Corre and co-workers.[79,109,132] The only published *in vitro* work on AfsA-type enzyme thus far is for AfsA itself.

Following from the initial observation that A-factor production is restored by insertion of the *afsA* gene into a *S. griseus* mutant lacking a DNA fragment that includes *afsA*, it has been shown that AfsA is required for A-factor biosynthesis.[133] More recent *in silico* modelling studies suggested AfsA contains a tunnel capable of accepting an acyl-ACP acyl chain, implying AfsA may be involved in catalysing the condensation of a β-ketoacyl chain and 3-carbon containing compound to produce an A-factor precursor.[109] Kato and co-workers confirmed *in vitro* that AfsA is the key enzyme for A-factor biosynthesis, as it catalyses the condensation of an intermediate from fatty acid metabolism – 8-methyl-3-oxononanoyl-acyl carrier protein (ACP) **65** – and dihydroxyacetone phosphate (DHAP) **60** to form an 8-methyl-3-oxnonanoyl-DHAP ester intermediate **66**, shown in Scheme 1.2.

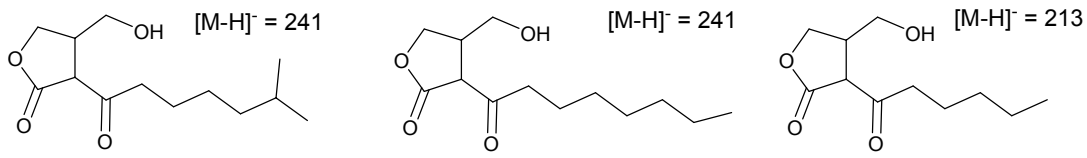
Regarding the downstream steps from the phosphorylated ester intermediate **66**, three further reactions are required for A-factor biosynthesis – an intramolecular aldol reaction that results in the butenolide, a dephosphorylation step and a step involving reduction of the C2-C3 double bond in the butenolide intermediate. Two initial pathways for these three steps were proposed by Sakuda and co-workers in which: i) dephosphorylation of the ester intermediate is followed by the aldol reaction and final



Scheme 1.2 – Biosynthetic route to A-factor as proposed by Kato and co-workers[109.134]

reduction of the butenolide to A-factor and ii) intramolecular aldol reaction occurs before the reduction step, which is followed by the dephosphorylation (Scheme 1.2).[134] Indeed, Kato and co-workers presented evidence for both of these pathways occurring during their investigations of AfsA and A-factor biosynthesis *in vitro* (Scheme 1.2).[109]

An enzymatic assay utilising His-tagged AfsA, ^{32}P-labelled DHAP and the synthetic 8-methyl-3-oxononanoyl-acyl ACP analogue, 8-methyl-3-oxononanoyl-*N*-



(19) **(71)** **(72)**

A-factor from *S. griseus*

A-factor homologues from *E. coli* expressing *afsA*

Figure 1.17 – A-factor and A-factor analogues produced by *E. coli* when expressing *afsA* as proposed by Kato and co-workers[109]

acetylcysteamine (NAC) led to the detection of a radioactive ester intermediate on thin layer chromatography (TLC). This product was compared with an authentic synthetic standard of the ester intermediate **66**. After dephosphorylation of the ester, analysis by LC-MS confirmed the presence of the butenolide expected to be formed by a non-enzymatically catalysed intramolecular aldol reaction of the dephosphorylated ester intermediate **67** to yield **68** (Scheme 1.2).

This synthetic ester **66** was also added to a lysate of a *S. griseus ΔafsA* mutant, and resulted in production of A-factor thus confirming that AfsA is required for A-factor biosynthesis and that this ester intermediate could be converted into A-factor by other enzymes present in the *ΔafsA* mutant lysate. Additionally, when testing other 3-carbon substrates in the AfsA assay, only upon addition of DHAP did the initial concentration of 8-methyl-3-oxononanoyl-*N*-acetylcysteamine (NAC) decrease with time, implying that the enzyme is specific for condensation of DHAP with the fatty acid derivatives to yield A-factor **19**.

This result was complemented by introduction of *afsA* into *E. coli*, which resulted in production of A-factor like molecules **71** and **72** (one of these molecules **71** exhibited the same *m/z* = 241 for its $[M-H]^-$ ion as does A-factor, the other **72** had *m/z* = 213 for its $[M-H]^-$ ion), shown in Figure 1.17. Both A-factor-like GBLs fragmented with the

same patterns as does A-factor when analysed by tandem mass spectrometry. The structural difference between A-factor and **71** and **72** was attributed to the differences in fatty acid metabolism between *Streptomyces* species and *E. coli*. *Streptomyces* species can produce branched fatty acids from starter units such as isobutyryl-CoA, 2-methylbutyryl-CoA and 3-methylbutyryl-CoA, whereas *E. coli* predominantly uses acetyl CoA.[135,136] Only *afsA* was expressed in the *E. coli* strain that produced these A-factor-like molecules. This observation implied that downstream steps in A-factor biosynthesis (the reduction of the butenolide and the dephosphorylation steps) are catalysed by enzymes that are non-specific to the A-factor biosynthetic pathway and such promiscuous phosphatases and reductases may be present across a wider range of bacteria than just *Streptomyces*.

Furthermore in the region close to *afsA*, an NADPH-dependent specific reductase – butenolide phosphate reductase, BprA – was identified, encoded by *bprA*. Recombinant BprA was shown to catalyse the reduction of the butenolide phosphate **69** to a butanolide phosphate **70** *in vitro*, but was also shown not to be able to catalyse the reduction of the dephosphorylated butenolide **68**, suggesting that there is a non-specific reductase in *E. coli* that can catalyse the reduction of the butenolide **68** to butanolide **69**. In both cases, it was suggested that there were other enzymes to catalyse the dephosphorylation step present in the *asfA* mutant lysate. This series of experiments therefore provided additional evidence for the biosynthetic pathway initially proposed by Sakuda and co-workers for VB biosynthesis.[131,134]

Biosynthesis of SCBs has been attributed to the ScbA protein in *S. coelicolor* (64% identity and 73% similarity to AfsA).[132,137] *In silico* analysis of ScbA and its homology to FabA and FabZ from fatty acid biosynthesis were used as a predictive tool for

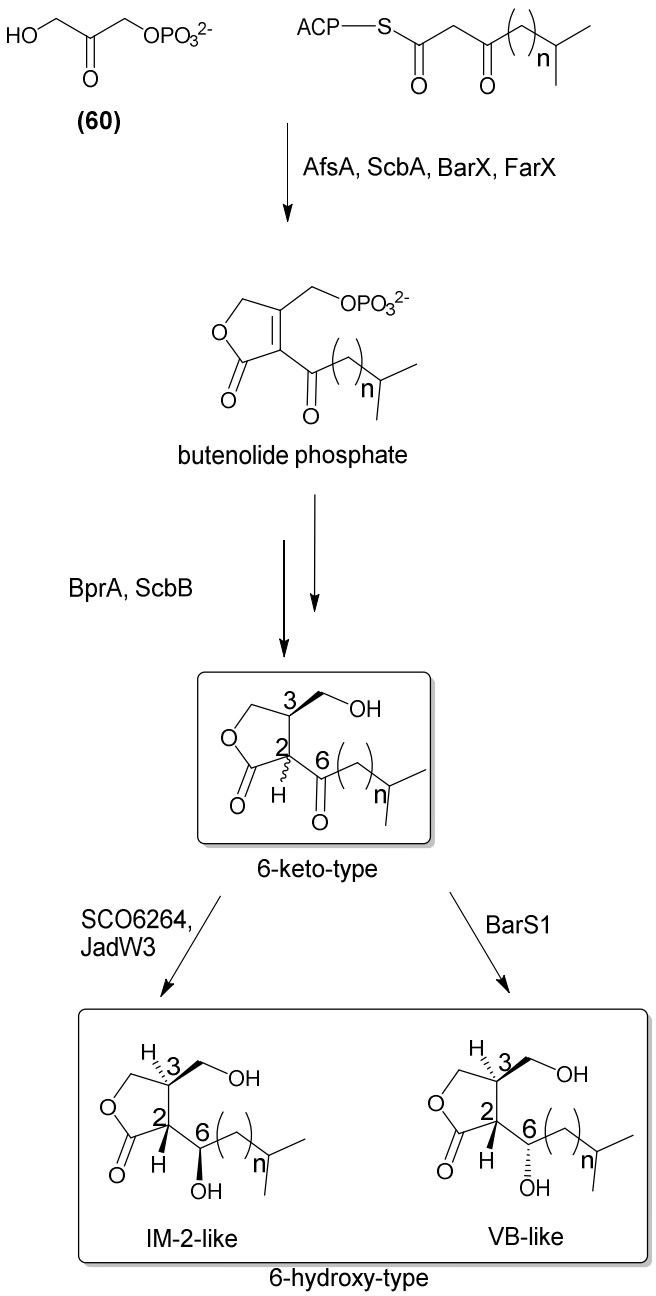
identification of key glutamate and arginine residues in the proposed active site.[132] Indeed, when these mutations were made, ScbA was no longer able to synthesise SCBs *in vivo*. These observations were accompanied by RT-PCR experiments to verify that the *scbA* mutant genes were being expressed. Attempts to reconstitute this pathway *in vitro* with recombinant ScbA were hindered by the poor solubility and expression levels of ScbA in heterologous hosts.[132]

1.5.2 Reduction of 6-keto-type GBLs to 6-hydroxy-type GBLs

All characterised 6-keto and 6-hydroxy-type GBLs have *R* stereochemistry at position 3 (Scheme 1.3). It was suggested by Kato and co-workers that BprA and its orthologue ScbB (SC06267), which has 76% identity and 84% similarity to BprA, convert the butenolide phosphate **69** into a butanolide phosphate **70** in a conserved stereospecific fashion, generating the 3*R* configuration in A–factor and SCB type GBLs, respectively.[109]

The reduction of the keto groups at position 6 and the resulting stereochemistry have been suggested to arise due to action of additional, NADPH-dependent reductases. One such enzyme has been characterised, BarS1, in *S. virginiae*.[138,139] The *barS1* gene is essential for VB biosynthesis and BarS1 has been shown to be specific for 6-keto-like butanolides with 3*R* stereochemistry, yielding only 6-hydroxy-type butyrolactones with 2*R*, 6*S* stereochemistry as in the natural metabolites VB-A-E (Scheme 1.3).

BarS1 orthologues are encoded for by *sco6264* in *S. coelicolor* (31% identity and 46% similarity at amino acid level) and *jadW3* in *S. venezuelae* (29% identity and 42% similarity at amino acid level).[95,139] *jadW3* has been shown to play a role in butyrolactone biosynthesis in the regulatory cascade controlling jadomycin production



Scheme 1.3 – General scheme of proposed steps in GBL biosynthesis, showing a conserved pathway to 6-keto type GBLs, where the stereochemistry of carbon 3 is determined by BprA (or homologous proteins), then the absolute configuration of carbons 2 and 6 are determined by either SCO6264/JadW3 or BarS1 enzymes[109,132,138]

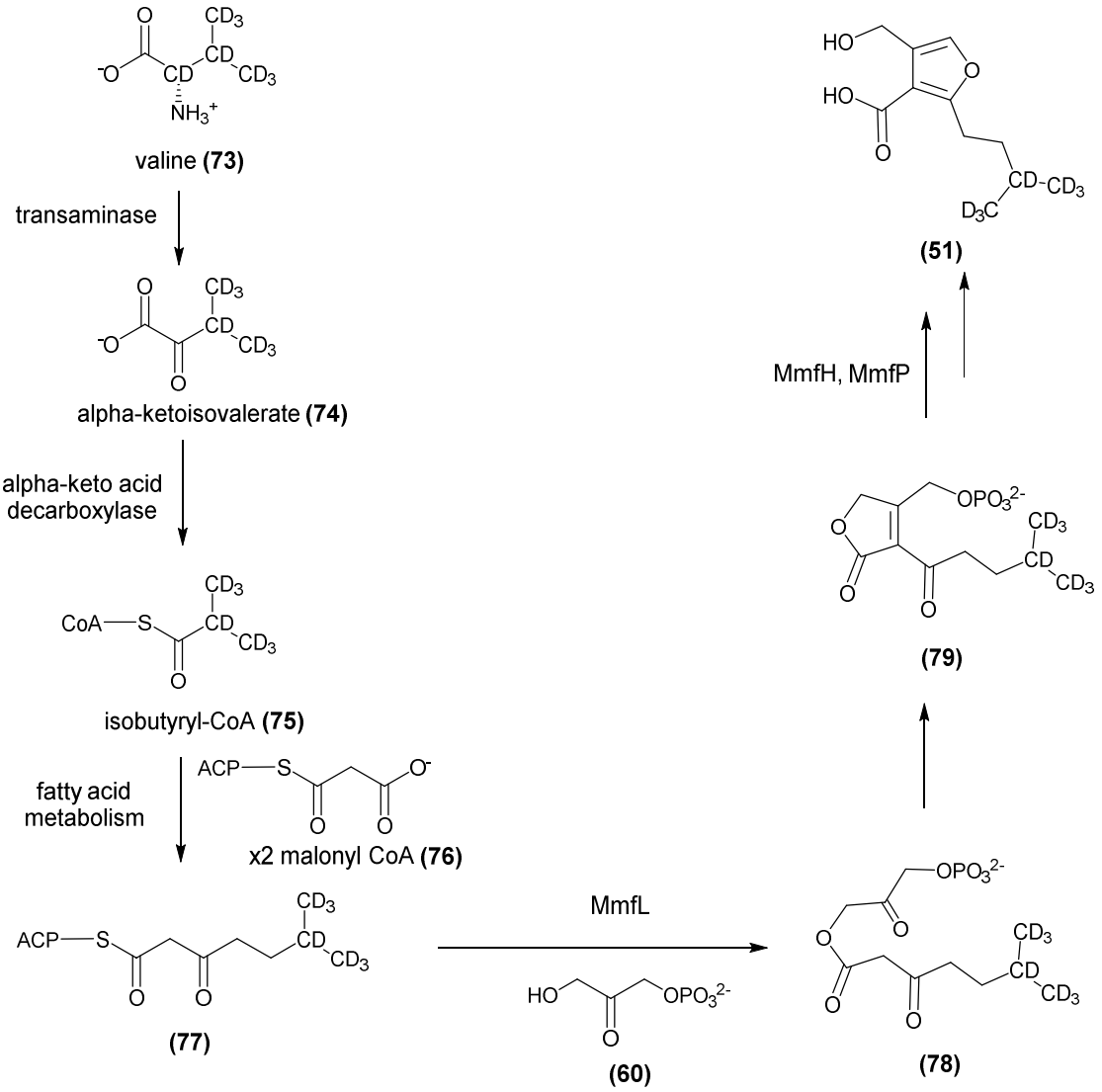
in *S. venezuelae*, however deletion of *jadW3* did not completely abolish jadomycin production.[95] SCO6264 has also been implicated in SCB biosynthesis in *S. coelicolor* but has not been investigated thoroughly.[132]

Therefore, it is possible that the reductases SCO6264, JadW3, BarS1 are stereoselective for keto-type butanolides with specific absolute configurations at carbons 2 and 3. Hence, the substrate specificity of these reductases may determine the stereochemistry at carbon 2 in GBL; they also determine the stereochemistry of carbon 6 by their catalytic activity. More work is needed in order to ascertain the precise function of SCO6264 or JadW3, however.[132,139]

1.5.3 AHFCA biosynthesis

The *afsA* homologue on the SCP1 plasmid of *S. coelicolor*, *mmfL*, has been shown to be essential for biosynthesis of methylenomycin.[79,85] Indeed, an *mmfL* deficient strain was created that no longer produced methylenomycin. However, production of methylenomycin was restored upon addition of ethyl acetate extract from a strain in which *mmfL* was present.[79] Upon introducing the *mmfLHP* genes from *S. coelicolor* A3(2) into a heterologous host which lacks the SCP1 plasmid, it was reported that the new strain constructed produced a series of novel inducer molecules, the 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids (AHFCAs), the structures of which are shown in Figure 1.14.

After the discovery of AHFCAs, Corre and co-workers proposed that the function of MmfL was to catalyse an analogous reaction to that of AfsA, by which the hydroxyl group of dihydroxyacetone phosphate (DHAP) undergoes a condensation reaction with β-ketothioester intermediates from fatty acid metabolism, yielding an ester intermediate **78**. Intermediate **78** then undergoes intramolecular aldol condensation to form a butenolide phosphate intermediate **79** (Scheme 1.4).[79,140] It was also hypothesised that subsequent dephosphorylation and reduction steps, catalysed by MmfP and MmfH



Scheme 1.4 – Proposed biosynthetic route to AHFCAs isolated from *Streptomyces coelicolor* W74, showing the proposed origin of the isobutyryl-CoA incorporated into AHFCA3[79,100,140]

respectively, yield the AHFCA molecules from the butenolide intermediate. MmfH is a flavin-dependent putative dehydrogenase and MmfP is a phosphatase.[140]

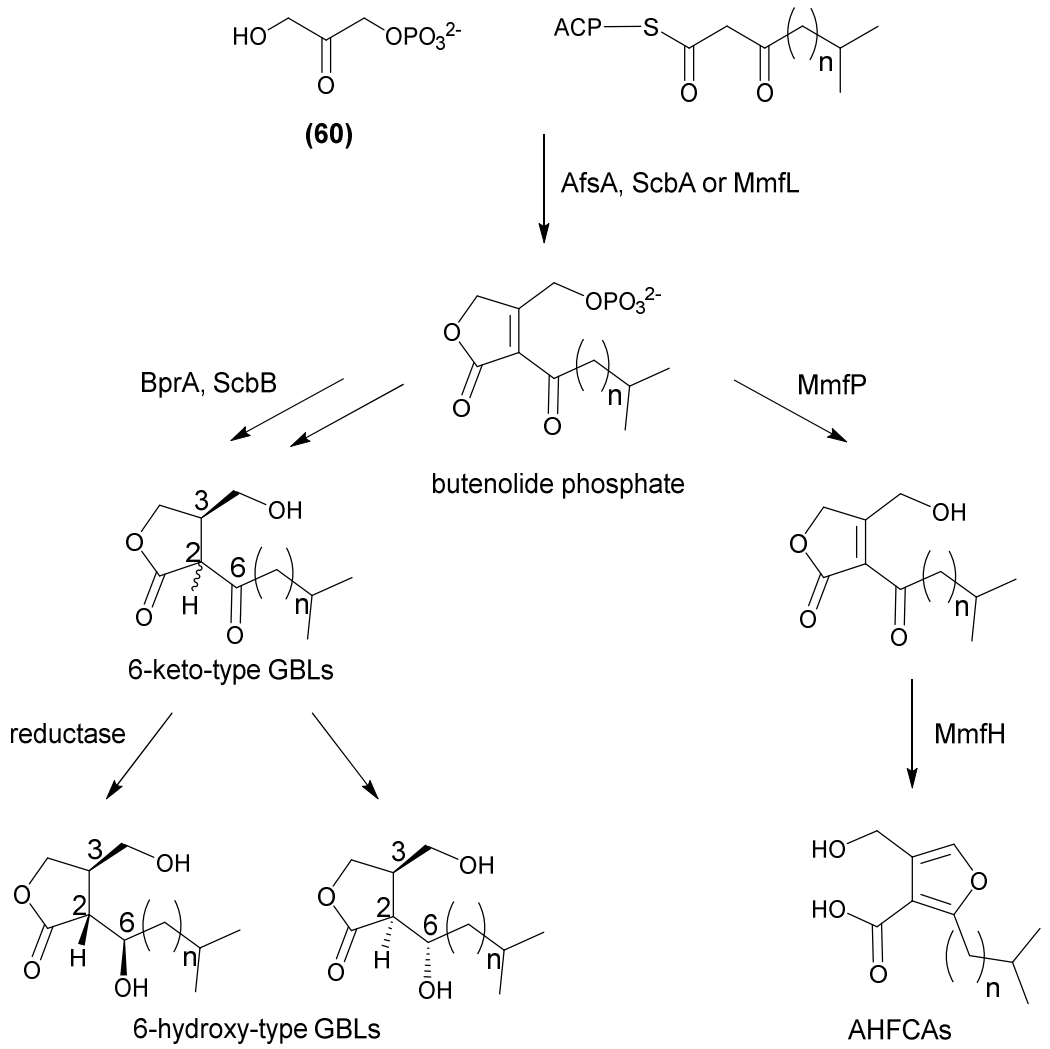
Evidence for this hypothesis was provided by the initial structure elucidation studies of the AHFCA compounds, during which deuterium-labelled precursors for starter units of fatty acid biosynthesis were added to the culture medium of the *S. coelicolor* W74. Mass spectrometry analyses revealed that seven of the deuterium labels from d_8-valine **73** were incorporated into AHFCA3 **51** (Scheme 1.4).[79] Analogous experiments were

performed with each of the proposed precursors that yield the different alkyl chains found in the other AHFCA compounds. Therefore it was deduced that the AHFCAs incorporate intermediates in fatty acid metabolism that are derived from with 3-methylbutyryl-CoA for AHFCA1, butyryl-CoA for AHFCA2 and AHFCA5 and propionyl-CoA for AHFCA4.

Expression of *mmfL* in the heterologous host *E. coli* also provided further evidence for AHFCAs being derived from intermediates in fatty acid metabolism, as only AHFCA2 and AHFCA5 with linear alkyl chains were produced by *E. coli* carrying *mmfL*, consistent with the previously established preference of *E. coli* to utilise acetyl-CoA as the starter unit for fatty acid biosynthesis.[136] This is also consistent with the deduction made by Kato and co-workers made about the A-factor analogues **71** and **72** produced when *afsA* was heterologously expressed in *E. coli* (Figure 1.17).[109] The reactions catalysed by MmfL, MmfH and MmfP have also been studied *in vitro* (N. Malet).[100]

It was suggested by Davis and co-workers that the divergence between GBL and AHFCA biosynthesis may arise due to the orientation of DHAP relative to the β-ketothioester precursors in the reactions catalysed by different AfsA-like enzymes.[141] This hypothesis implied that AHFCA biosynthesis would not proceed via a butenolide intermediate. However, subsequent incorporation experiments performed by Corre and co-workers confirmed the orientation of DHAP with respect to the β-ketothioester, thus providing evidence for AHFCA biosynthesis proceeding via a butenolide intermediate.[140]

Because *in vivo* DHAP is derived from glycerol, as is acetyl CoA (the precursor from which β-ketothioesters in fatty acid biosynthesis are derived), the incorporation of



Scheme 1.5 – General structures of *Streptomyces* signalling molecules which are proposed to share an intermediate butenolide phosphate.[138,140] Labels of enzymes catalysing each step are shown; the specificity of each type (AfsA, MmfL, ScbA, for example) is presumably determined by the length of the alkyl chain in the β-ketothioester precursors

carbon atoms (^{13}C) into the AHFCAs using ^{13}C-labelled glycerols labelled at the pro-*R* and pro-*S* carbon atoms – and subsequent analysis of which fragments the ^{13}C were retained in – was used to confirm the orientation of DHAP with respect to the β-ketothioesters.[140]

Therefore, the formation of GBLs and AHFCAs are proposed to proceed via a shared intermediate butenolide phosphate intermediate.[140] Scheme 1.5 summarises the biosynthetic routes to GBLs and AHFCAs from DHAP and β-ketothioesters from fatty

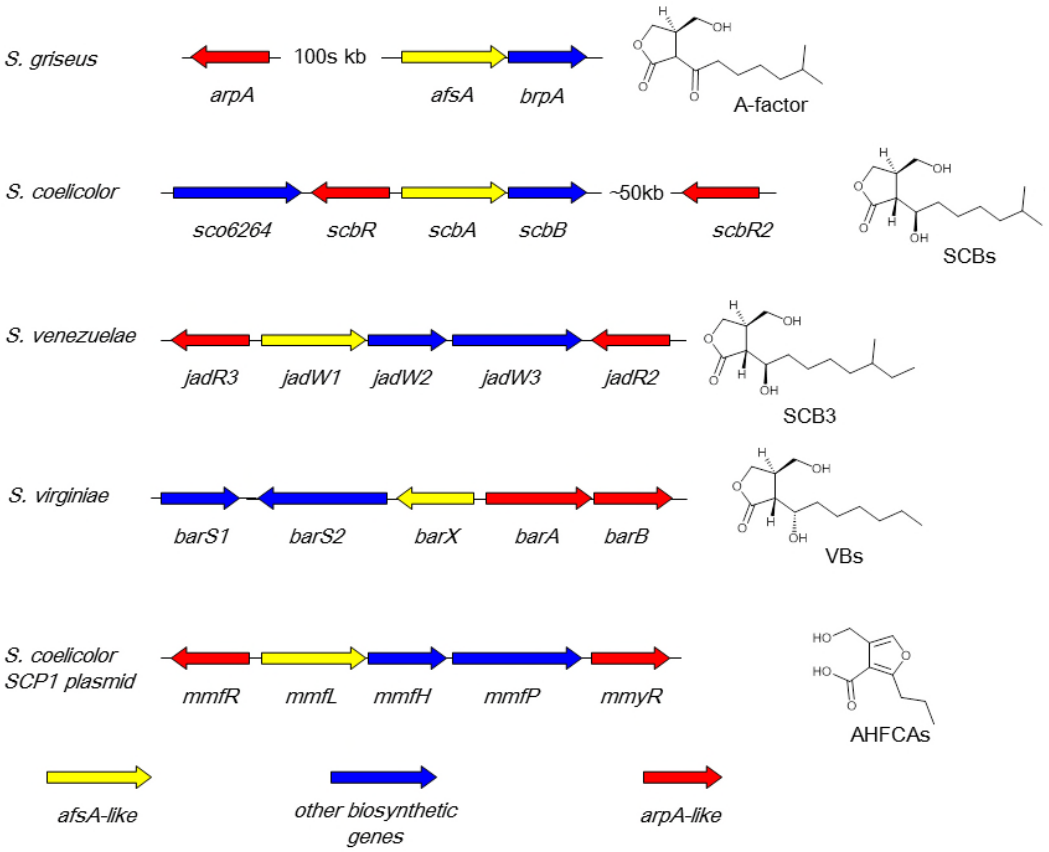


Figure 1.18 – Examples of genetic arrangement *arpA-afsA* like homologues in different *Streptomyces* signalling molecule systems.[79,89,95,117,139] The diversity of structures produced by each system arise from the action of the biosynthetic enzymes shown in yellow/blue

acid metabolism. The genetic arrangement of several clusters involved in GBL-like signalling molecule biosynthesis are shown in Figure 1.18.

1.6.Ligand-protein and protein-DNA interactions of ArpA-like repressors

1.6.1 Determination of ligand-repressor specificity

As discussed in Section 1.4, genes required for signalling molecule biosynthesis are located adjacent to genes encoding for specific ArpA-like repressors. In most cases, there are multiple *arpA*-like genes, one of which is situated divergently adjacent to the *afsA* homologue.[114] Figure 1.18 illustrates examples of the relative arrangement of genes proposed to be involved with signalling molecule biosynthesis and the structures of the signalling molecule(s) they produce.

Ligand-repressor interactions have been investigated in several ArpA-like systems, and have been shown to be highly selective. For example, ScbR from *S. coelicolor* has been shown to have a much greater affinity for 6-hydroxy-type GBLs with (2*R*, 3*R*, 6*R*) stereochemistry compared to 6-hydroxy-type GBLs with (2*R*, 3*R*, 6*S*) stereochemistry.[132] Furthermore, SCB1 with native stereochemistry (2*R*, 3*R*, 6*R*) was shown to be much more effective at releasing ScbR from its target sequence in the *cpkO* promoter than SCB1 with (2*S*, 3*S*, 6*R*) stereochemistry.[132] Furthermore, of the SCB analogues tested with the preferred (2*R*, 3*R*, 6*R*) stereochemistry, there was also a preference for alkyl chains containing between 7 and 9 carbons in the chain, and this correlated well with the structures of the characterised SCBs 1-3 from this system, which all contain 8 or 9 carbons in their alkyl substituents.

In addition to these *in vivo* methods, ligand binding of ArpA-like regulators has been studied *in vitro*. For example, radioisotope-labelled ligands have been used in competition assays to determine the ligand specificity of the *S. lavendulae* ArpA homologue, FarA. These experiments demonstrated that FarA is highly selective for IM-2 **35** (a 6-hydroxy-type GBL with (2*R*, 3*R*, 6*R*) stereochemistry) compared to GBLs with (2*R*, 3*R*, 6*S*) stereochemistry or 6-keto type GBLs.[142] One more widely used method for assaying ligand-repressor interactions is electrophoretic mobility shift assays (EMSAs). The binding of a specific ArpA protein to a particular DNA sequence can be observed by gel electrophoresis by DNA staining in a native polyacrylamide gel. DNA fragments to which a protein binds can be identified by running parallel samples with and without protein. If the protein binds to the DNA, there will be retardation of the DNA migration through the gel, therefore a band shift will be observed (Figure 1.19).[87,125,126] Upon addition of the cognate ligand, the DNA binding affinity of the

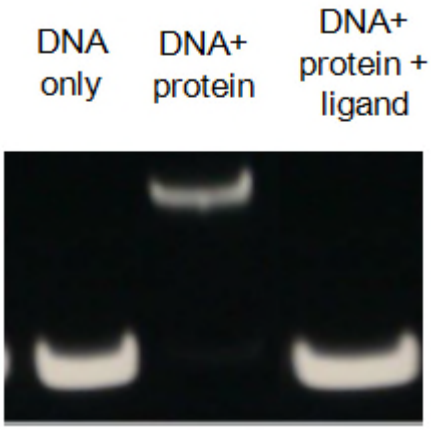


Figure 1.19 –Example of electrophoretic mobility shift assay (EMSA)[100]

protein will be reduced. Therefore, the DNA will be released from the protein and its migration through the gel will no longer be retarded.

EMSAs have been used to demonstrate specificity of *S. virginiae* BarA for 6-hydroxy-type GBLs with (2*R*, 3*R*, 6*S*) stereochemistry over 6-hydroxy-type GBLs with (2*R*, 3*R*, 6*R*) stereochemistry or 6-keto type GBLs.[126] Furthermore, EMSAs have been used in our group to probe the ligand specificity of MmfR for different AHFCA ligands, illustrating that AHFCAs with alkyl chains between 3 and 5 carbons long were most effective at attenuating the ability of MmfR to bind its DNA target in the *mmfR-mmfL* intergenic region.[100] The specificity of each ArpA-like repressor shown by EMSAs correlated well with the structures of the ligands extracted from these organisms in both these cases.

The MmfR structure with AHFCA2 bound is the first example of an ArpA-like repressor structure being solved with its ligand bound, and allowed for the AFHCA binding pocket in MmfR to be identified.[100] Key residues involved with interactions with AHFCA ligand were identified, in particular tyrosines at positions 85 and 144, and

the effect of mutation of these amino acids is ongoing (C. Corre and G. L. Challis groups).

1.6.2 A potential role for ArpA ligands in interspecies signalling

If several different ArpA-like repressors can respond to the same ligand, there is precedent for the GBL or GBL-like signalling molecules to be involved in interspecies signalling.[143] Indeed, SCB3 **38** has recently been shown to be the elicitor of jadomycin biosynthesis.[95] The role of SCB3 in jadomycin biosynthesis was confirmed by chemically complementing a *jadW123* deletion mutant unable to produce SCB3 and jadomycin, upon addition of purified SCB3. In the presence of SCB3, jadomycin production was restored in this strain. Indeed, when growing the *jadW123* deletion mutant adjacent to *S. coelicolor* M145 (a known producer of SCB3), production of jadomycin was also restored.[95]

A complementary experiment was also performed in which the *S. venezuelae* wild type was grown in the presence of a *S. coelicolor* M145 strain in which *scbA* (required for SCB biosynthesis) had been deleted. However, enhanced antibiotic production was not shown in this strain. This was attributed to the concentration of SCB3 produced by the different strains; however, as it was shown that 1 nM SCB3 was sufficient to induce jadomycin biosynthesis in *S. venezuelae*, compared to 128-256 nM required for induction of antibiotic biosynthesis in *S. coelicolor* M145.[95]

These experiments demonstrate that GBL-mediated interspecies signalling may occur between different *Streptomyces* species in nature. Indeed, Sakuda and co-workers previously reported that VB-A **39** can be produced by *S. antibioticus*.[131] A-factor has also been implicated in an indirect mechanism of activation of pimaricin biosynthesis

in *S. natalensis*, even though there is no evidence that *S. natalensis* can itself produce A-factor.[143] Furthermore, A-factor, or an A-factor-like GBL has been implicated in regulation of pristinamycin by *S. pristinaespiralis*.[94,144]

The addition of the A-factor analogue 1,4-butyrolactone has recently been shown to increase validamycin A production in *S. hygroscopicus* and an ArpA-like receptor ShbR3 was identified as a potential target for 1,4-butyrolactone.[145] However, 1,4-butyrolactone was shown to be a very poor ligand for ShbR3 when analysed in EMSAs. Furthermore, the concentration of 1,4-butyrolactone used (10 mM) is unlikely to be biologically relevant.[145] Despite the growing amount of evidence for interspecies signalling, addition of an ArpA ligand to induce production of cryptic metabolite biosynthesis has yet to be demonstrated.

1.6.3 Repressor-DNA interactions

As described in Section 1.4, ArpA binds to a specific DNA sequence in the *adpA* operator region in *S. griseus*. By comparison of the DNA sequences upstream of genes to which ArpA is proposed to bind, a general consensus was proposed by Folcher and co-workers and is shown in Figure 1.20.[144] Collectively, these specific DNA sequences are known as autoregulatory response elements (AREs), due to these being the DNA elements to which the repressor proteins bind, regulating their own expression as well as that of other biosynthetic genes. For example the *mmfL-mmfR* intergenic region is the target for MmfR. AREs are often situated in the operator regions upstream of genes encoding for transcriptional activators, for example *mmyB* in the methylenomycin cluster, and *cpkO* in the coelimcyin cluster.[85,88]

Intergenic region	Sequence	Gene cluster
ARE consensus	**AWACNNACYNNNCGGTTT**	consensus from Folcher[144]
mmyB-mmyT	**AAACCTTCGGGAAGGTTT**	*mmy* (methylenomycin)
gbnA	**ATACCTTCGTGAAGGTTT**	*gbn* (*sven_4179-sven_4181*)
jadR1	**AAACCGGCGCAGCGGTTT**	*jad* (jadomycin)
vmsR	**AAACCGTATAGTCTGTTT**	*vms* (virginiamycin)
tylS	**AAACCGTCCGCTCCGTTT**	*tyl* (tylosin)
alpV	**ACAAACCGACTGTGCTGT**	*alp* (alpomycin)

Figure 1.20 – Examples of ARE consensus sequences from operator regions of transcriptional activators in gene clusters from different *Streptomyces* species.[85,95,114,122,126,144] Conserved nucleotides are highlighted in yellow

The presence of AREs in gene clusters regulated by ArpA-like repressors also demonstrates a further level of complexity within these signalling pathways. For example, the jadomycin biosynthetic (*jad*) cluster in *S. venezuelae* contains four AREs to which the repressor JadR3 binds with different affinities. Indeed, the *far* cluster in *S. lavendulae* also contains five AREs. Furthermore, multiple repressors may also bind the same AREs. In the *scb* cluster for example, ScbR and ScbR2 both bind to one of the AREs in the *scbA-scbR* intergenic region, yet only ScbR has been shown to bind to the other.[149]

The regulatory mechanisms involved can also be complicated by the presence of multiple regulatory proteins. For example in the *jad* cluster, there are four regulators, JadR1, JadR2, JadR3 and JadR*. JadR1 is an OmpR-type transcriptional activator and is negatively regulated by JadR2, but positively regulated by JadR3. JadR2 and JadR3 are both ArpA-like repressors, and JadR2 has been proposed to respond to antibiotics such as jadomycin and chloramphenicol at 500 μM concentrations, whereas JadR3 has been shown to bind a GBL with the same structure as SCB3.[95,146]

Due to the high degree of conservation between the AREs in different systems, AREs can be identified by using bioinformatics tools such as MEME (multiple Em for motif elucidation, http://meme.nbcr.net/meme/).[147] The likelihood of highly similar DNA sequences to occur in specific regions is determined, compared to their likelihood of appearing randomly in nature. Some examples of ARE consensus sequences for GBL systems are shown in Figure 1.20. Experimentally, these elements can be identified by DNaseI footprinting (although this technique shows which DNA regions are protected from being digested by DNaseI, the individual proteins protecting each DNA region must be determined later).[87,95,119]

Therefore, identification of AREs that ArpA-like repressors bind to represents a powerful tool for predicting which genes are controlled by these regulators. Specific inactivation of an ArpA-like repressor may result in de-repression of gene expression and can constitute a strategy to activate silent biosynthetic pathways. However, identifying which specific ArpA-like repressor to delete in order to get overproduction of the specific compound of interest may be very challenging. Indeed, systematic deletion of genes encoding for ArpA-like repressors may have unexpected effects.[150]

1.6.4 Phenotypes of *arpA* deletion mutants

Deletion of *arpA*-like repressors has been shown to regulate antibiotic activity in a variety of *Streptomyces* species.[120,121,144,148] For example, in *S. venezuelae*, deletion of *jadR3* decreases jadomycin production, whereas deletion of *jadR2* increases jadomycin production.[95,149] In *S. coelicolor*, deletion of *scbR2* increases coelimycin production by de-repression of transcription of the *cpkO* gene. It has been shown that several different repressors may bind the same operators, as in the case of the *scb* cluster, in which *scbR* and *scbR2* both bind to the *scbA* operator.[88,149] The ability of JadR2 and JadR3 to bind

Table 1.3 – Effect of deletion of genes encoding for ArpA-like repressors on antibiotic production in host strains

Streptomyces *strain*	*gene cluster*	*Ligand acceptor*	*Natural product biosynthesis regulated*	*Effect of deletion on production*	*Refs*
S. griseus	N/A	ArpA	streptomycin	increased	102
GBL receptor					
S. coelicolor	*cpk/scb*	ScbR	coelimycin	no production	88
S. venezuelae	*jad*	JadR3	jadomycin B	decreased	95
S. lavendulae	*far*	FarA	showdomycin	increased	125
S. virginiae	*vms/bar*	BarA	virginiamycin	increased	126
S. coelicolor	*mmy/mmf*	MmfR	methylenomycin	decreased	85
pseudo receptors					
S. coelicolor	*cpk/scb*	ScbR2	coelimycin	increased	88
S. venezuelae	*jad*	JadR2	jadomycin B	increased	146
S. lavendulae	*far*	FarR2	showdomycin	unknown	125
S. virginiae	*vsm/bar*	BarB	virginiamycin	increased	148
S. coelicolor	*mmy/mmf*	MmyR	methylenomycin	increased	85

to the *jadR1* operator in the *jad* cluster has also been demonstrated.[95,149] Due to the complexity of these regulatory networks, there is no general 'rule' for predicting the phenotype of the *arpA*-deletion mutant by simply choosing to inactivate the 'GBL-receptor' or the 'pseudo-GBL-receptor' (Table 1.3).

In the methylenomycin gene cluster in *S. coelicolor*, there are genes encoding for two repressors, *mmfR* and *mmyR*. Deletion of these two repressors gives distinct phenotypes: a deletion mutant of *mmfR* was shown to be deficient in methylenomycin production, whereas a deletion mutant of *mmyR* overproduced methylenomycin.[79,85]

1.6.5 Regulatory cascade controlling methylenomycin biosynthesis

Methylenomycin production was also abolished in a strain in which *mmfLHP* had been deleted. Furthermore, when *mmfR* and *mmyR* were each introduced into this *mmfLHP* deletion mutant, it was found that methylenomycin production was only restored in the

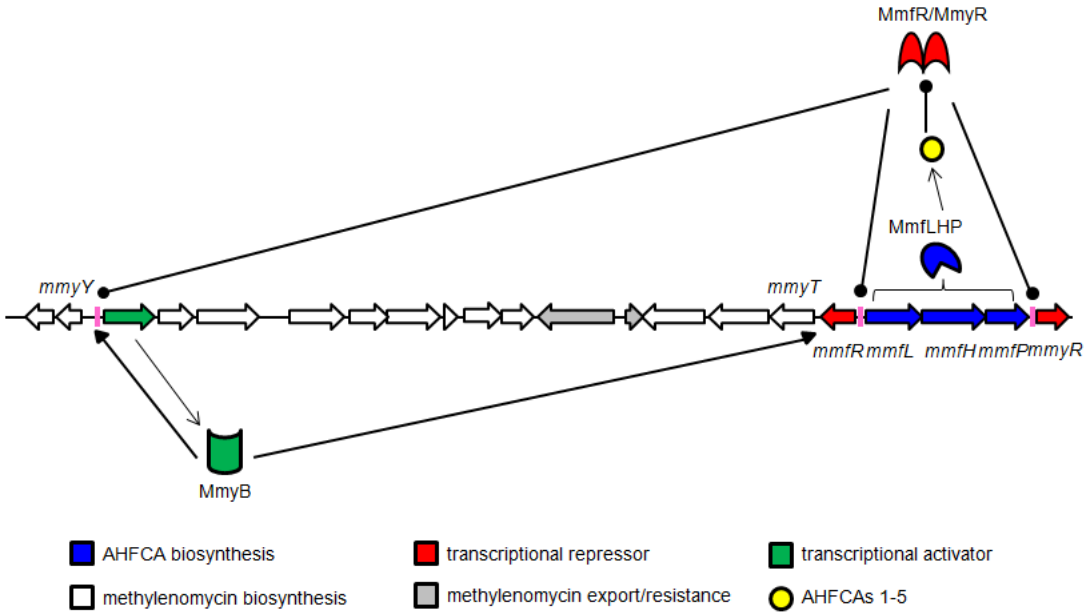


Figure 1.21 – Signalling cascade involved in methylenomycin biosynthesis.[85] ARE sequences are shown in pink. Stimulatory events are denoted by blocked arrowheads, inhibitory events are denoted by circular arrowheads

mmfR-containing strain when grown in close proximity to an AHFCA-secreting strain. The same inductive effect of AHFCAs was not observed in the *mmfLHP* strain to which *mmyR* had been introduced, implying that MmfR is the effector protein which binds AHFCAs, not MmyR.[85]

As in other *afsA*-*arpA*-like repressor gene clusters, the *afsA* homologue *mmfL* is divergent to one of the *arpA* homologues, *mmfR*. The AREs identified in the *mmy*-*mmf* gene cluster are situated in intergenic regions of *mmfR*-*mmfL*, *mmyB*-*mmyY* and the region upstream of *mmyR* (Figures 1.20 and 1.21). MmyB is a transcriptional activator belonging to the unusual MltR (MmyB-like transcriptional regulator) family of regulators.[151] It was hypothesised that the ArpA homologues MmfR and/or MmyR repress biosynthesis of the *mmyB*, *mmfL*, *mmfR* and *mmyR*, regulating methylenomycin biosynthesis.[85]

Additional transcriptional analyses revealed that there was a small amount of *mmfL*, *mmyR* and *mmfR* transcription between 12 and 36 hours, with *mmyB* beginning to increase between 36 and 48 hours, before *mmy* biosynthetic genes were transcribed. The presence of TTA codons in the *mmyB* and *mmfL* genes was also noted, implying that translation of MmfL and MmyB may be delayed until the late stages of growth, as the leucinyl-tRNA corresponding to a TTA codon is not produced until a later stage in the developmental cycle.[85]

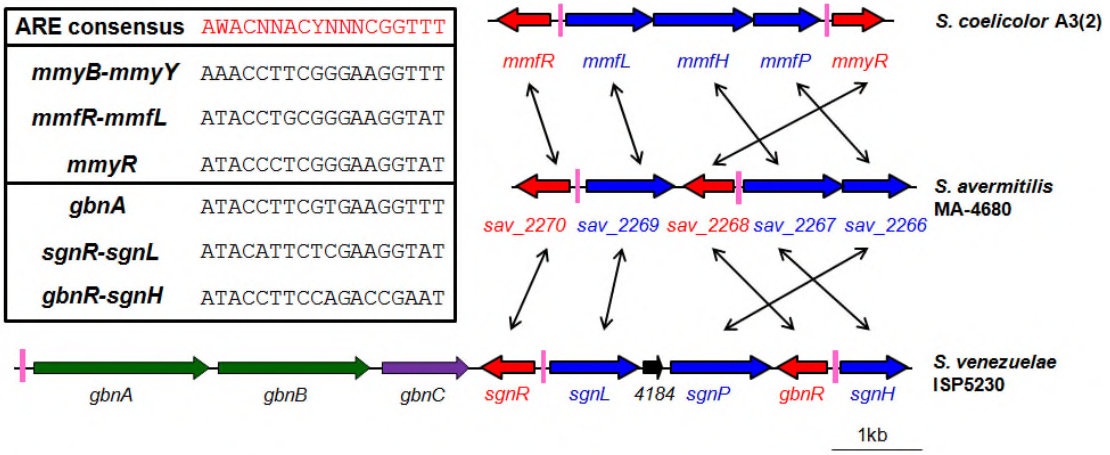
Based upon these observations, a sequence of regulatory events in methylenomycin biosynthesis was proposed. Initially, there is a small amount of *mmfL*, *mmfR* and *mmyR* transcription. This leads to a low concentration of AHFCA molecules being synthesised by MmfLHP. A small amount of MmfR and MmyR protein is also being produced, which is enough to repress transcription of *mmyB*. Eventually, a critical AHFCA concentration is reached, where there is a sufficiently high concentration to cause de-repression of MmfR binding to the *mmfL-mmfR* intergenic region. This leads to a rapid increase in the amount of *mmfLHP* transcription, therefore the intracellular concentration of AHFCAs increases. AHFCAs are then exported into the extracellular milieu, thus synchronising the de-repression of the *mmfL-mmfR* promoter between neighbouring cells. The intracellular concentration of AHFCAs is now great enough to cause de-repression of the *mmyB* promoter, therefore MmyB can be produced. MmyB is proposed to bind to the operator regions of *mmyB-mmyY* (in pseudo-palindromic regions known as a 'B-boxes', which are distinct from the ARE) and *mmyT*, therefore activating transcription of the structural genes in the *mmy* cluster, leading to methylenomycin production.[85] This regulatory cascade is summarised in Figure 1.21.

It has recently been shown (in our group and G. L. Challis group) that MmfR is responsive to AHFCAs *in vitro*. Addition of the AHFCAs 1-5 isolated from *S. coelicolor* to MmfR and a DNA fragment containing the *mmfL-mmfR* intergenic region caused disassembly of the MmfR protein-DNA complex, confirming that MmfR is a furan-responsive regulator.[100] Furthermore, the crystal structure of MmfR with AHFCA2 bound has also been solved, allowing key residues involved in AHFCA binding to be proposed. These residues are not conserved between MmfR and MmyR, which is consistent with the hypothesis that MmyR is unresponsive to AHFCAs.

1.7 *mmfLHP*-like biosynthetic gene clusters

1.7.1 Identification of other *mmfLHP* clusters

During the initial study in which AHFCAs were established as being the products of the *mmfLHP* genes, it was noted that an analogous set of genes, *sav_2269, sav_2267 and sav_2266* is present in *Streptomyces avermitilis* MA-4680 (Figure 1.22). This mini-cluster of genes also contains genes encoding for two putative transcriptional regulators, *sav_2270* and *sav_2268*.[79] BLAST analyses revealed the presence of a similar cluster of genes in *S. venezuelae*, named *sgnL (sven_4183)*, *sgnP (sven_4185)* and *sgnH (sven_4188)*, after their homologues in *S. coelicolor* (Figure 1.22).[152] These genes are located in a mini-cluster which also contains genes encoding for homologues of MmfR and MmyR, named SgnR (SVEN_4182) and GbnR (SVEN_4187) respectively (Figure 1.22). The *mmfLHP*-like system in *S. venezuelae* is situated adjacent to the *sven_4179-sven_4181* gene cluster, which had not been previously investigated. It was hypothesised that if the key repressor was deleted, this would lead to overexpression of the *sven_4179-sven_4181* gene cluster, revealing new natural products.[152]



Protein	Homologue (% identity, % similarity)	Proposed function
SgnR	MmfR (54, 71)	AHFCA-binding transcriptional repressor
SgnL	MmfL (33, 44)	butenolide synthase
SgnP	MmfP (42, 56)	phosphatase
GbnR	MmyR (47, 54)	transcriptional repressor
SgnH	MmfH (45, 57)	flavin-dependent dehydrogenase

Figure 1.22 – Analogous *mmfLHP* systems identified in *S. avermitilis* and *S. venezuelae*. The *mmfLHP* system in *S. venezuelae* is located adjacent to the cryptic *gbnABC* cluster.[79,152] AREs are highlighted in pink, and *S. coelicolor* and *S. venezuelae* AREs compared with the general consensus proposed by Folcher and co-workers[144] are shown are shown. W = A/T, Y = C/T

The genome of *S. venezuelae* was recently sequenced, revealing it to contain 31 gene clusters encoding for secondary metabolite gene clusters (Figure 1.23).[63] Interestingly, the *sven_4179-sven_4181* gene cluster was not identified by antiSMASH, as these genes are atypical in secondary metabolite pathways.

S. venezuelae is known to produce chloramphenicol, jadomycin and the macrolides pikromycin, narbomycin and methymycin.[13,39,153] Recently, a series of phenoxazine alkaloids and aminophenols, venezuelines A-G, have also been shown to be produced by *S. venezuelae*.[154] This relatively short list of natural products produced by *S. venezuelae* illustrates that the biosynthetic potential for *S. venezuelae* is still largely unexplored.

Identified secondary metabolite clusters

Cluster	Type	From	To
The following clusters are from record FR845719.1:			
Cluster 1	Ectoine	237842	248258
Cluster 2	Terpene	274553	296739
Cluster 3	T1pks-t3pks-nrps	504136	604067
Cluster 4	Lantipeptide-terpene	614220	645285
Cluster 5	Lantipeptide	707463	730315
Cluster 6	Indole	867489	890695
Cluster 7	Other	1031023	1073914
Cluster 8	Other	2055965	2096690
Cluster 9	Siderophore	2794973	2806751
Cluster 10	Bacteriocin	3028824	3039972
Cluster 11	Other	4408196	4451900
Cluster 12	Butyrolactone	4522134	4533171
Cluster 13	Melanin	5003818	5014228
Cluster 14	Butyrolactone	5477370	5502716
Cluster 15	Thiopeptide	5516950	5557501
Cluster 16	T3pks	5785193	5826323
Cluster 17	Siderophore	5869901	5883169
Cluster 18	Siderophore	5935668	5950407
Cluster 19	Bacteriocin	6350466	6361866
Cluster 20	Butyrolactone-t2pks	6494470	6531416
Cluster 21	Other	6672467	6716369
Cluster 22	Nrps	6720590	6793506
Cluster 23	Nrps	6800192	6855167
Cluster 24	Terpene	7021575	7048100
Cluster 25	Bacteriocin	7128838	7139692
Cluster 26	T2pks	7421589	7464101
Cluster 27	Melanin	7484949	7495338
Cluster 28	Nrps	7706602	7760938
Cluster 29	Terpene	7788497	7809951
Cluster 30	T3pks	7946080	7987237
Cluster 31	Terpene-nrps	8189935	8226158

Figure 1.23 –Secondary metabolite gene clusters identified in *S. venezuelae* by antiSMASH (October 2014)

1.7.2 Previous work on *gbnABC* system in *S. venezuelae*

By comparing the promoter sequences of the *gbnABC* genes to those in the *sgnR-sgnLPH-gbnR* mini-cluster in *S. venezuelae*, AREs were identified in the operator regions of *gbnA*, *sgnR-sgnL* and *gbnR-sgnH* (Figure 1.22). These AREs identified in *S. venezuelae* are almost identical in sequence to the ARE consensus from the *mmf-mmy* cluster proposed by O'Rourke and co-workers (Figure 1.21). The presence of an ARE in the *gbnA* promoter implied that transcription of *gbnA* is regulated by SgnR/GbnR, as in the case of MmfR/MmyR-mediated regulation of *mmyB* in the methylenomycin cluster.

Therefore, it was hypothesised that *sgnLPH* are transcribed, leading to an increase in the intracellular concentration of AHFCAs. At a critical concentration, AHFCAs bind to SgnR, causing it to dissociate from its DNA targets (the AREs located in the *gbnA*, *sgnR-sgnL* and *gbnR-sgnH* operators). As a result, transcription of *gbnABC* can occur, leading to biosynthesis of new natural product(s). It is proposed that GbnR would bind to the ARE located in the *gbnA* promoter and switch off transcription of *gbnABC* (Figure 1.24).

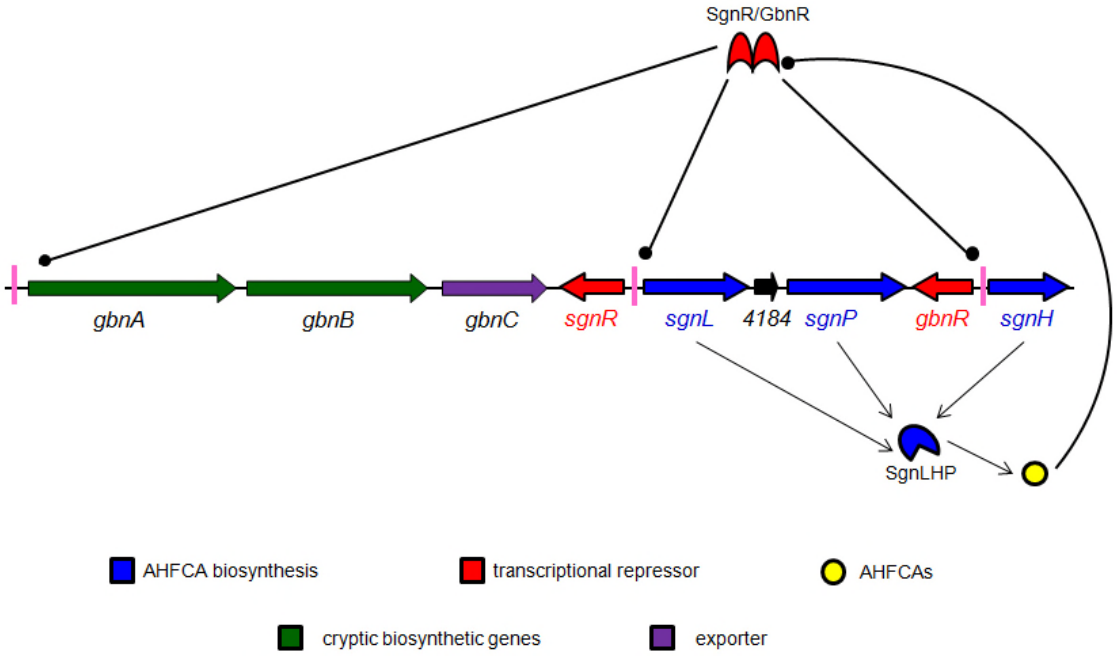


Figure 1.24 – Proposed regulatory mechanism involved in regulation of the cryptic *gbn* cluster

As reported previously, deletion of the transcriptional repressor protein *mmyR* in *S. coelicolor* leads to overproduction of AHFCAs by de-repression of the *mmfR*/*mmfL* promoter, in addition to triggering overproduction of methylenomycin via de-repression of transcription of *mmyB*. As the *S. venezuelae sgnR-sgnL* and *gbnA* promoters both contain AREs, it seemed likely that deletion of the *mmyR* homologue *gbnR* in *S. venezuelae* could therefore lead to overproduction of AHFCA-like signalling molecules and also to overproduction of the natural products encoded for by the *gbnABC* cluster, as the 'off switch' GbnR protein would no longer be produced.

If this hypothesis proved to be true, then deletion of *mmyR*-like transcriptional repressors from gene clusters which contain the *mmfR-mmfLHP-mmyR*-like mini-cluster represents a powerful new strategy for natural product discovery in a range of *Streptomyces* species. Therefore, as a proof-of-concept study, a *S. venezuelae gbnR* deletion mutant was constructed and the metabolites produced by the *gbnR* deletion mutant were compared with the wild type.[152]

LC-MS analyses of the metabolites extracted from these cultures revealed that new metabolites **80-85** were produced by the *gbnR* mutant. A common neutral loss of 129 Da corresponding to a $C_5H_7NO_3$ fragment was observed, and tandem mass spectrometry revealed each $[M–C_5H_7NO_3+H]^+$ ion to correspond to a different amino acid. Partial NMR structures elucidated for the metabolites **80**, **81/82**, **83** and **85** were found to contain a conserved $CH_2CH_2CH_2CO$ moiety.[152]

1.8 Aims of project

- Complete structure elucidation of gaburedins A-F (**80**-**85**) by MS, NMR and HPLC analyses and by synthesis of an authentic standard.
- Link the production of gaburedins to the *gbn* gene cluster by comparing metabolites produced by *S. venezuelae* wild type, *gbnR* mutant and a double-deletion *gbnB*/*gbnR* strain.
- Propose a biosynthetic route to gaburedin biosynthesis based upon bioinformatics and experimental analysis including incorporation of proposed precursor molecules and analogues.
- Reconstitute the GbnA, GbnB, GbnC biosynthetic pathway by heterologous expression of proposed biosynthetic genes *gbnA* and *gbnB* in *E. coli*.
- Determine the nature of any signalling molecules overproduced by the *S. venezuelae gbnR* mutant and develop an assay to investigate their role in regulation of gaburedin biosynthesis.

1.9 Outline of thesis

Chapter 2 focuses on the structure elucidation of the gaburedins – synthesis of an authentic standard of gaburedin A is discussed, along with feeding experiments to verify the structure of each of the gaburedins A-F. Chiral HPLC analysis of culture extracts is also discussed, to verify the stereochemistry of the cognate amino acids.

Chapter 3 includes bioinformatic analyses of each of the GbnA, GbnB and GbnC proteins. The identification of a key gene for gaburedin biosynthesis is discussed and a biosynthetic route to gaburedins is presented. Incorporation experiments with stable isotope precursor molecules are used to probe the proposed biosynthetic route. The substrate tolerance of enzymes GbnA and GbnB is also discussed. Chapter 4 includes feeding experiments to produce a library of gaburedin analogues to further investigate diversity of aminoacyl moieties which can be incorporated into gaburedin biosynthesis.

Chapter 5 details attempts to reconstitute the *S. venezuelae* GbnA, GbnB biosynthetic pathway *in vitro* by heterologous expression of *gbnA* and *gbnB* in *E. coli* and overproduction of recombinant proteins. The construction of a series of vectors for heterologous expression of the complete *gbnABC* gene cluster is also reported.

Chapter 6 deals with the characterisation of the signalling molecules produced by the *S. venezuelae gbnR* deletion mutant. The development of an assay to demonstrate the ability of some of these signalling molecules to induce gaburedin biosynthesis is also discussed. Finally, Chapter 7 summarises the findings of this PhD project and discusses the context of the results herein. Some analogous *mmfLHP*-like systems in other *Streptomyces* species identified during the course of this project are also discussed.

2. Identification and characterisation of the novel natural products gaburedins A-F from *Streptomyces venezuelae*

Preliminary work comparing the metabolites produced by the *Streptomyces venezuelae* wild type and *gbnR::apra* strain revealed that there is a series of metabolites overproduced by the *gbnR::apra* strain (Section 1.7). These metabolites were shown to have *m/z* values corresponding to $[M+H]^+$ of 294.9 (compound **80**), 261.0 (compounds **81** and **82**), 247.0 (compound **83**), 279.0 (compound **84**) and 292.9 (compound **85**). These compounds all exhibited a common fragmentation pattern in electrospray ionisation mass spectrometry, with loss of a neutral 129 Da fragment corresponding to a $C_5H_7NO_3$ moiety.[152]

Fully elucidating the structures of these new metabolites would provide the first example of new natural products to be discovered by deletion of an ArpA-like transcriptional regulator. The diverse structures of this family of metabolites are discussed in Chapters 3 and 4.

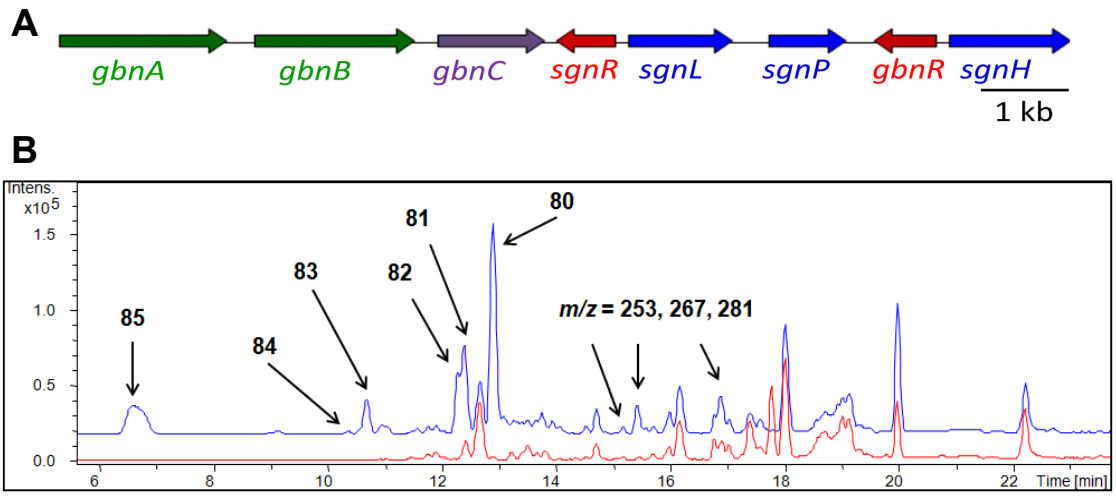


Figure 2.1 – Panel A: genetic arrangement of the *gbn/sgn* cluster in *S. venezuelae*. Panel B: Base peak chromatograms of metabolites extracted at pH 3 from the *Streptomyces venezuelae* wild-type (red trace) and *gbnR* mutant (blue trace) after being grown on SMMS for 3 days[152]

A series of additional metabolites unrelated in structure to those mentioned above were also were overproduced in the *gbnR* strain, eluting between 14 and 18 minutes (Figure 2.1). The nature of these additional molecules will be discussed in Chapter 6.

2.1 Growth conditions

2.1.1 Production of metabolites **80-85** in *S. venezuelae gbnR::apra* mutant

In order to confirm the presence of compounds **80-85**, the *S. venezuelae gbnR* mutant was grown on both mannitol-yeast medium (MYM) and supplemented minimal medium (SMM). MYM is a rich medium typically used for spore production, with mannitol being the primary carbon source and SMM is a simple production medium for secondary metabolites, with glucose as the primary carbon source.[155] Richer media contain more sources of carbon and nitrogen, therefore are favoured for activating the biosynthesis of secondary metabolites. Compounds **80-85** were produced on both media, however, there were fewer different metabolites produced by *S. venezuelae* when grown on SMMS. Therefore, the *gbnR::apra* strain was grown on SMMS in future experiments.

To confirm that the overproduction of these new metabolites **80-85** in the *gbnR::apra* strain is only observed in the *gbnR* deletion mutant, this strain was grown for 3 days on supplemented minimal media solid (SMMS), cultures acidified to pH 3, and the metabolites extracted. The metabolites produced were compared to those produced by the *S. venezuelae* wild-type and *sgnL-P::apra* strains when grown on SMMS for 3 days. Metabolites were analysed by liquid chromatography-mass spectrometry (LC-MS) analyses, confirming that the metabolites **80-85** are overproduced in the *gbnR::apra* strain relative to the wild type and the *sgnL-P::apra* strain (Figure 2.2).

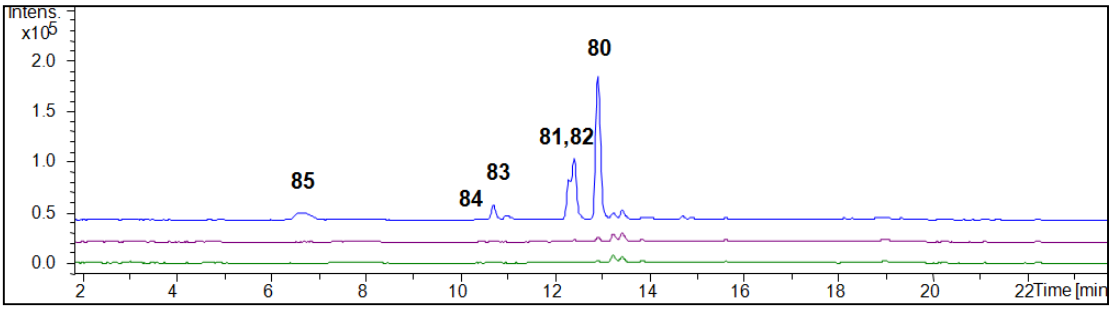


Figure 2.2 – Extracted ion chromatograms for *m/z* = 247.0, 261.0, 278.9, 292.9 and 294.9 for metabolites **80-85** extracted from the *Streptomyces venezuelae* wild type (green trace), *gbnR* mutant (blue trace) and *sgnL-P* mutant (purple trace) at pH 3 after being grown on SMMS for 3 days

As discussed in Section 1.6.2, it was expected that the wild type would not overproduce the metabolites of interest due to the regulatory gene *gbnR* being present, therefore GbnR is still able to act as an 'off' switch by binding to the *gbnA* promoter, preventing transcription. Therefore, transient transcription of *gbnABC* genes was predicted. Similarly, the *sgnL-P* mutant is missing the *afsA*-like gene *sgnL* thought to be essential for the biosynthesis of the AHFCA signalling molecules controlling the transcriptional repressor/DNA interactions, therefore SgnR protein will stay bound to the ARE in the *gbnA* promoter, and will not be dissociated to allow transcription of *gbnABC*. Therefore in both the wild type and *sgnL-P* mutant, the natural products biosynthesised by GbnABC are expected to be observed in much smaller quantities than in the *gbnR* mutant (if they are observed at all).

It was found that after both 3 and 6 days of growth, the yield of the metabolites extracted was greater when the culture media was acidified prior to extraction with ethyl acetate. Furthermore, the relative abundance of the metabolites of interest after 3 and 6 days growth appeared to be largely unchanged. Hence, in future experiments on SMMS, *S. venezuelae* strains were only grown for 3 days.

2.1.2 Culture conditions for maximising production of *S. venezuelae gbnR::apra* metabolites **80-85**

This experiment was extended to monitor the production of these metabolites as a function of time; plates were inoculated with *gbnR* mutant spores and grown for 24, 48 or 72 days before the metabolites were extracted.

Upon extraction, it was found that the yield of these metabolites was almost unchanged after 1 day of growth compared with 3 days of growth (Figure 2.3). However, there were notable changes in the base peak chromatograms and sporulation phenotypes between days 1 and 3 of growth (data not shown). It was therefore decided to grow future cultures for 3 days.

S. venezuelae is known to grow well in liquid culture compared to other *Streptomyces*

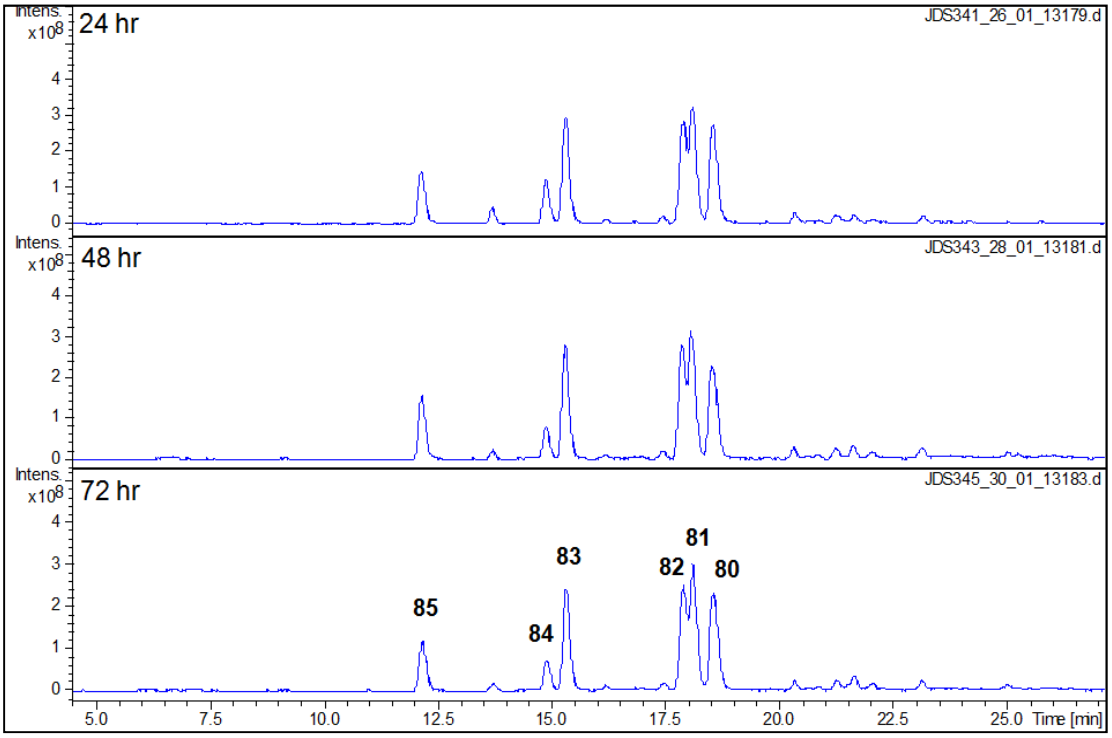


Figure 2.3 – Extracted ion chromatograms *m/z* = 247.0, 261.0, 278.9, 292.9 and 294.9 for metabolites **80-85** extracted from the *Streptomyces venezuelae gbnR* at pH 3 after being grown on SMMS for 24, 48 and 72 h

species, which are prone to clumping or do not sporulate efficiently.[5] However it was discovered that production of the metabolites **80**-**85** was not consistent in liquid media, therefore future attempts to scale up production of these metabolites were focussed on solid SMM.

2.2 Structure elucidation

2.2.1 Liquid chromatography-mass spectrometry (LC-MS) analysis

The metabolites **80**-**85** overproduced by the *S. venezuelae gbnR* mutant were analysed by ultra-high resolution LC-MS (UHR-LC-MS). The mass spectra of these compounds all displayed the characteristic fragmentation pattern corresponding to the presence of the $[M+H]^+$ and $[M+Na]^+$ species, in addition to a $[M–129+H]^+$ fragment (Figure 2.4).

UHR-LC-MS analyses also allowed for molecular formulae to be obtained for each of the species of interest. For example, compound **80** – the metabolite in this series produced in the greatest abundance – has a molecular ion $[M+H]^+$ corresponding to $C_{14}H_{19}N_2O_5$ (calculated *m/z* = 295.1288, found 295.1287), a $[M+Na]^+$ species with formula $C_{14}H_{18}N_2O_5Na$ (calculated *m/z* = 317.1108, found 317.1105), and also a $[M–C_5H_7NO_3+H]^+$ fragment ion corresponding to a molecular formula $C_9H_{12}NO_2$ (calculated *m/z* = 166.0863, found 166.0864), verifying the loss of a neutral $C_5H_7NO_3$ fragment consistent with the mass loss of 129 Da. Compounds **81**-**85** exhibit similar fragmentation patterns, confirming the loss of a neutral $C_5H_7NO_3$ fragment from all six compounds, summarised in Table 2.1. The mass spectrum fragmentation patterns observed are consistent with previously reported MS-MS experiments by which a parent protonated or sodiated adduct of each compound **80**, **81**/**82**, **83** and **85** were isolated and then fragmented further and shown to lose this 129 Da fragment.

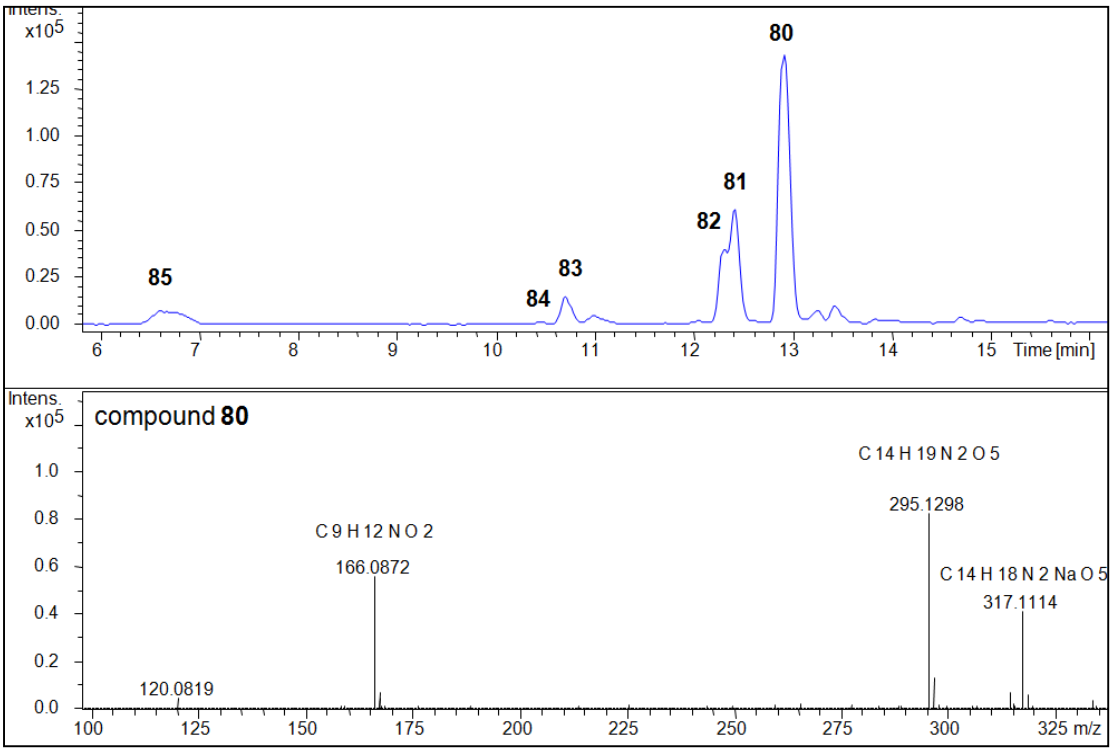


Figure 2.4 – UHR-LC-MS trace of metabolites extracted from *S. venezuelae gbnR::apra* mutant. Extracted ion chromatogram corresponds to *m/z* = 295.1, 261.1, 247.1, 279.1, 293.1 for metabolites **80-85**. The mass spectrum of compound **80** is also shown, with molecular formulae for molecular and fragment ions

It was noted that the fragments remaining after loss of the $C_5H_7NO_3$ fragments from compounds **80-84** corresponded to *m/z* values and molecular formulae of protonated amino acids. For compound **85**, $C_{10}H_{17}N_2O_6S$ an additional fragment ion corresponding to a molecular formula of $C_3H_8NO_2S$ was also observed (expected *m/z* = 122.0270, found 122.0281) and was assigned as a $[M–C_5H_7NO_3–C_2H_2O+H]^+$ daughter ion (Table 2.1) corresponding to $C_3H_8NO_2S$, the molecular formula of protonated cysteine. A summary of the parent and fragment ions of **80-85** is shown in Table 2.1.

The observation that all metabolites **80-85** lose a conserved $C_5H_7NO_3$ fragment strongly implied that they belong to the same series of natural products. The difference in the retention times on the reverse phase HPLC column also correlate with the relative hydrophobicity of the amino acid fragments of these metabolites. Indeed, the

Table 2.1 – UHR-MS assignments of molecular and fragment ions for metabolites **80-85**

calculated m/z *value*	*observed* m/z *value*	*Molecular formula*	*Assignment*
compound **80**			
317.1108	317.1114	$C_{14}H_{18}N_2NaO_5$	$[M+Na]^+$
295.1288	295.1298	$C_{14}H_{19}N_2O_5$	$[M+H]^+$
166.0863	166.0872	$C_9H_{12}NO_2$	$[M+H–C_5H_7NO_3]^+$, phe
compound **81**			
283.1264	283.1260	$C_{11}H_{19}N_2NaO_5$	$[M+Na]^+$
261.1445	261.1441	$C_{11}H_{20}N_2O_5$	$[M+H]^+$
132.1019	132.1019	$C_6H_{14}NO_2$	$[M+H–C_5H_7NO_3]^+$, leu/ile
compound **82**			
283.1264	283.1260	$C_{11}H_{19}N_2NaO_5$	$[M+Na]^+$
261.1445	261.1440	$C_{11}H_{20}N_2O_5$	$[M+H]^+$
132.1019	132.1018	$C_6H_{14}NO_2$	$[M+H–C_5H_7NO_3]^+$, leu/ile
compound **83**			
269.1108	269.1114	$C_{10}H_{18}N_2NaO_5$	$[M+Na]^+$
247.1288	247.1295	$C_{10}H_{19}N_2O_5$	$[M+H]^+$
118.0863	118.0874	$C_5H_{12}NO_2$	$[M+H–C_5H_7NO_3]^+$, val
compound **84**			
301.0829	301.0834	$C_{10}H_{18}N_2NaO_5S$	$[M+Na]^+$
279.1009	279.1015	$C_{10}H_{19}N_2O_5S$	$[M+H]^+$
150.0583	150.0591	$C_5H_{12}NO_2S$	$[M+H–C_5H_7NO_3]^+$, met
compound **85**			
315.0621	315.0622	$C_{10}H_{16}N_2O_6SNa$	$[M+Na]^+$
293.0802	293.0809	$C_{10}H_{17}N_2O_6S$	$[M+H]^+$
164.0376	164.0382	$C_5H_{10}NO_3S$	$[M+H–C_5H_7NO_3]^+$, acetyl-cys
122.0270	122.0281	$C_3H_8NO_2S$	$[M+H–C_5H_7NO_3\text{-}C_2H_2O]^+$, cys

more polar, hydrophilic metabolites **84** and **85** which appear to contain methionine or an acetyl derivative of cysteine, respectively, elute sooner than the metabolites **80-83** which appear to be derived from more hydrophobic, non-polar amino acids such as phenylalanine and leucine/isoleucine.

To acquire enough material for NMR characterisation, the *gbnR* mutant was grown on 800 mL SMMS (16 x 50 mL plates) for 3 days, the agar acidified and ethyl acetate extraction performed at pH 3. The residue was analysed by LC-MS in order to confirm the presence of the metabolites of interest, before the extract was introduced to a

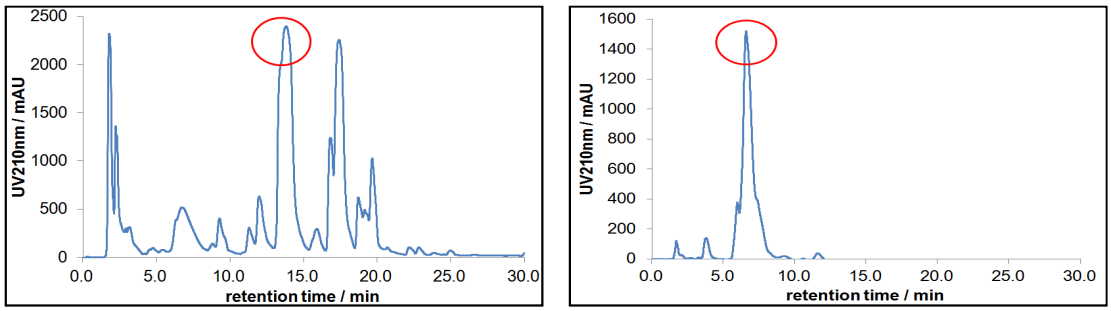


Figure 2.5 – HPLC traces from purification of compound **80** from 300 µL injection of crude *S. venezuelae gbnR::apra* extract. The peak collected in red during the first round (left panel) was collected and re-injected (right panel)

preparative high performance liquid chromatography (HPLC) system. LC-MS analyses had previously shown the compounds of interest to absorb at 210 nm (Appendix 1), therefore absorption at 210 nm was used to guide the HPLC purification. The closely eluting compounds **80**, **81** and **82** proved difficult to separate, so were isolated as a mixture, re-extracted from the HPLC solvents and re-injected into the HPLC (Figure 2.5).

2.2.2 Characterisation of metabolite **80** by nuclear magnetic resonance (NMR) spectroscopy

NMR spectra (1H, COSY, PENDANT, HSQC, HMBC) were obtained for the compound **80** and a structure was proposed (Figure 2.6, see Section 8.1 for full assignments). The structure proposed was consistent with previous data obtained for crude samples of compounds **80**-**83** and **85** illustrating the presence of the conserved γ-aminobutyrate (GABA)-like moiety (H-l, H-m, and H-n all representing 2 protons in a $CH_2CH_2CH_2$ chain located adjacent to an amide proton H-k). Furthermore, COSY correlations between the aromatic protons H-d and H-c, HMBC correlations between H-d and aromatic carbons C-e and C-f suggested the presence of a phenylalanine-like moiety also to be present in compound **80**. This observation is consistent with the fragment ion *m/z* = 166.0872 observed in the mass spectrum of compound **80**.

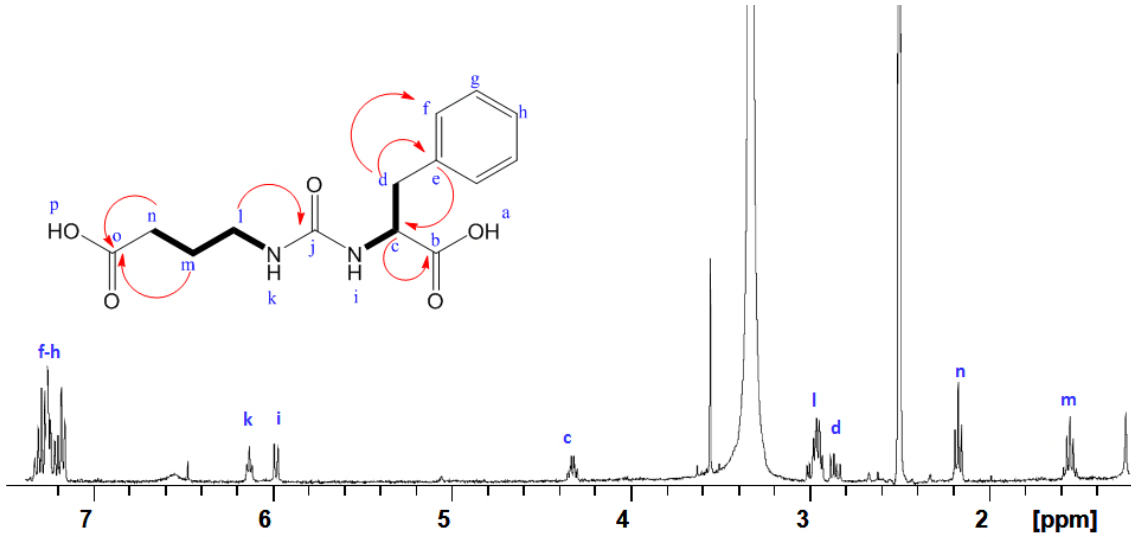


Figure 2.6 – Proton NMR spectrum and assignments of gaburedin A (**80**) in d_6-DMSO (700 MHz). 2D correlations are also shown (COSY correlations in bold bonds, HMBC correlations denoted by red arrows)

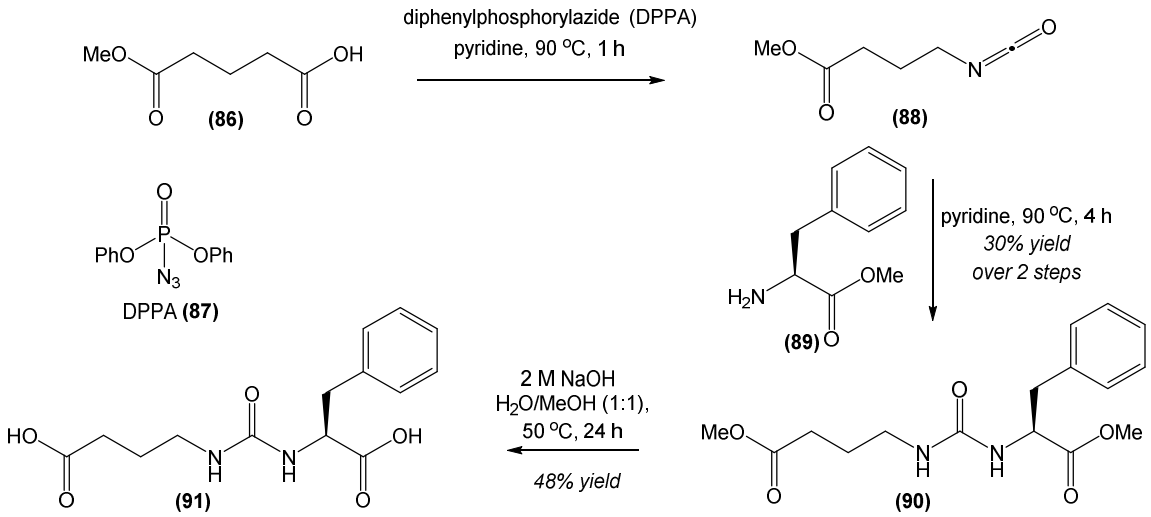
Another interesting feature is the presence of a quaternary carbon C-j at 158 ppm consistent with an ureido carbonyl shift, linking the GABA and phenyalaninyl moieties.[156-158] The conservation of a ureido-linked GABA-like moiety across this series of metabolites **80-85** therefore led to them being named gaburedins.

These initial NMR data could not be used to unequivocally elucidate the structure of gaburedin **80** due to the absence of key correlations. Cross peaks expected in the HMBC spectrum between the proton H-c and ureido carbonyl C-j were missing, therefore the aminoacyl moiety could not be confirmed as being attached to the ureido carbonyl as proposed in Figure 2.6. Some HMBC correlations expected to be seen in the GABA-like moiety between protons H-l and H-m with carbon C-n were also not observed (Figure 2.6 shows all HMBC correlations observed). Furthermore, impurities in the NMR spectrum (signals at 1.24 ppm, 6.45 ppm) and the overlaps of signals for protons d and l at 2.84-3.01 ppm complicated the spectra.

It was therefore decided that an authentic standard corresponding to the proposed structure of gaburedin **80** should be synthesised. Using well documented reactions is also useful as it allows one to reliably predict the structure of the product from the reagents and conditions used.[159] Total synthesis also results in higher purity of the compound of interest, as the resultant mixture contains fewer compounds compared with the complex mixture of different compounds present in microbial culture extracts.

2.2.3 Synthesis of authentic standard of gaburedin A

Synthesis of ureido containing molecules has been widely reported in the literature. Such ureas can be synthesised in a few steps *via* the corresponding isocyanate intermediates.[159,160] The required intermediate isocyanate was synthesised *via* a modified Curtius rearrangement, starting from *mono*-methylglutaric acid **86** and diphenylphosphoryl azide (DPPA) **87** under basic conditions to yield methyl 4-isocyanatobutanoate **88** (Scheme 2.1).[160] The consumption of *mono*-methyl glutarate was monitored by TLC after which L-phenylalanine-*O*-methyl ester hydrochloride **89** was added in order to yield a dimethyl ester protected gaburedin **90** (Scheme 2.1).



Scheme 2.1 – Synthetic route to gaburedin **1** from *mono*-methyl glutaric acid

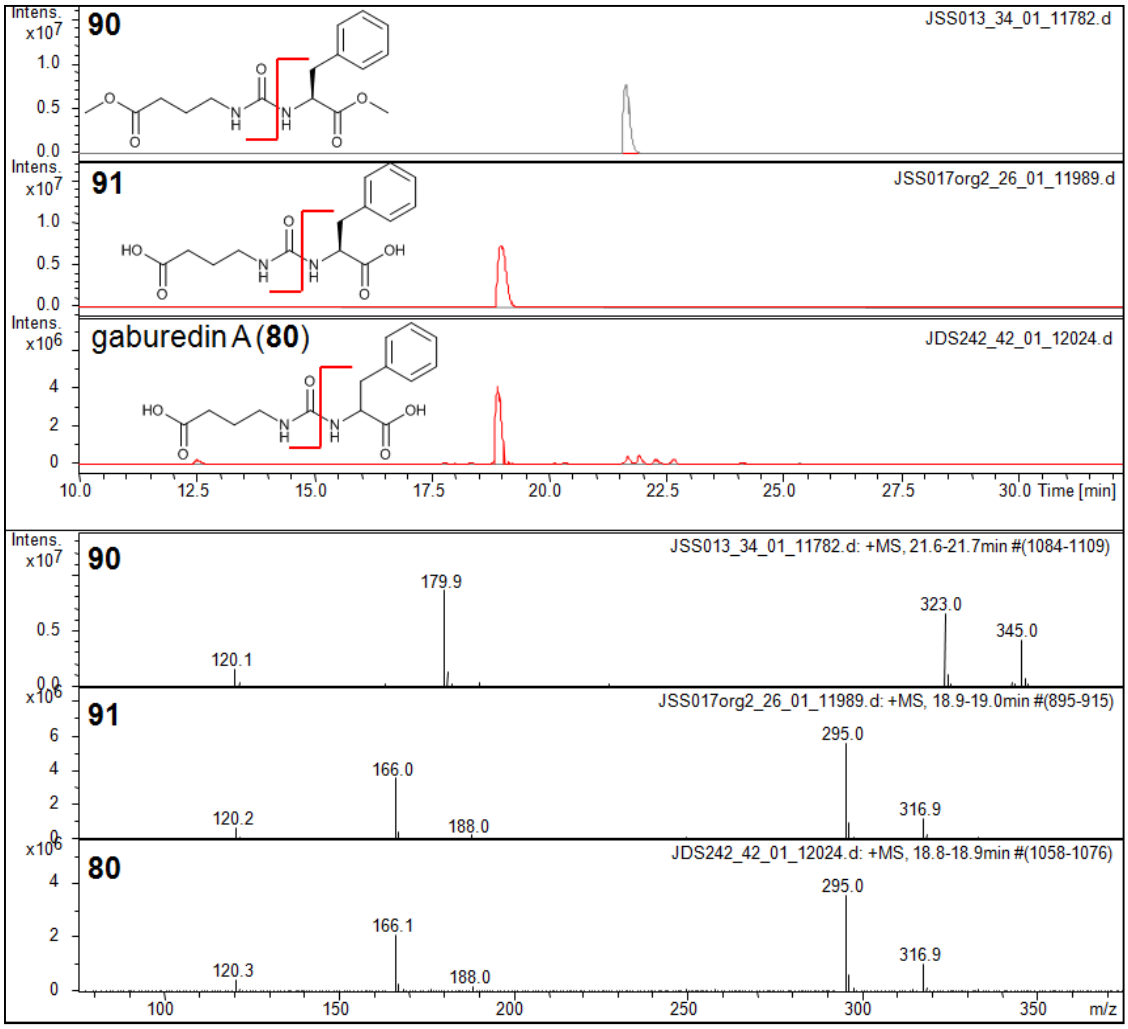


Figure 2.7 – Extracted ion chromatogram for *m/z* = 323.0 for the intermediate **90** and extracted ion chromatograms *m/z* = 294.9, corresponding to the authentic standard **91** and proposed structure of gaburedin A **80** in metabolites extracted from *S. venezuelae gbnR* mutant after 3 days (upper panel). Mass spectra of the **90**, **91** and **80** are also shown (lower panel)

Intermediate **90** was isolated by flash column chromatography and characterised by LC-MS and NMR. LC-MS analysis showed that compound **90** fragmented the same way as the gaburedins (Figure 2.7), with a molecular ion peak *m/z* = 323.0 as expected for a dimethyl ester of gaburedin A, and a fragment ion *m/z* = 179.9 which would be expected for L-phenylalanine-*O*-methyl ester remaining after cleavage of the same ureidyl C-N bond that is observed upon fragmentation of gaburedins A to F.

Double deprotection of the diester was achieved by heating **90** overnight in the presence of sodium hydroxide in order to furnish 12.6 mg final product **91** (Scheme

2.1). Compound **91** was purified by HPLC, solvents removed and the white solid obtained was analysed by LC-MS alongside crude *S. venezuelae gbnR::apra* extract (Figure 2.7).

Extracted ion chromatograms for *m/z* = 294.9 confirm the synthesised compound **91** has the same retention time and mass spectrum fragmentation pattern as gaburedin A **80** isolated from the crude microbial extract, suggesting that the compounds are identical. The absolute stereochemistry at C-c in the natural gaburedin **80** had not yet been confirmed, however, and will be discussed in Section 2.2.5.

Comparison of the ^{1}H NMR spectrum of compound **91** (in d_6-DMSO) shows signals corresponding to the ^{1}H spectrum of gaburedin A **80** purified from the culture extract

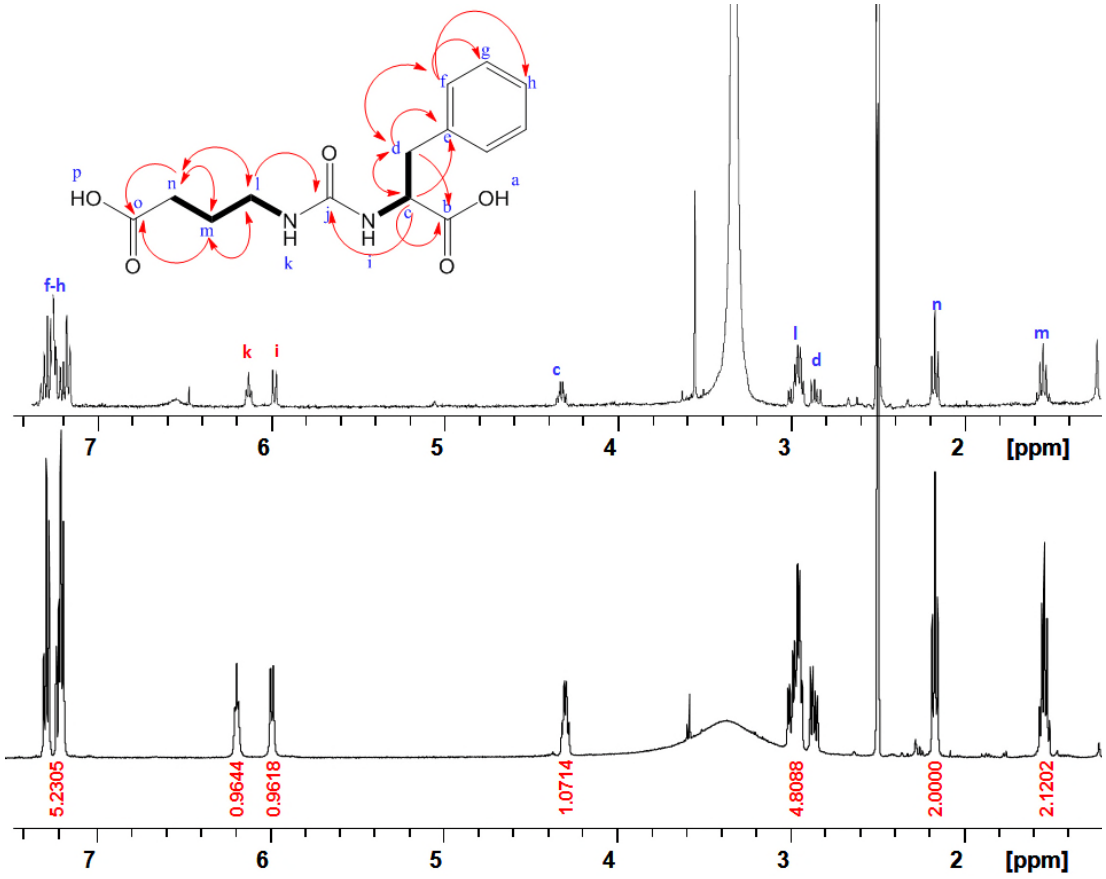


Figure 2.8 – 1H NMR spectra in d_6-DMSO of **80** isolated from *S. venezuelae* grown for 3 days on SMMS (upper spectrum) and authentic standard of **91** (500 MHz, lower spectrum). 2D COSY and HMBC correlations for **91** are also shown

(Figure 2.8). Furthermore, HMBC correlations between the ureido C-j and the protons H-c and H-l confirm that the ureido bridge links the phenylalaninyl and GABA moieties. The presence of HMBC correlations between C-b and H-c and H-d also allowed unequivocal confirmation of the phenylalaninyl portion of the molecule.

As shown in Figure 2.8, the signals corresponding to H-k in the synthetic **91** and bacterial-derived gaburedin **80** samples have slightly different chemical shifts (6.13 for **80** compared to 6.17 for **91**). However this may be attributed to the precise chemical shift of amide protons being variable dependent on the pH of the individual samples, or intramolecular hydrogen bonding due to the difference in concentrations between the two samples.[161]

2.2.4 Feeding experiments to overproduce gaburedins **80-85**

As discussed in Section 2.2.1, it would appear that each gaburedin **80-85** contains a different amino acid moiety. Therefore, it was hypothesised that the relative abundances of gaburedins extracted from *S. venezuelae gbnR::apra* cultures may be altered by growing the bacteria on media enriched with each amino acid proposed to be incorporated into the structures of gaburedin **80-85**.

To test this hypothesis, L-valine was added to SMMS to final concentrations of 1 mM, 5 mM and 20 mM and the pH altered to 7.2 as normal prior to autoclaving. Metabolites were extracted from the acidified *gbnR::apra* culture extract after being grown on these media for 3 days, and analysed by LC-MS. Extracted ion chromatograms for m/z = 247.0 (corresponding to gaburedin C **83**) illustrate a marked increase in the production of **83** as the concentration of L-valine in the media was increased (Figure 2.9).

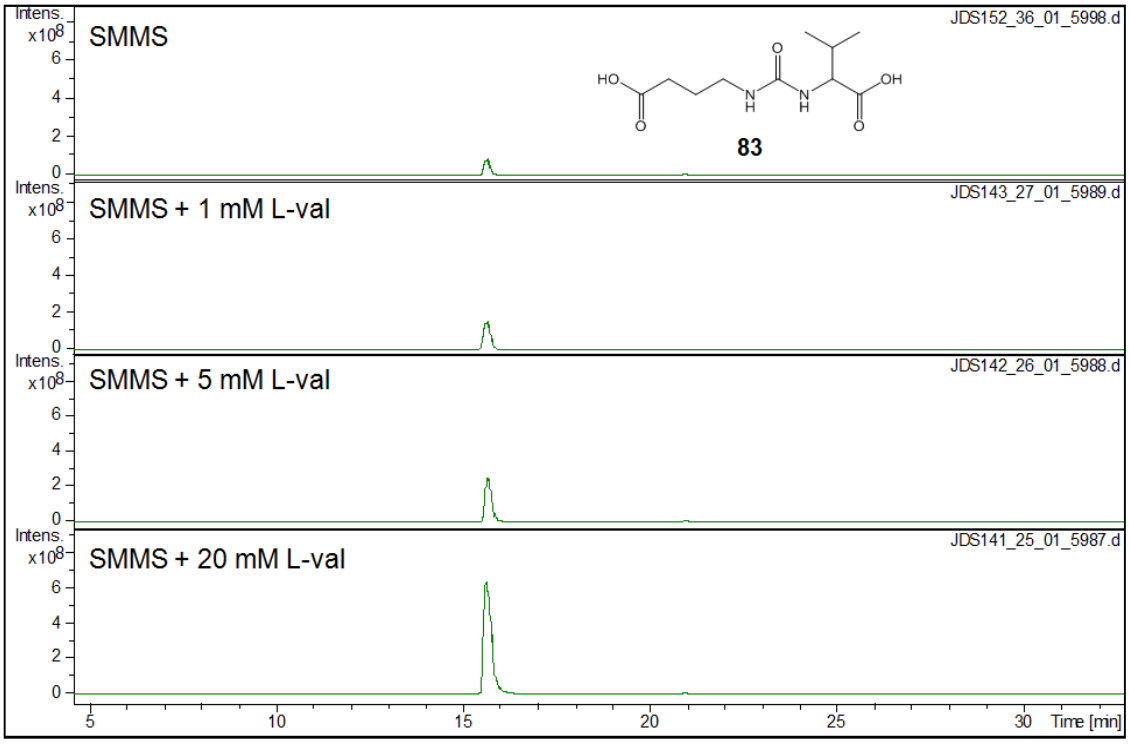


Figure 2.9 – Extracted ion chromatograms for *m/z* = 247.0 corresponding to gaburedin C **83** for metabolites extracted from LC-MS analysis of the metabolites extracted at pH 3 from the *S. venezuelae gbnR::apra* strain when grown on SMMS for 3 days with different concentrations of L-valine added to the media

In order to further verify the incorporation of valine into gaburedin C, the *gbnR::apra* strain was grown on SMMS enriched with 20 mM d_8-DL-valine. If d_8-DL-valine is incorporated into gaburedin C, a mass shift of 8 Da should be observed. Indeed, LC-MS analysis revealed that the *m/z* of the $[M+H]^+$ species of gaburedin C increased by 8 Da from 246.9 to 254.9 upon feeding with d_8-DL-valine, and that the $[M–129+H]^+$ fragment ion *m/z* also increased by 8 from *m/z* = 118.1 to 126.1 (Figure 2.10).

The assignments of these ions were confirmed by UHR-LC-MS analysis, which confirmed the presence of molecular ions $[M+H]^+$ corresponding to $C_{10}H_{11}D_8N_2O_5$ (calculated *m/z* = 255.1791, found 255.1802) and $[M+Na]^+$ corresponding to $C_{10}H_{10}D_8N_2O_5Na$ (calculated *m/z* = 277.1610, found 277.1617). An additional $[M–C_5H_7NO_3+H]^+$ fragment ion corresponding to a molecular formula $C_5H_4D_8NO_2$ was

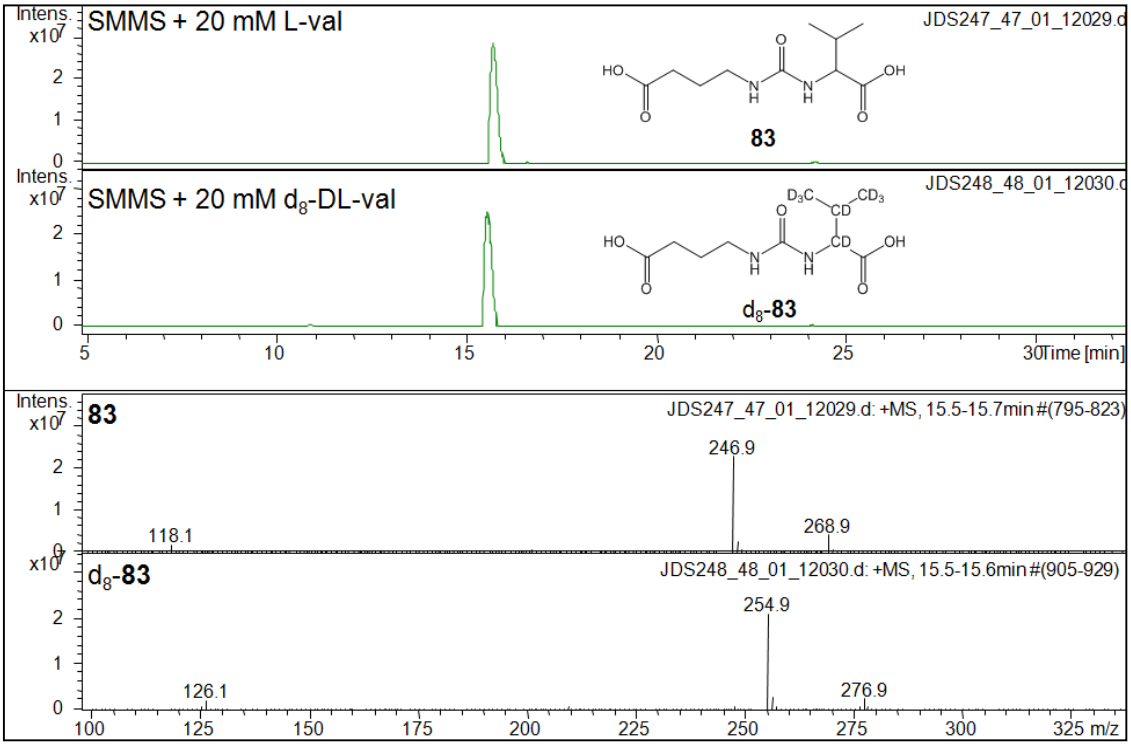


Figure 2.10 – Extracted ion chromatograms for *m/z* = 247.0 for metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS enriched with 20 mM L-valine and 20 mM d_8-DL-valine (upper panels) and mass spectra of the unlabelled gaburedin C (lower panels)

also observed, representing a protonated d_8-valine ion (calculated *m/z* = 126.1365, found 126.1372).

As gaburedins B and C (**81** and **82**) both have the same fragmentation patterns and daughter ions with the same molecular formula of $C_6H_{14}NO_2$, the *gbnR::apra* strain was grown on media enriched with L-leucine and with L-isoleucine (20 mM each). In both cases, there was a dramatic increase in the intensity of the peaks corresponding to *m/z* = 260.9 at retention time = 17.0-17.3 min. The microbial extracts were therefore diluted 10 fold and re-analysed. After dilution, there were two resolvable peaks for *m/z* = 260.9 (Figure 2.11) showing that gaburedin B (t_R = 17.2 min), is overproduced upon addition of L-leucine, and gaburedin C (t_R = 17.0 min), is overproduced upon addition of L-isoleucine into the growth media. This is consistent with previous reports of other natural products only differing in the branching of leucine/isoleucine moieties, for

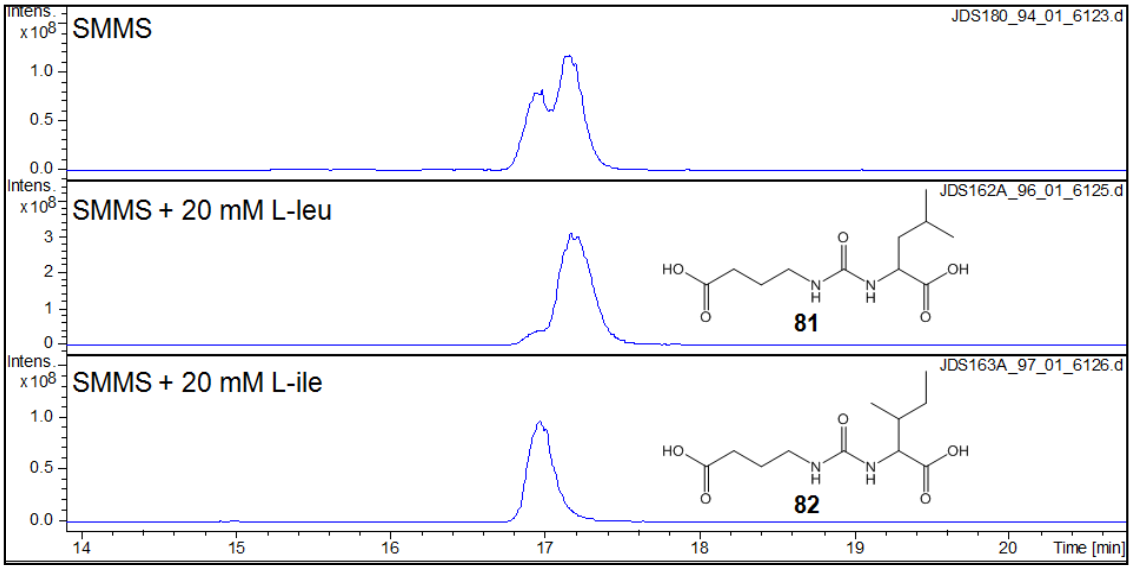


Figure 2.11 – Extracted ion chromatograms for *m/z* = 260.9 for metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS and SMMS enriched with 20 mM L-leucine and 20 mM isoleucine. Culture extracts were diluted by factor of 10 prior to LC-MS analysis

example in the case of germicidins where the isoleucine-derived compound elutes earlier than the leucine-derived compound.[83]

To confirm which of the gaburedins B and C contained leucine or isoleucine, this experiment was repeated with SMMS enriched with 10 mM each of d_{10}-L-leucine or d_{10}-L-isolecucine. It was predicted that the gaburedins B and C extracted from these cultures would each be specifically labelled in one of the samples, but not the other. In order to determine which gaburedin was labelled with which precursor molecule, extracted ion chromatograms (*m/z* = 271.0) were generated from LC-MS analysis of the metabolites extracted from the *gbnR::apra* strain after feeding with 10 mM d_{10}-L-leucine and d_{10}-L-isoleucine.

Figure 2.12 shows the difference in retention times for ions with *m/z* = 271.0, confirming that the slightly earlier eluted compound, gaburedin C (**82**), contains 10 deuterium labels upon feeding with d_{10}-L-isoleucine. Conversely, the compound that

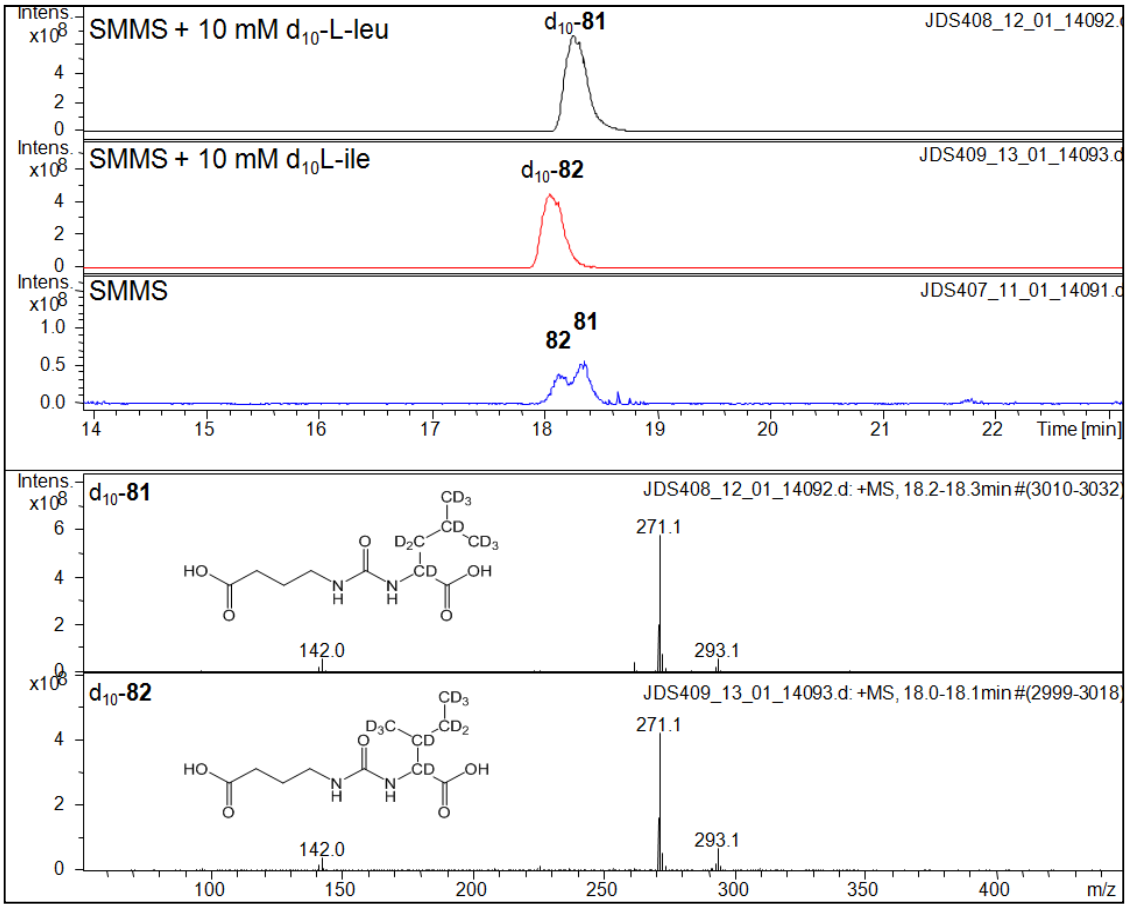


Figure 2.12 – Extracted ion chromatograms for *m/z* = 271.0, corresponding to d_{10}-gaburedin B (**81**) and d_{10}-gaburedin C (**82**) extracted from the *S. venezuelae gbnR::apra* strain after growing on SMMS for 3 days upon addition of 10 mM d_{10}-leucine (black trace) 10 mM d_{10}-isoleucine (red trace). Extracted ion chromatogram for *m/z* = 261.0, corresponding to unlabelled gaburedins B and C extracted from *S. venezuelae gbnR::apra* strain after growing on SMMS for 3 days (blue trace). Mass spectra of the labelled gaburedins B and C are also shown

elutes later, gaburedin B (**81**), contains 10 deuterium labels from feeding with d_{10}-L-leucine. Furthermore, the labelling can be seen in the $[M–129+H]^+$ fragment ions of gaburedins B and C (their *m/z* values increased from 131.9 to 142.0) demonstrating that the deuterium labels are in the amino acid moiety of the molecules. These data confirm the structures of gaburedins B and C contain leucine and isoleucine, respectively.

Similar feeding experiments were performed with SMMS supplemented with 1 mM, 5 mM and 20 mM phenylalanine in order to demonstrate a specific increase in the yield of gaburedin A (**80**). Furthermore, to demonstrate specific incorporation of L-

methionine into gaburedin E (**84**), LC-MS analysis of metabolites extracted upon feeding the *gbnR::apra* strain with 20 mM L-methionine showed there to be overproduction of gaburedin E.

It was unclear from NMR experiments with a crude sample of gaburedin F (**85**) whether or not the structure contained *N*- or *S*-acetyl cysteine. Tandem mass spectrometry had only previously show loss of an acetyl group from the $[M–129+H]^+$ at *m/z* = 163.9 to leave a fragment with *m/z* = 122.1 (Table 2.1). In order to confirm the structure of gaburedin F, feeding experiments were performed in which the *gbnR::apra* mutant was grown on media supplemented with *N*-acetyl cysteine at 5 mM and 10 mM final concentrations.

If gaburedin F contained *S*-acetyl cysteine then it would be expected that upon feeding with *N*-acetyl cysteine that there would be no increase in the amount of gaburedin F produced. Alternatively, there would potentially be the appearance of an additional peak with *m/z* of $[M+H]^+$ = 292.9 with a retention time distinct from that of gaburedin F due to the different regiochemistry of these two cysteine derivatives.

Analysis of metabolites extracted from the media containing different concentrations of *N*-acetyl cysteine (Figure 2.13) showed that increasing the concentration of *N*-acetyl cysteine in the growth medium specifically increased the production of gaburedin F, suggesting that gaburedin F contains *N*-acetyl cysteine (Figure 2.13). Interestingly, the yields of the gaburedins A-E decreased as the concentration of *N*-acetyl cysteine in the media was increased. This may simply be due to a competitive effect when there is more *N*-acetyl cysteine available to react with any intermediates in gaburedin biosynthesis, and will be discussed in more detail in Chapter 3.

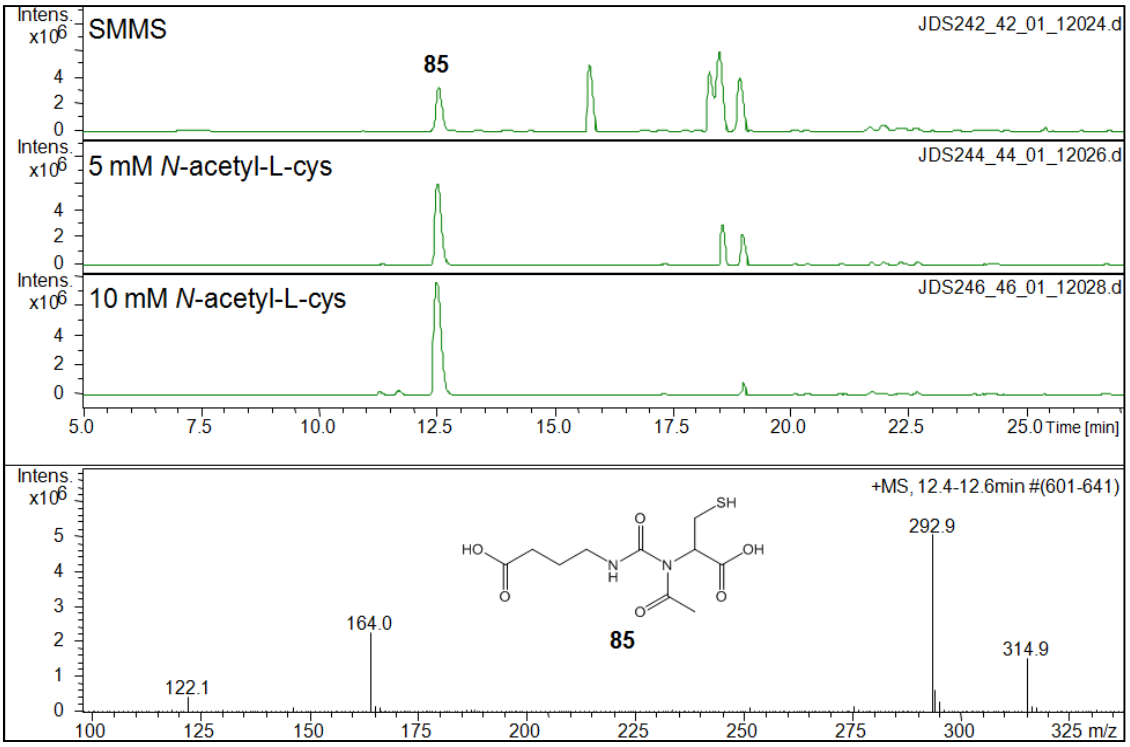


Figure 2.13 – Extracted ion chromatograms for *m/z* = 247.0, 261.0, 279.0, 293.0 and 295.0 for gaburedins A-F in metabolites extracted from *S. venezuelae gbnR::apra* grown on SMMS and SMMS enriched with 5 mM and 10 mM *N*-acetyl-L-cysteine. The mass spectrum and proposed structure for gaburedin F derived from *N*-acetyl-L-cysteine (**85**) are also shown

2.2.5 Chiral HPLC analysis of L- and D-gaburedin A

The feeding experiments in Section 2.2.5 illustrate the incorporation of L-amino acids into gaburedins A-F. To provide further evidence for the absolute stereochemistry of the amino acid moieties found in the natural gaburedins, this series of feeding experiments were extended to include amino acids with D- configuration as well as amino acids of L- configuration.

For example, both 20 mM L- and D-valine were added separately to SMMS, and the *gbnR::apra* strain grown for 3 days, the metabolites extracted and analysed by LC-MS. Interestingly, the addition of both L- and D-valine resulted in overproduction of gaburedins **92** and **93** with identical retention times as gaburedin C (**83**) (Figure 2.14). Incorporation of both L- and D-phenylalanine into gaburedins **94** and **95** was also

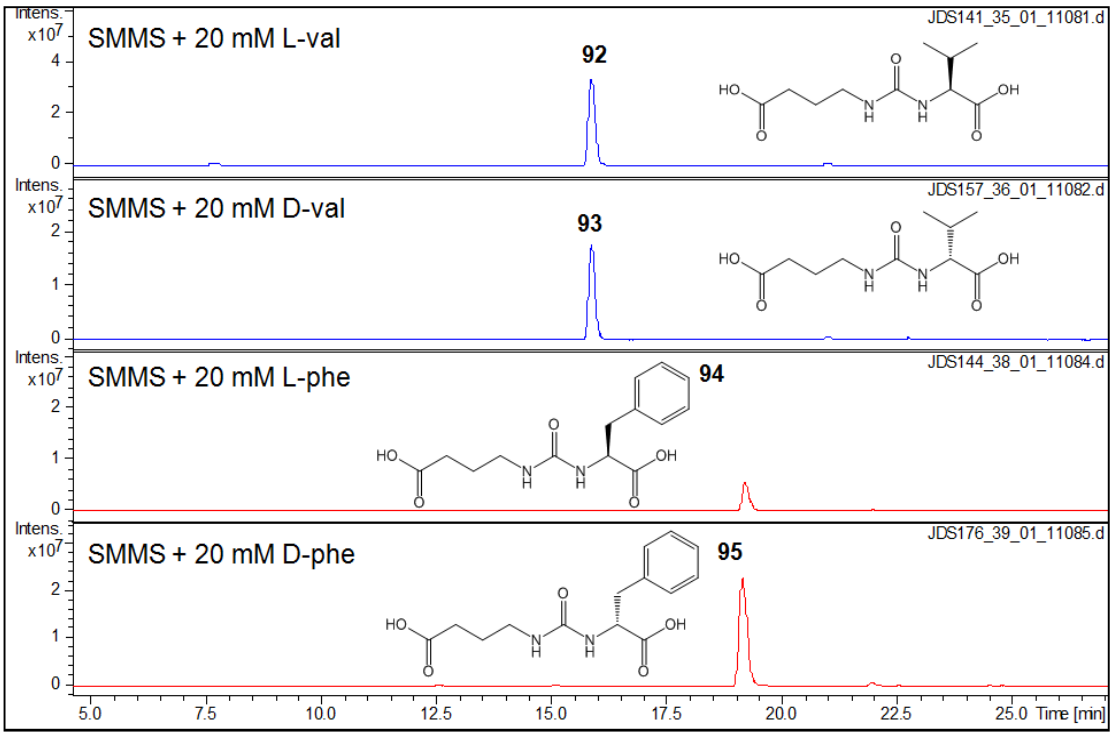


Figure 2.14 – Extracted ion chromatograms for *m/z* = 247.0 corresponding to metabolites **92** and **93** extracted from the *S. venezuelae gbnR* mutant when grown on SMMS enriched with 20 mM L-valine and 20 mM D-valine (blue traces) and extracted ion chromatograms for *m/z* = 295.0 for metabolites **94** and **95** extracted from the *S. venezuelae gbnR* mutant when grown on SMMS enriched with 20 mM L-phenylalanine and 20 mM D-phenylalanine (red traces)

observed (Figure 2.14). The amino acid source in the media (casamino acids) are derived from casein, therefore contain L-amino acids (hence it was proposed that gaburedin D **83** has identical stereochemistry to **92**, and that gaburedin A **80** has identical stereochemistry to **94**.[162] However there may be intracellular epimerisation of the amino acids that are incorporated into the gaburedin natural products observed.

In order to determine if there was epimerisation of one amino acid enantiomer to the other *in vivo*, chiral HPLC analysis of the culture extracts from feeding with both 20 mM L-phenylalanine and D-phenylalanine were performed (L. Song) using a Chiral-PAK IA column with a chiral polysaccharide stationary phase. Both extracts were injected separately, followed by co-injection of both extracts in order to determine whether or not the gaburedins **94** and **95** had different configurations in the phenylalaninyl moiety. Figure 2.15 demonstrates that there are two distinct compounds

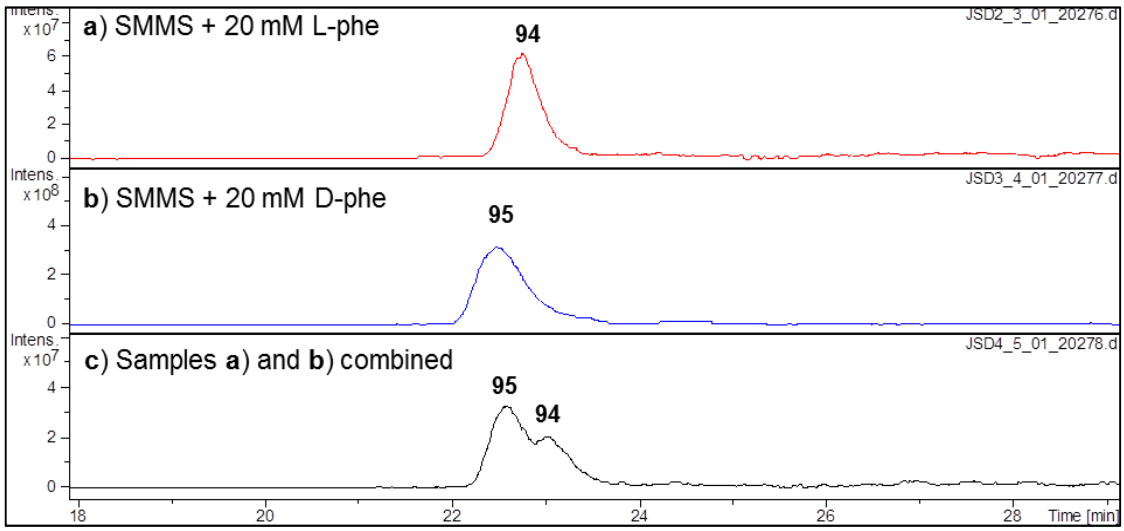


Figure 2.15 – Extracted ion chromatograms from chiral HPLC-MS analysis of metabolites extracted from the *S. venezuelae gbnR::apr* mutant grown on 20 mM L- or D-phenylalanine for *m/z* = 295, corresponding to L- and D- enantiomers of gaburedin A, (panels **a** and **b**, respectively). Coinjection of extracts **a** and **b** (panel **c**) confirmed that the gaburedins **94** and **95** produced upon supplementation with L- or D-phenylalanine are different; as gaburedin **94** (panel **a**) and its enantiomer **95** (panel **b**) have different retention times

with *m/z* = 295.0 in the co-injected sample, confirming the incorporation of both L- and D-phenylalanine into gaburedins **94** and **95** respectively. This result has implications for how gaburedins are biosynthesised, and will be discussed further in Section 3.2.

Furthermore, this also implies that the stereochemistry of the amino acids found in the gaburedins **80-85** isolated from the *gbnR::apra* strain is L-, as the casamino acids present in the SMMS are of L- configuration (Figure 2.16).[162] The result that modification of the media to contain D- as well as L- amino acids can lead to incorporation of either enantiomer into gaburedins implies that there is unprecedented substrate tolerance by one of the enzymes involved (one would expect an enzyme that is specific for an amino acid would only be able to tolerate one enantiomer). Alternatively, it could imply that the incorporation of amino acids into gaburedins is not enzyme catalysed.

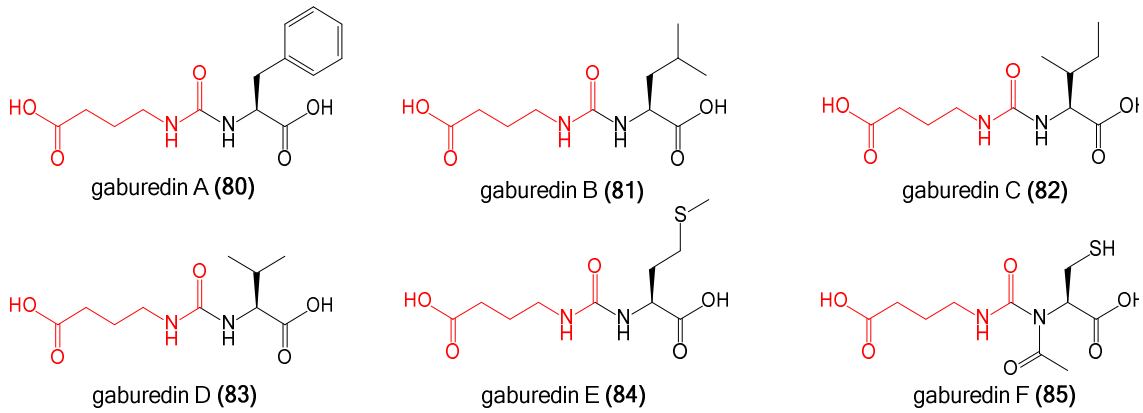


Figure 2.16 – Proposed structures of gaburedins A-F (**80**-**85**)

2.3 Discussion

The gaburedins reported here represent one of the first examples of novel natural products being discovered by deletion of a gene encoding for an ArpA-like transcriptional repressor. In the previously mentioned examples, deletion of *scbR2* in *S. coelicolor* led to overproduction of a yellow pigment, coelimycin (**30**), which had been observed previously.[88,89] In the example of deletion of *ksbC* being deleted in *K. setae*, the motivation behind deletion of *ksbC* was to determine the effect of that repressor on morphological differentiation and bafilomycin production, but had the additional effect of leading to identification of the previously uncharacterised kitasetaline.[87] The strategy employed herein contrasts these other examples, as a targeted gene deletion of *gbnR* was made to specifically de-repress a cryptic natural product gene cluster, *gbnABC*. Production of the gaburedins in the *gbnR* deletion mutant had not been directly attributed to the *gbnABC* cluster, however.

Contrary to the hypothesis that translation of MmfL may only occur once the leucinyl-tRNA responsive to the TTA codon becomes available, it would appear that *sgnL* (which also has a TTA codon) is translated earlier in the *gbnR* mutant, as gaburedins

are observed as being produced in the culture extract of the *gbnR* mutant after just 24 h. Furthermore, the relative yields of gaburedins A-F extracted from the *gbnR* mutant culture did not vary much between 24 and 72 hours, suggesting that there may be an additional mechanism for switching off gaburedin biosynthesis than just the GbnR protein itself.

The ureido bridge found in gaburedins A-F is an interesting functional group as there are several other examples of natural products in which ureido bonds are present. For example, the antibiotic pacidamycins **96** produced by *Streptomyces coeruleorubidus*, the protease inhibitor syringolin **97** isolated from *Pseudomonas syringae*, and the potential HIV-1 protease inhibitor Mer-N5075A **98** which is biosynthesised by *Streptomyces chromofuscus* Mer-N5075 (Figure 2.17).[163-166] Natural products such as the potential anticancer bromophenol **99** and derivatives have also been isolated from the red alga *Rhodomela confervoides*. Grateloupine **100** and lividine **101**, isolated from red algae *Grateloupia filicina* and *Grateloupia livida*, also bear striking resemblance to gaburedins, although no potential biological role for these compounds has been reported.[156,157] Based upon some of these ureido containing compounds as having antibacterial activity, culture extracts containing gaburedins were tested in preliminary antibiotic assays against *E. coli*, *B. subtilis*, and *S. coelicolor* M145, but no antibacterial activity was observed.

Ureido bridges have been used synthetically to produce several interesting biomolecules, such as the glutamate carboxypeptidase inhibitor **102** synthesised by Kozikowski and co-workers, the epoxide hydrolase inhibitor 4-(3-cyclohexylureido)butanoic acid **103** produced by Gomez and co-workers, the series of aminopeptidase N inhibitors **104** produced by Su and co-workers, and the first-

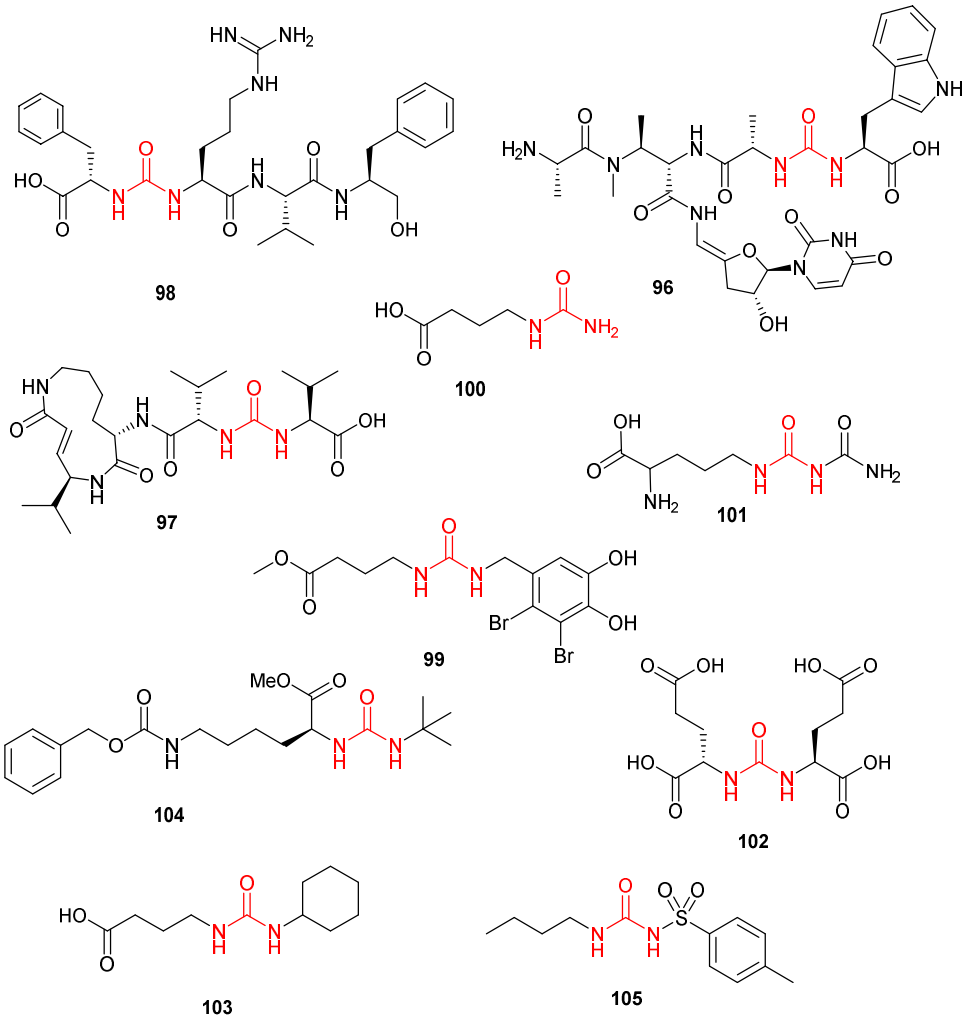


Figure 2.17 – Examples of natural products and synthetic candidate drug molecules containing ureido bridges

generation diabetes drug tolbutamide, **105**.[158,167-169] Ureido bridges are an interesting functional group, as they offer additional stability compared with amide bonds (they will not be hydrolysed by proteases).[159] The enzymes proposed to catalyse the formation of ureido linkages have received little attention, however, with the only mechanistic studies to date involving the enzyme SylC which has been shown to catalyse the formation of the ureido bond in syringolin biosynthesis.[165] Therefore, investigation of the mechanism by which gaburedins are biosynthesised could provide insight into how such functionalities arise in nature and is discussed further in Section 3.3.3.

3. Towards gaburedin biosynthesis

As demonstrated by results in Section 2.2.4, the incorporation of a variety of amino acids the observation that both L- and D-amino acids can be incorporated into gaburedins suggest that a diverse range of substrates can be tolerated by the biosynthetic enzyme(s) involved. It will therefore be interesting to probe the substrate specificity of the enzymes GbnA and GbnB, to determine the range of structures to which a GABA moiety can be introduced.

Primarily, it will be important to establish that the genes *gbnA* and/or *gbnB* confer the gaburedin-producing chemotype (from the results presented thus far, it could be argued that *gbnR* inactivation induces various metabolic changes which indirectly result in gaburedin biosynthesis). In addition, it would be interesting to investigate whether the *gbnABC* cluster is present in any other organisms, or whether it is unique to *Streptomyces venezuelae*. Therefore, Chapter 3 will focus on elucidating a biosynthetic route to gaburedins by investigating the origin of the GABA and ureido groups (blue in Figure 3.1); Chapter 4 will focus on the diversity of amino acids and amino acid analogues that can be incorporated into the aminoacyl moiety (red and green in Figure 3.1).

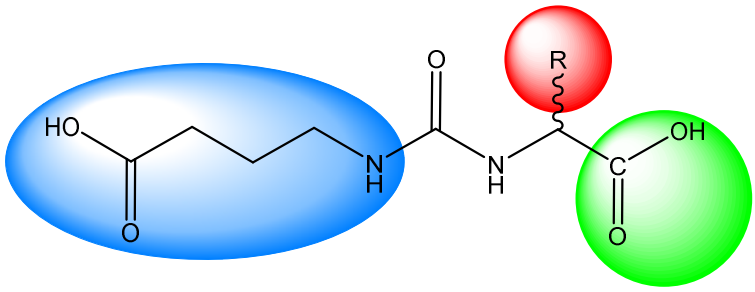


Figure 3.1 – General structures of gaburedins **80-85** characterised in Chapter 2. Chapter 3 will focus on the origin of the GABA-like moiety (blue) and ureido bridge. Chapter 4 will focus on the diversity of substrates with different amino acid side chains (red) and different groups in place of the aminoacyl carboxylic acid group (green) which can be incorporated into gaburedins

3.1 Bioinformatic analyses of GbnA, GbnB, GbnC

3.1.1 BLAST analysis

BLAST analysis of GbnA, GbnB and GbnC individually revealed homologous *gbnABC* gene clusters in other organisms (Figure 3.2 and Table 3.1). The arrangement of the genes in several of these organisms, namely *Streptococcus mutans* GS-5, *Salmonella enterica* subspecies serovar Agona strain SL483, *Pantoea ananatis* LMG 5342, *Azosporillum brasilense* sp245 and *Vibrio vulnificus* MO6_24/O were easily confirmed, as their entire genome sequences are available on NCBI.

Nine other systems were identified in other microorganisms by a BLAST search of the protein sequences of GbnA, GbnB and GbnC from *S. venezuelae* as a single query. This search – followed by manual annotation of the partial genome sequences from which these proteins originated – revealed *gbnABC* gene clusters in three other *Streptomyces* species; *Streptomyces globisporus* subspecies NRRL B-2293, *Streptomyces* Amel2xE9 B065 and *Streptomyces* sp. NRRL B-24482. *gbnABC* systems were also identified in the Gram-positive organisms *Lactobacillus paraplantarum* strain L-ZS9 and *Clostridium papyrosolvens* DSM 2782 as well as in the Gram-negative species *Enterobacter cloacae* (from a multi-species source), *Marinomonas* sp. MED121, *Providencia alcalifaciens* Dmel2 and *Photobacterium halotolerans* DSM 18316. The different *gbnABC* gene clusters identified from these analyses are represented in Figure 3.2 The sequence identity and similarity to GbnA, GbnB and GbnC from *S. venezuelae* are also shown in Table 3.1.

As shown in Figure 3.2, in five of the *gbnABC* systems present in Gram-negative bacteria there appear to be larger proteins (approximately 800 amino acids) comprised

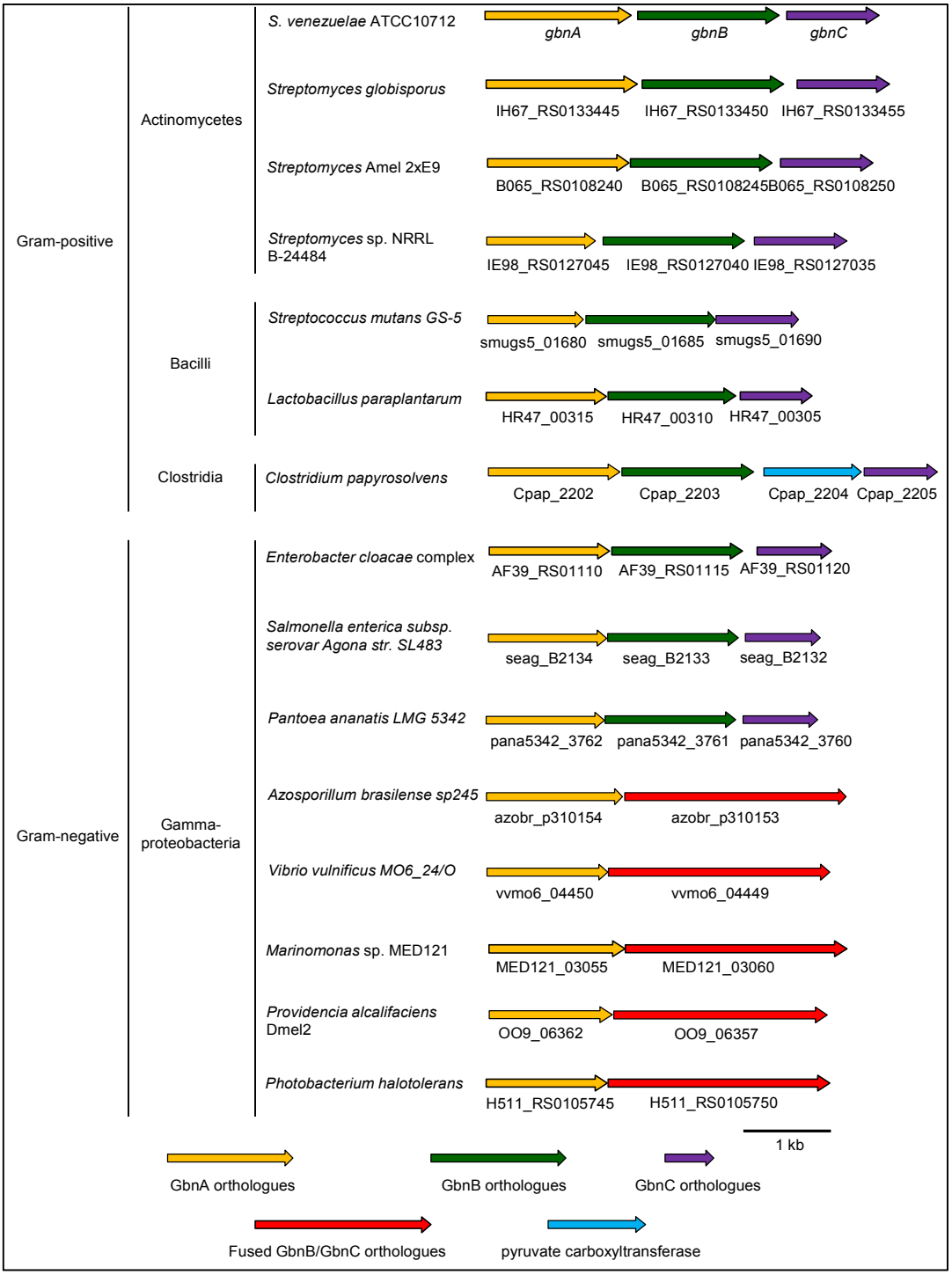


Figure 3.2– *gbnABC* clusters identified by BLAST searches. More detailed analysis of these gene clusters is shown in Appendix 2

of domains which are analogous to both GbnB and GbnC from *S. venezuelae*. As GbnC are membrane-associated exporters of small amino acid-like molecules (see Section 3.2.1 below), this observation could also imply that GbnB and GbnC function together as a complex for gaburedin biosynthesis/export in the organisms in which *gbnB* and *gbnC* are encoded by distinct genes. Furthermore, the cluster in *Clostridium*

Table 3.1 – Sequence homology of *S. venezuelae* GbnA, GbnB and GbnC and orthologous proteins present in other systems as revealed by BLAST searches. Numbers denoted in each column indicate number of amino acids, % identity, % similarity compared with *S. venezuelae* GbnA, GbnB and GbnC

Organism	*GbnA homologue*			*GbnB homologue*			*GbnC homologue*		
	N	%I	%S	N	%I	%S	N	%I	%S
Other *Streptomyces* species									
Streptomyces venezuelae	558	100	100	532	100	100	354	100	100
Streptomyces globisporus	579	93	95	532	96	98	345	96	98
Streptomyces amel 2xE9	530	83	87	528	79	88	345	83	91
Streptomyces NRRL B-24484	465	80	87	537	64	78	338	77	86
Gram positive species									
Streptococcus mutans GS-5	324	27	48	487	27	48	311	41	60
Lactobacillus paraplantarum	460	33	54	492	27	45	308	41	62
Clostridium papyrosolvens DSM 2782	500	33	55	487	30	49	319	40	62
Gram negative species									
Enterobacter cloacae complex	461	35	54	498	32	48	321	55	69
Salmonella enterica serovar Agona SL483	460	35	55	498	31	49	320	56	71
Pantoea ananatis LMG 5342	461	37	55	498	31	48	321	55	70
Gram negative species with fused *gbnBC*									
Azosporillum brasilense sp245	494	48	61	846	37	52	846	36	52
Vibrio vulnificus MO6_24/O	465	28	49	834	31	51	834	37	54
Marinomonas sp. MED121	491	29	52	836	31	52	836	36	57
Providencia alcalifaciens Dmel2	451	26	90	812	29	50	812	37	56
Photobacterium halotolerans	467	25	51	833	31	52	833	35	56

papyrosolvens has an additional gene encoding for a pyruvate carboxyltransferase situated between the *gbnB* and *gbnC* orthologues.

3.1.2 *gbnA* encodes for a PLP-binding putative decarboxylase

The NCBI database was used to analyse the protein sequences of GbnA, GbnB and GbnC. Analysis of GbnA (encoded for by *sven_4179* gene) revealed it to belong to the pyridoxyl 5'-phosphate (PLP)-dependent class of transaminase/decarboxylase enzymes.[170] This analysis highlighted a conserved domain (residues 26-420, pfam01276) including a PLP-binding pocket composed of eight conserved residues T116, S117, N120, S197, D227, W230, S268 and the catalytic lysine residue, K271.[171,172] Alignment of GbnA and its orthologues revealed that the eight amino acid

A

Organism	116	117	120	197	227	230	268	271
Streptomyces venezuelae	T	S	N	S	D	W	S	K
Streptomyces globisporus	T	S	N	S	D	W	S	K
Streptomyces amel 2xE9	T	S	N	S	D	W	S	K
Streptomyces NRRL B-24484	T	S	N	S	D	W	S	K
Streptococcus mutans GS-5	-	-	-	N	D	W	S	K
Lactobacillus paraplantarum	T	S	N	N	D	W	S	K
Clostridium papyrosolvens DSM 2782	T	T	N	N	D	W	S	K
Enterobacter cloacae complex	T	S	N	T	D	W	S	K
Salmonella enterica serovar Agona SL483	T	S	N	T	D	W	S	K
Pantoea ananatis LMG 5342	T	S	N	T	D	W	S	K
Azosporillum brasilense sp245	T	T	N	N	D	W	S	K
Vibrio vulnificus MO6_24/O	S	T	N	N	D	W	S	K
Marinomonas sp. MED121	S	S	N	N	D	W	S	K
Providencia alcalifaciens Dmel2	S	T	N	N	D	W	S	K
Photobacterium halotolerans	S	T	N	N	D	W	S	K

B

```
Streptomyces venezuelae SVEN_4179              IIAYPPGIPITVPGEKLCERTLSQIDTLR
Streptomyces globisporus IH67_RS0133445        IIAYPPGIPITVPGEKLCERTLSQIDTLR
Streptomyces Amel2Xe9 B065_RS0108240           IIAYPPGIPIAVPGERVCDRTLGEIDALR
Streptomyces NRRL_B24484 IE98_RS0127045        IIAYPPGIPITVPGEKLCDATRAEIRALR
Streptococcus mutans GS-5 SMUGS5_01680         IIPYPPGVPIAVPGEVMTTGLKQRIEEYR
Lactobacillus paraplatarum HR47_00315          IIPYPPGVPLIVPGEKFNSSIIAQLSVYE
Clostridium papyrosolvens CPAP_2202            VIPYPPGVPLLFKGDIINRKEAEKIKILK
Enterobacter cloacae AF39_RS01110              IIPYPPGVPLVFPGEIIDDGIRNKIDEYR
Salmonella enterica SL483 SEAG_B2134           IISYPPGVPLVFPGDVISKDVRNKINECK
Pantoea ananatis LMG 5342_3762                 IIPYPPGVPLVFPGEVIDEVIKSKIEKYR
Azosporillum brasilense sp245 AZOBR_p310354    IIPYPPGVPIIVPGEEIGGDTLRRIEAIR
Vibrio vulnificus MO6_24/O VVMO6_04450         YILYPPGIPILTPGQTICNNVITKINQSI
Marinomonas MED121_03055                       FILYPPGIPVLKPGQRVCENALSKIN
Providencia alcalifaciens Dmel2 OO9/006362     VIPYPPGVPIVFPGQNISSLDLEKINKHK
Photobacterium halotolerans H511_RS0105745     FILYPPGTPVLTPGQTVCKFDLDKINNSL
```

Figure 3.3 – A) Conservation of amino acid residues involved in PLP-binding domains in GbnA and orthologues B) C-terminal motifs of GbnA and orthologues, illustrating the IXYPPGXP consensus

residues associated with PLP-binding are well conserved across all 15 GbnA proteins (Figure 3.3). The N-terminus of the *Streptococcus mutans* orthologue SMUGS5_01680 is truncated, however, and it is missing several PLP-interacting residues.

The observation that all the natural products **80**-**85** contain a conserved GABA-like moiety (blue in Figure 3.1) suggests that the decarboxylase GbnA might catalyse glutamate decarboxylation rather than Orn/Lys/Arg decarboxylation. Decarboxylation of glutamate would generate GABA, which could be then used by GbnB as a precursor to gaburedins. However, the C-terminal domain of GbnA was identified as belonging to

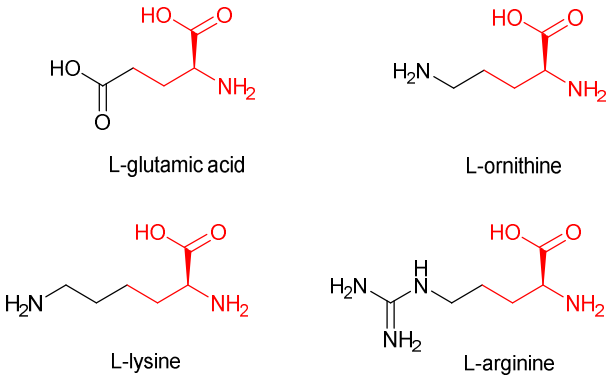


Figure 3.4 – Structures of L-glutamic acid, L-ornithine, L-lysine and L-arginine. Conserved structural motif is highlight in red

pfam03711 (ornithine/lysine/arginine substrate binding motif, amino acid residues 452-481), and this region was found to contain a highly conserved IXYPPGXP motif that is also present in the 1ORDA crystal structure for ornithine decarboxylase from *Lactobacillus* 30a.[171] Ornithine, lysine, arginine and glutamic acid are structurally related and all share the substructure represented in Figure 3.4, which could explain the observation that the C-terminal motif of GbnA and its orthologues share this common OKR-binding motif.

Other, more specific, residues in the ligand binding pocket would vary to accommodate the amine groups of ornithine, lysine or arginine, or the carboxylate group of glutamate, therefore conferring specificity for one substrate over another. However, these residues could not be identified.

3.1.3 *gbnB* encodes for a putative AMP-ligase

GbnB (encoded for by *sven_4180* gene) is labelled as containing an adenylate forming domain (positions 169-505, CD04433) and this domain contains thirteen out of fourteen of the conserved residues involved in adenylate-monophosphate binding at positions T170, G285, G286, L307, Y308, Y309, G310, T311, T312, D394, L406,

Name	Start	*p*-value	Sites
Marinomonas	169	3.80e-33	IQHRAFKSIS FTSGTSGIPKSVVRNKSFDARRFAFFTARY GFSSDDQHLL
Photobacterium	165	1.04e-31	LQHRPFKSIS FTSGTSGIPKPVIRNKSFDGRRFSFFTARY GFSSEDRHLL
Sglobisporus	169	1.04e-31	PVEQPFEAFS FTSGTSGVPKMVVRRTSFEARRFADLVDQY AFDEEDVHLV
Sven_4180	169	1.04e-31	PVEQPFEAFS FTSGTSGVPKMVVRRTSFEARRFADLVDQY AFDEEDVHLV
Azbor_p310153	164	1.71e-31	APARPFRAIS FTSGTSGTPKPVIRTASFDARRFAYFTARY GFSSADRHLA
S_Amel2Xe9	165	1.71e-31	PVEQPFEAFS FTSGTSGIPKLVIRRSSFEARRFADLVDQY SFDEDDVHLV
Vvmo6_04449	165	1.93e-31	AQHRPFRSIS FTSGTSGTPKTVVRSKSFDARRFSFFTARY GFSSEDVHLL
S_NRRL_B24484	172	3.84e-29	PVEQPFTAFS FTSGTSGIPKLVIRNSSFEARRLSDLVDQY SFDEDDVHMV
Clostrid_papyrosolvens	159	5.15e-29	GISRPFESIS FTSGTSGIPKVVYRDKSFDGRRFPYLIDRF GFTELDLFLV
PANA5342_3761	164	3.41e-28	PNASSFRAVA LTSGTSALPKIALRYRSFDARRFAWFTDRF GFSDSDGFML
Providencia	164	1.01e-27	DYSRPFNAIS FTSGTSGVPKAAIRSSSFDLRRFSYFTLKY GFNSKDRHLL
SeAg_B2133	164	1.58e-27	HIPAPFRSVS LTSGTSSAPKIVLRYNSFDARRFDWFTQRF NFTHHDGFLL
Enterobacter	164	2.25e-27	YTQPPYRGLS LTSGTSSVPKVALRYRSFDTRRFAWFTRQF GFSDADGFML
Lactobacillus_paraplantarum	161	1.68e-25	NHNDDILSFS FTSGTTGLPKCIYRKSSFDKRRLGIFTNKY EITSDDIYLV
Smugs5_01685	156	3.38e-22	KHSDSIVSFG FTSGTTGLPKCIYRDYSFATERMKELTKLY NFNATDVFLV

Figure 3.5 –Sequence alignment of GbnB and orthologues, illustrating the acyl-activating motif TSGTSXXPK identified during conserved domain (CD) searches

R409 and K502.[77,170] A conserved motif found in acyl-activating enzymes can also be mapped onto the GbnB sequence at the nine positions L150, T170, S171, G172, T173, S174, G175, P177 and K178 (Figure 3.5).[173,174] Furthermore, ten conserved residues involved in coenzyme A substrate binding can also be mapped to the GbnB sequence at positions V210, P259, V260, R262, G285, A418, G419, K420, A480 and L486. These are distinct from the acyl-activating consensus, therefore suggesting that the specific substrate for GbnB may only have some structural features shared with coenzyme A.[173]

GbnB and its orthologues were aligned using Clustal Omega, revealing highly conserved residues at the positions identified as being associated with AMP-binding. In addition, all fifteen proteins contain the conserved TSGTSXXPK motif involved acyl-activation (Figure 3.5). These observations together imply that the substrate and/or product of GbnB are structurally similar to AMP or a derivative of AMP.

The specific residues mentioned above are all part of large domains that belong to classes of adenylate-forming enzymes that are shared between the A-domain of non-ribosomal peptide synthetases (NRPSs) as well as enzymes involved in Coenzyme A

binding. Specific residues relating to a particular substrate were not identified, however.

3.1.4 *gbnC* encodes for a transport protein

GbnC (encoded for by the *sven_4181* gene), belongs to the EamA-like family of membrane transport proteins (pfam00892) characterised by two 129 amino acid repeated regions (residues 19-147 and 174-302) which are similar in sequence to other members of this family, that includes amino acid, peptide and amine transporters.[77,175] This analysis implies that GbnC does not have a catalytic role in gaburedin biosynthesis, but would be responsible for the export of gaburedins from the cells to the extracellular milieu.

3.2 GbnB is essential for gaburedin biosynthesis

3.2.1 Characterisation of the *gbnB/gbnR* double mutant strain

After characterisation of the gaburedin natural products overproduced by the *gbnR* mutant strain, it was important to establish if the *gbnABC* gene cluster was directly responsible for gaburedin assembly and export. Deletion of *gbnR* was predicted to directly control the expression of the *gbnABC* operon (Section 1.7). Strains lacking the *gbnR* gene (hence able to transcribe *gbnABC* and hence are able to produce gaburedins) would lose the gaburedin-producing phenotype if one of the essential biosynthetic genes in the *gbnABC* cluster was also deleted.

The *gbnB/gbnR* strain was constructed by replacing (and therefore disrupting) the *gbnB* gene with an apramycin resistance gene in the SV4 F01 cosmid carrying the complete *gbn-sgn* gene clusters. The *gbnR* gene was also removed, and replaced by an 81 bp

Template	B,R	B,R	WT	R	WT
Screening primers used	H/H'	F/F'	H/H'	F/F'	F/F'
Predicted size / bp	1884	323	2112	323	863

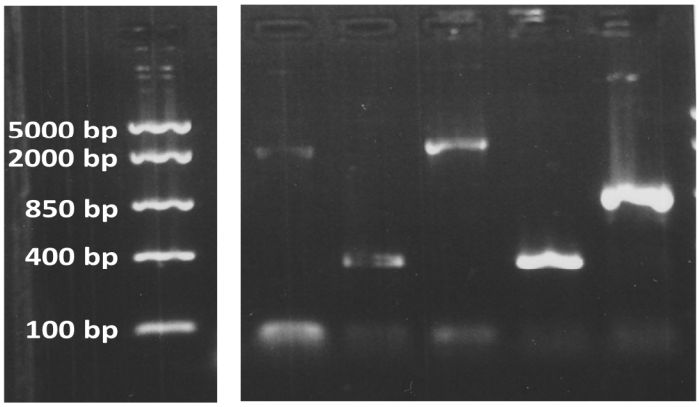


Figure 3.6 **–** Agarose gel of PCR products used to verify the correct construction of the *S. venezuelae gbnB/gbnR* mutant strain. The templates used were genomic *gbnB::apra, gbnR::scar* mutant (B, R) and cosmids SV4 F01 containing the wild-type *gbn-sgn* gene clusters (WT) and SV4 F01 *gbnR::scar* (R). Different screening primers were used to amplify the genes shown

'scar' mutation. This cosmid was then introduced into the *S. venezuelae* wild-type strain and apramycin resistant clones were isolated (O. Lazos, G. L. Challis group).

To identify *S. venezuelae gbnB::apra, gbnR::scar* strains with the correct genotype, genomic DNA was extracted and polymerase chain reactions (PCRs) were performed with specific primers F/F' and H/H' to screen for the disruption of *gbnR* and of *gbnB*, respectively (primer sequences can be found in Section 8.5). The SV4 F01 and SV4 F01 *gbnR::scar* cosmids were used as control templates with each set of primers. The PCR products were analysed by agarose gel electrophoresis (Figure 3.6).

The PCR products obtained using the *gbnR* screening primer pair F/F' were the sizes predicted (323 bp for the *gbnR::scar* mutation in the *gbnB/gbnR* double mutant genomic DNA and SV4 F01 *gbnR::scar* cosmid control, and 863 bp for the undisrupted *gbnR* gene from the SV4 F01 cosmid DNA template). Figure 3.6 illustrates bands at just above the 2000 bp marker for the PCR product using primer pair H/H' to amplify

the *gbnB* gene from the SV4 F01 cosmid (expected 2122 bp), and a band at just below the 2000 bp marker for the *gbnB::apra* cassette amplified from the *gbnB/gbnR* double mutant genomic DNA (1884 bp predicted), therefore confirming the genotype of this new strain.

3.2.2 Chemotyping of the *gbnB/gbnR* double mutant strain by LC-MS analyses

After confirming that the *gbnB/gbnR* strain had the desired genetic modifications, spore stocks were prepared and this strain was grown for 3 days on SMMS. The metabolites produced were extracted at pH 3, and compared with metabolites produced by the wild-type and the *gbnR::apra* single deletion mutant strains by LC-MS in order to determine the effect of *gbnB* gene deletion on gaburedin production (Figure 3.7). As illustrated by the extracted ion chromatograms for the gaburedins **80-85**, in the *gbnB::apra,gbnR::scar* mutant strain, gaburedin production was completely abolished. This result confirmed that *gbnB* is essential for gaburedin biosynthesis. The potential role of GbnB in gaburedin biosynthesis is discussed further in Section 3.4.5.

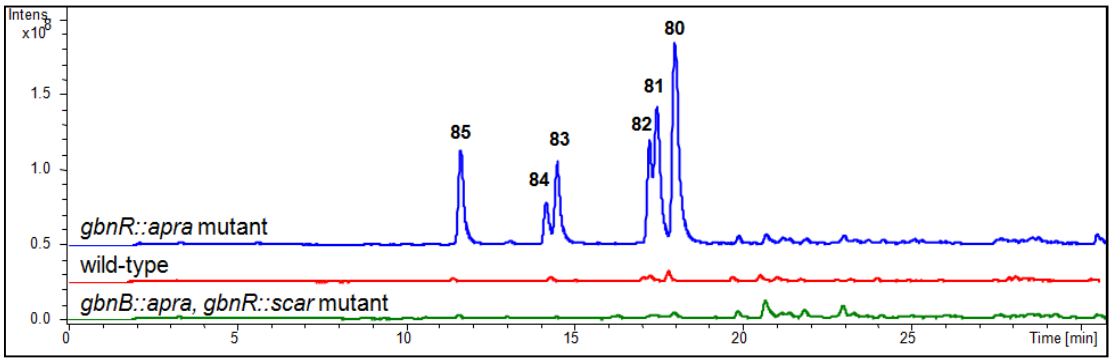


Figure 3.7 – Extracted ion chromatograms *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 of gaburedins A-F (**80-85**) extracted at pH 3 from the *Streptomyces venezuelae* wild-type (red trace), *gbnR* mutant (blue trace) and *gbnB/gbnR* mutant (green trace) after being grown on SMMS for 3 days

3.3 Proposed biosynthetic route to gaburedins

3.3.1 Feeding experiments with putative precursor molecules

As mentioned in Section 3.1.2, the observation that all the gaburedins produced by the *gbnR::apr* mutant contain a conserved GABA moiety implies that the substrate of GbnA could be L-glutamate, rather than L-lysine, L-ornithine or L-arginine. To test this hypothesis, the *gbnR* mutant was grown on SMMS supplemented with 10 mM L-glutamic acid. Extracellular metabolites produced after 3 days were then extracted and submitted to LC-MS analyses. It was expected that by enriching the culture media with L-glutamic acid that the yield of gaburedins would be increased. Instead, two additional new gaburedin-like molecules **106** and **107** were observed (Figure 3.8).

It was observed that there was a new gaburedin **106** with *m/z* = 277.0, as would be expected for a gaburedin containing L-glutamic acid (Figure 3.8). Molecular formulae of the ions of $C_{10}H_{17}N_2O_7$ for the $[M+H]^+$ ion (calculated *m/z* = 277.1030, found 277.1030), $C_{10}H_{16}N_2O_7Na$ for the $[M+Na]^+$ ion (calculated *m/z* = 299.0850, found 299.0858), and $C_5H_{10}NO_4$ for the $[M+H\text{-}C_5H_7NO_3]^+$ fragment ion (calculated *m/z* = 148.0604, found 148.0602) were confirmed by UHR-LC-MS. Interestingly, an additional gaburedin **107** with *m/z* = 232.9 for $[M+H]^+$ 254.9 and $[M+Na]^+$ and a small fragment ion $[M+H\text{-}C_5H_7NO_3]^+$ *m/z* = 104.0 were also observed in this sample, suggesting that a new gaburedin containing GABA was also formed (Figure 3.8 and Scheme 3.1).

This result is consistent with the hypothesis that GbnA acts as a glutamate decarboxylase, as providing the bacteria with a greater concentration of glutamate would lead to an increase in the amount of available substrate for GbnA to turn over

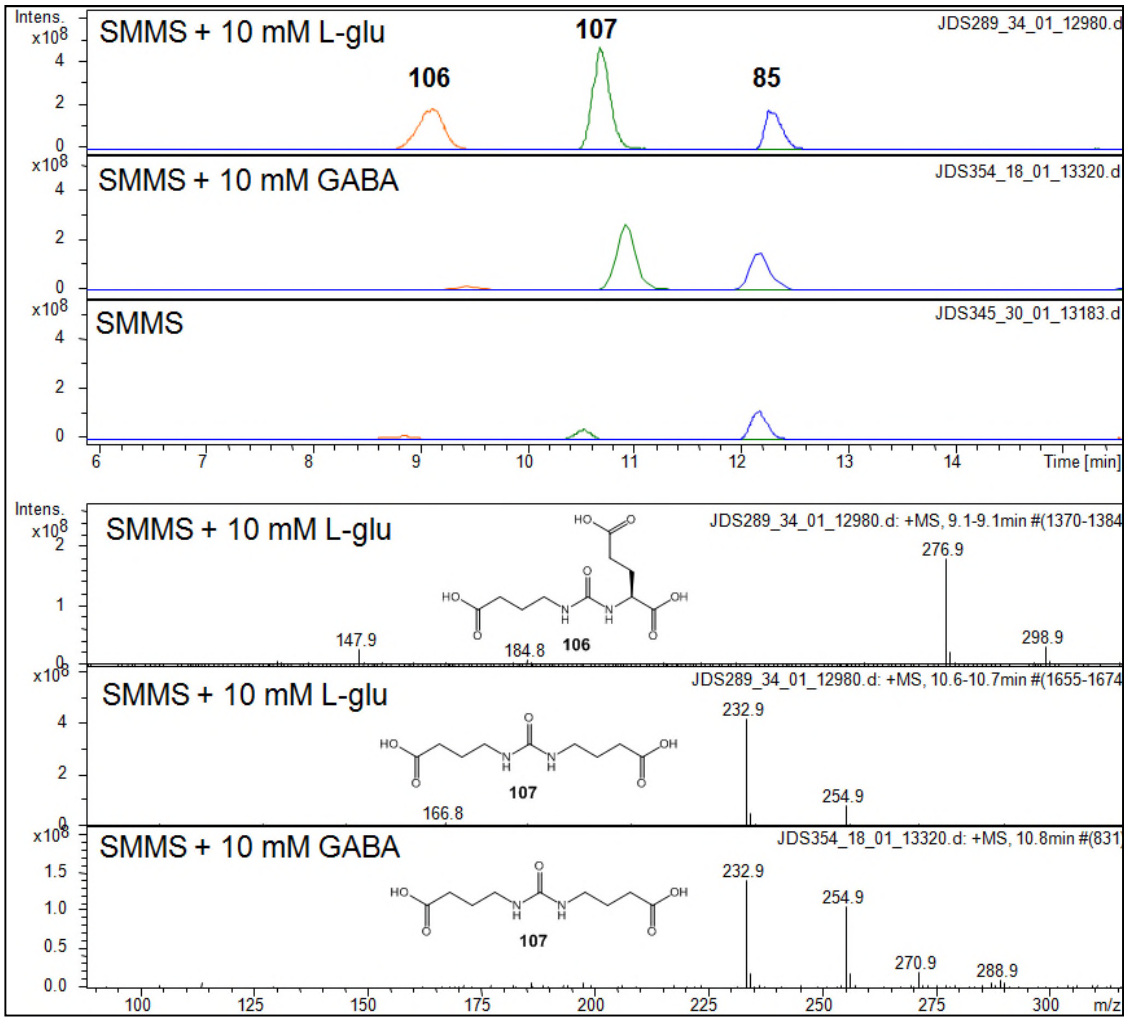
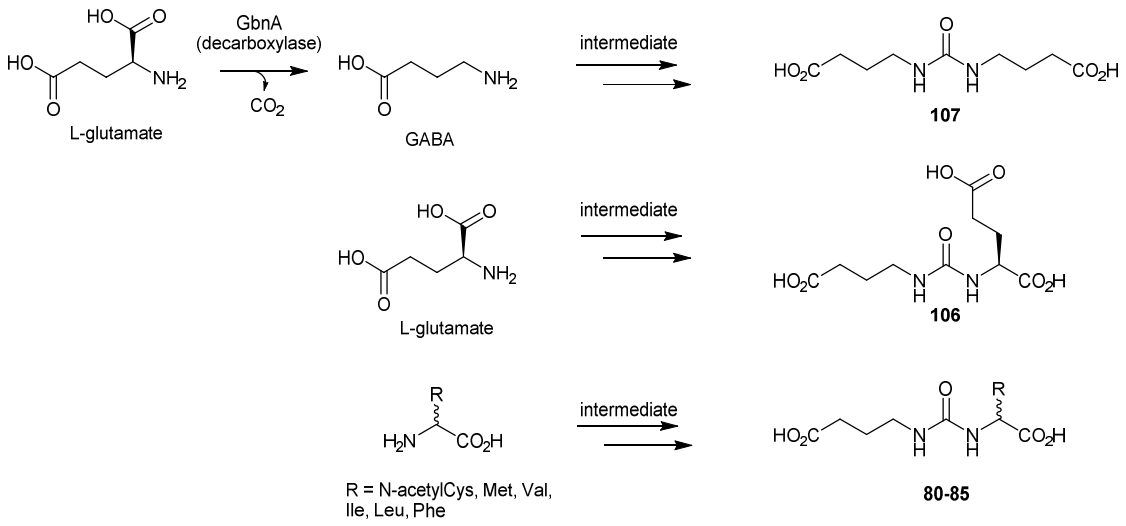


Figure 3.8 – Extracted ion chromatograms for *m/z* = 292.9 (blue) for gaburedin F (**85**) , *m/z* = 277.0 (green) for new gaburedin **106** and *m/z* = 233.0 (orange) for new gaburedin **107** for metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS enriched with 10 mM L-glutamate, 10 mM GABA and unmodified SMMS. Mass spectra of the new gaburedin **106** containing L-glutamate and gaburedin **107** containing GABA are shown

into GABA, hence there would be a greater intracellular concentration of GABA. However, the observation of a new gaburedin **107** (Figure 3.8, Scheme 3.1) resulting from from GABA being incorporated into gaburedins instead of an aminoacyl moiety, was unexpected.

In order to determine whether this additional gaburedin **107** with *m/z* = 232.9 for the $[M+H]^+$ ion was indeed a GABA-derivative of gaburedins, SMMS containing 10 mM (GABA) was prepared, the *gbnR::apra* strain grown and metabolites extracted and



Scheme 3.1 – Schematic to show incorporation of L-glutamate and GABA into gaburedins. Decarboxylation of L-glutamate by GbnA produces GABA, which is incorporated into gaburedins

analysed by LC-MS. A new compound was observed in the LC-MS trace with *m/z* = 232.9 and was shown to have the same retention time as the compound with *m/z* = 232.9 observed in the metabolites of the *gbnR::apra* strain when grown on media supplemented with increasing concentrations of L-glutamate (Figure 3.8).

This sample was analysed by UHR-LC-MS which confirmed the molecular formulae of **107** to be $C_9H_{17}N_2O_5$ for the $[M+H]^+$ ion (calculated *m/z* = 233.1132, found 233.1134), $C_9H_{16}N_2O_5Na$ for the $[M+Na]^+$ ion (calculated *m/z* = 255.0951, found 255.0949), and small peak corresponding to $C_4H_{10}NO_2$ for the $[M+H\text{-}C_5H_7NO_3]^+$ fragment ion corresponding to protonated GABA (calculated *m/z* = 104.0706, found 104.0714). Therefore, it would appear that both L-glutamate and GABA can replace phenylalanine, leucine, isoleucine, valine, methionine or acetylcysteine as the aminoacyl moiety of gaburedins (Scheme 3.1).

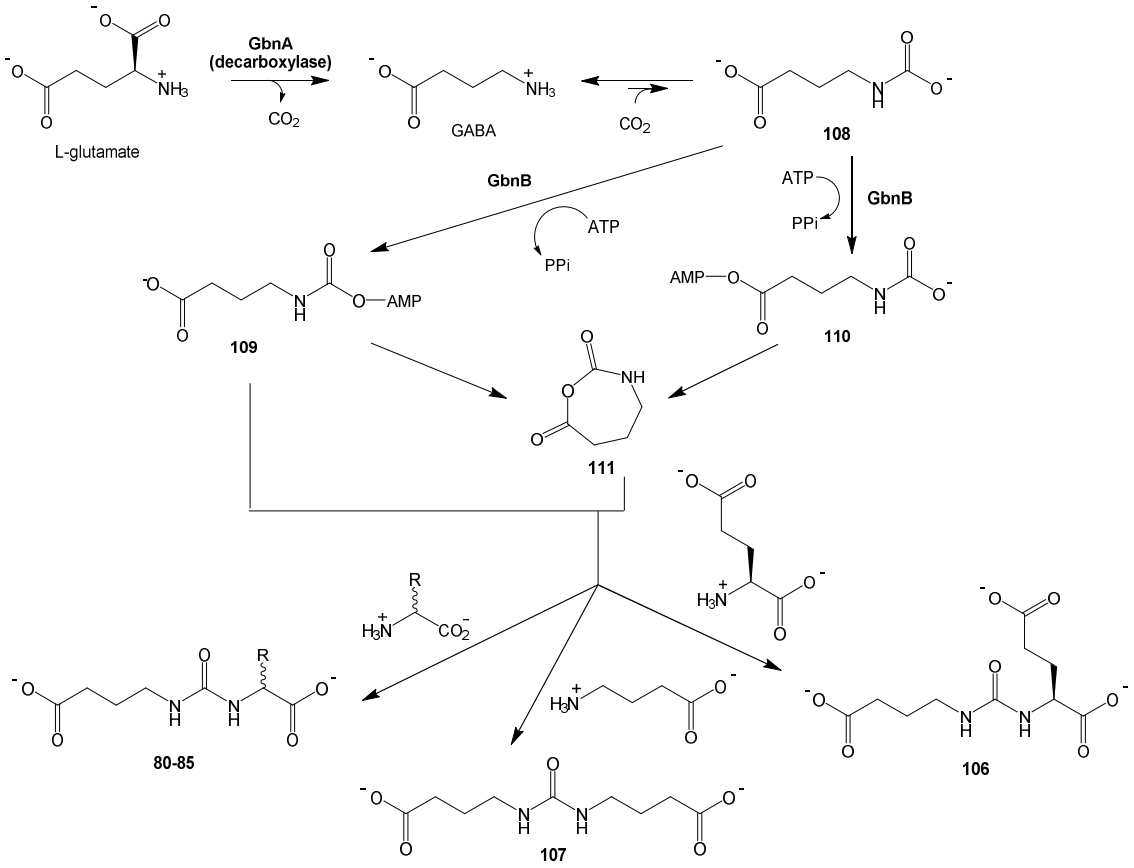
3.3.2 Proposed biosynthetic scheme

The combination of bioinformatic analyses and experimental results discussed thus far allow for a feasible biosynthetic route to gaburedins to be proposed. Primarily, the proposed decarboxylase function for GbnA, and the presence of a conserved GABA moiety across all of the urea natural products identified as being produced by the *gbnR::apra* strain, strongly suggest that L-glutamate is the cognate GbnA substrate.

The decarboxylation product of L-glutamate would be GABA, which could react spontaneously with CO_2 to yield a carbamate intermediate **108**, shown in Scheme 3.2. As this reaction would be reversible (one would expect a carbamate intermediate would readily be hydrolysed back to GABA in aqueous conditions) intermediate **108** could be trapped by activation with ATP, to form one of two AMP-ester intermediates, **109** or **110**.

Bioinformatic analyses suggested GbnB to be an AMP-ligase and experimental evidence suggested GbnB to be an essential enzyme for gaburedin biosynthesis. Hence GbnB was proposed to adenylate either the carboxyl or the carbamoyl group of compound **108**, (Scheme 3.2). Carboxyl adenylate **110** would need to undergo cyclisation to the *N*-carboxyanhydride **111**, which would then afford the gaburedins via ring opening with amino acids (this would involve nucleophilic attack at the less electrophilic carbonyl group of **111**, which is chemically unlikely). On the other hand, the carbamoyl adenylate **109** could react with the appropriate amino acids either directly, or via anhydride **111**, to form gaburedins.

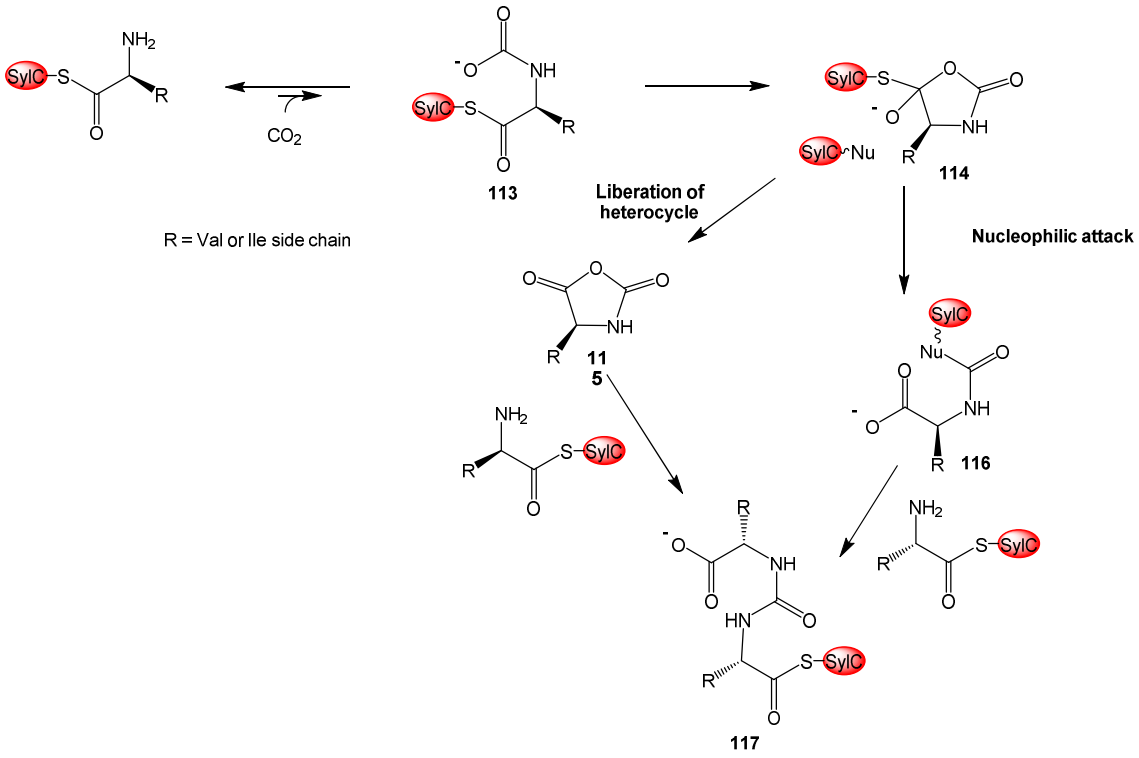
Other urea-containing natural products were discussed in Section 2.3 and include syringolins and pacidamycins.[163-165] Interestingly, the adenylation domain of SylC,



Scheme 3.2 **–** Proposed biosynthetic route to gaburedins. L-glutamate is decarboxylated to give GABA, which may proceed by one of two possible intermediates **109** or **111**, to furnish gaburedins

which has been shown to catalyse urea formation in syringolin biosynthesis, shares 39% similarity and 22% identity with GbnB.[165] The mechanism for SylC-catalysed urea formation has been proposed to involve a five-membered cyclic *N*-carboxyanhydride **115**, or a closely related species (Scheme 3.3). This anhydride is hypothesized to be formed via an analogous mechanism to that suggested for the formation of **111** (Scheme 3.2). Trapping of the anhydride with an isoleucyl or valinyl acyl carrier protein thioester would lead to formation of the ureido linkage of the syringolins. As noted above, an analogous mechanism for the formation of the gaburedins via intermediate **111** seems unlikely.

The AMP-carbamate intermediate **109** is sufficiently reactive that it would be a short-lived species in aqueous conditions; therefore it is likely that it would readily react with



Scheme 3.3 – Biosynthetic origin of ureido-linked dipeptides in syringolin biosynthesis as proposed by Imker and co-workers[165]

excess intracellular nucleophiles. Therefore, the incorporation of phenylalanine, leucine, isoleucine, valine, methionine and *N*-acetyl cysteine into the gaburedins A-F (**80-85**) may reflect the relative availability of amino acids in the cell at the time of formation of the AMP-intermediate **109**, rather than specific incorporation of these amino acids by an enzyme recognising particular structural features (GbnB). In the case of gaburedin F, this could be formed via acetylation of a cysteine-containing gaburedin, or by *N*-acetyl cysteine attacking the AMP-ester intermediate. However, both of these routes to gaburedin F would require a poorly-nucleophilic amide nitrogen lone pair (or an ureido-nitrogen lone pair) attacking an electrophile, both of which are chemically unlikely.

The presence of such a reactive intermediate species **109/111** is also consistent with the observation that a vast range of different substrates are incorporated into gaburedin

natural products; both L- and D- amino acids are tolerated aminoacyl moeities. The biosynthetic scheme proposed here would also account for the additional gaburedins **106** and **107** resulting from feeding experiments with L-glutamic acid both L-glutamate and the decarboxylation product GABA can also both act as nucleophiles to react with **109**/**111**.

Finally, the GbnC enzyme (belonging to a large family of small amino acid / peptide transporters) may be involved with export of the gaburedins into the extracellular milieu. Alternatively, GbnC may be involved in export of the AMP-carbamate intermediate **109**, allowing nucleophiles in the extracellular medium to react with it to furnish gaburedins.

3.3.3 Origin of the ureidyl carbon

To probe the biosynthetic origin of the central carbon atom and GABA moieties (blue in Figure 3.1), incorporation experiments with ^{13}C-labelled precursors were carried out. If the precursor to gaburedins is glutamate, one would expect that feeding the *gbnR::apra* mutant with ^{13}C-labelled L-glutamic acid would result in ^{13}C labelled gaburedins.

Initially, to probe the origin of the ureidyl carbonyl carbon, the *gbnR::apra* strain was grown on SMMS enriched with [1-^{13}C]-L-glutamic acid added to a 10 mM final concentration. If the ureidyl carbonyl is directly incorporated from the carboxylate group lost during the decarboxylation of L-glutamate to yield GABA (Scheme 3.2), then the ^{13}C label from [1-^{13}C]-L-glutamic acid added to the media would be expected to be retained through the biosynthesis, and therefore observed in the gaburedins extracted from this culture extract (Scheme 3.2). LC-MS analysis of the gaburedins

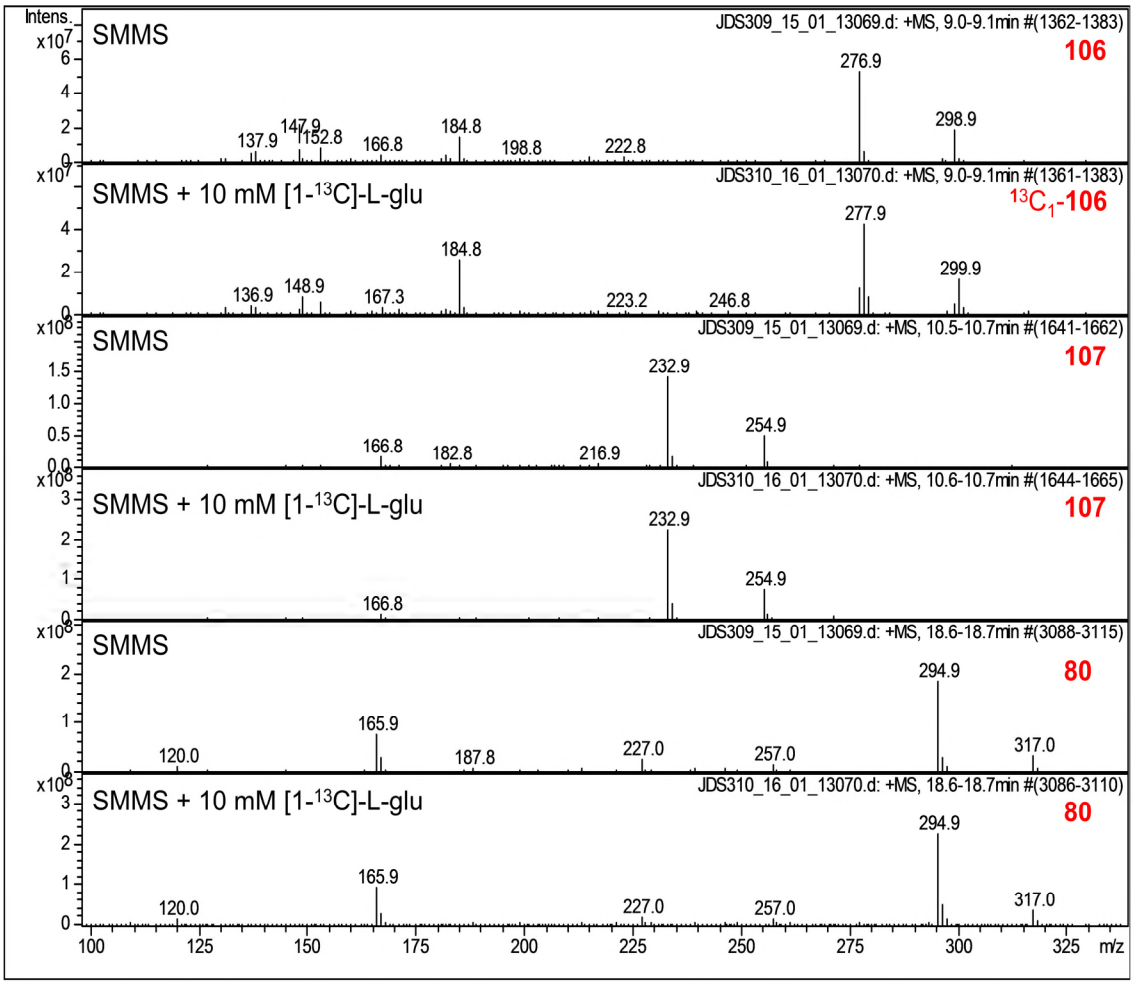


Figure 3.9 – Mass spectra of gaburedins **80**, **106** and **107** from metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS and SMMS enriched with 10 mM [1-^{13}C]-L-glutamic acid illustrating the only labelled gaburedin is gaburedin **106**, the product of the reaction of [1-^{13}C]-L-glutamic acid with the proposed biosynthetic intermediate **109/111**

extracted from the *gbnR::apra* strain grown on SMMS with [1-^{13}C]-L-glutamic acid revealed that there was no labelling of the gaburedins A-F (**80-85**) present in the culture extract (Figure 3.9). The only gaburedin extracted that contained a ^{13}C label was the gaburedin **106** derived from attack of the carbamate intermediate by [1-^{13}C]-L-glutamate, confirmed by the increase in the *m/z* values of the $[M+H]^+$, $[M+Na]^+$ and $[M+H–C_5H_7NO_3]^+$ ions by 1, compared with gaburedin **106** derived from unlabelled L-glutamic acid (Figure 3.9).

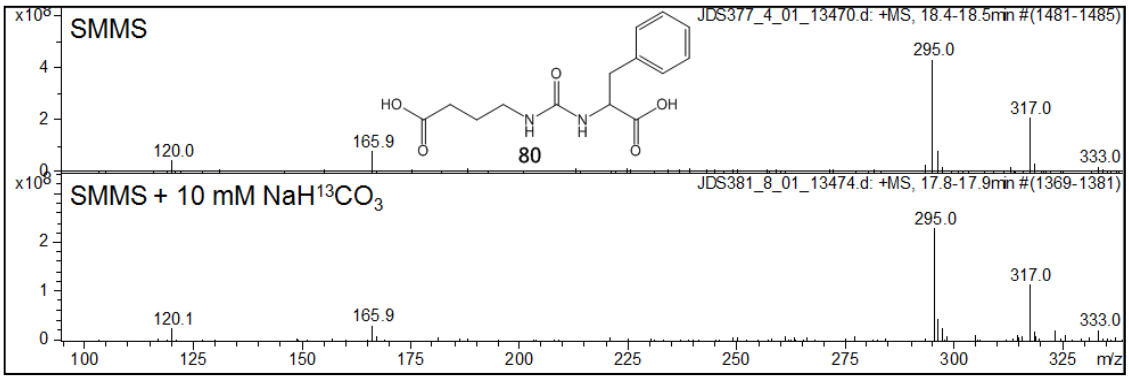


Figure 3.10 **–** Mass spectra of gaburedin A (**80**) in metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS (top panel) and SMMS enriched with 10 mM $NaH^{13}CO_3$ (bottom panel), illustrating that the ^{13}C label from $NaH^{13}CO_3$ is not retained in gaburedin biosynthesis

The result of this labelling experiment therefore confirms that the ureidyl carbon is not derived from L-glutamate. It was predicted that if the ureidyl carbon is derived from sodium hydrogen carbonate, that specific labelling of gaburedins would be observed when the *gbnR::apra* strain was grown on SMMS enriched with 10 mM $NaH^{13}CO_3$ as in the case of the ureidyl carbon present in syringolins.[165] LC-MS analysis of the culture extract revealed that the mass spectra of the gaburedins observed were unchanged compared to those extracted from unmodified SMMS (Figure 3.10), suggesting that the ureidyl carbon is not derived from sodium hydrogen carbonate.

3.3.4 Origin of the conserved GABA moiety

The *gbnR::apra* strain was then grown on SMMS enriched with [U-^{13}C]-L-glutamic acid, as it was hypothesised four of the ^{13}C labels would be retained throughout the biosynthesis of the gaburedins and therefore be detected in the gaburedins extracted. This would be consistent with the previous observation that the C-1 carbon lost in the decarboxylation step is not retained in the final gaburedin natural products observed in the culture extract, but the other four carbon atoms are retained in gaburedins produced.

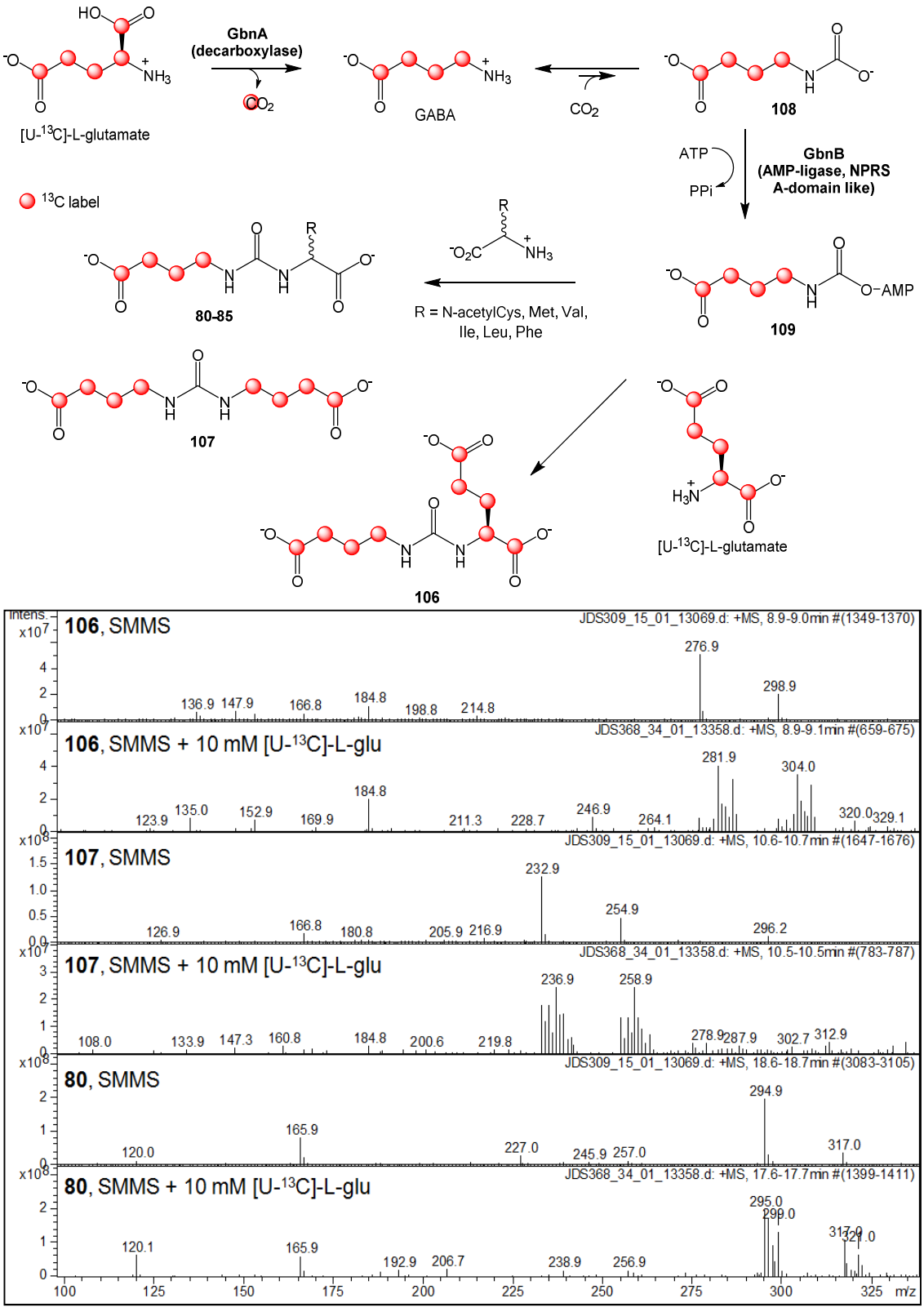


Figure 3.11 – Biosynthetic route to gaburedins, showing the proposed incorporation of ^{13}C labels upon feeding the *gbnR::apra* strain with [U-^{13}C]-L-glutamic acid (top panel). Mass spectra of gaburedins A **80**, **106** and **107** from metabolites extracted from the *S. venezuelae gbnR* mutant when grown on SMMS and SMMS enriched with 10 mM [U-^{13}C]-L-glutamic acid (bottom panel), showing retention of ^{13}C labels in gaburedins identified in this culture extract

Indeed, LC-MS analysis of the gaburedins **80-85** present in the culture extract of the *gbnR::apra* strain on SMMS enriched with 10 mM [U-^{13}C]-L-glutamic acid revealed that there is specific incorporation of four ^{13}C labels into all six gaburedins A-F (Figure 3.11 shows the mass spectra of gaburedin A upon feeding the *gbnR::apra* mutant with [U-^{13}C]-L-glutamic acid compared with feeding with 10 mM unlabelled L-glutamic acid). Although the isotope patterns in the mass spectra are complicated, the predominant *m/z* for the $[M+H]^+$ species is shifted from 295.0 to 299.0 for gaburedin A **80**, however the fragment ion $[M+H-C_5H_7NO_3]^+$ still has *m/z* = 166.0 as would be expected if the ^{13}C labels are in the conserved GABA moiety, and are therefore lost upon loss of the neutral $C_5H_7NO_3$ fragment.

In addition, consistent with previous data from feeding with unlabelled L-glutamic acid, the gaburedin **107** derived from GABA was also observed, again the *m/z* of its molecular ion being increased in value by 4 from *m/z* = 232.9 to 236.9, and the $[M+Na]^+$ peak increased in value from *m/z* = 254.9 to 258.9.

Furthermore, the mass spectrum of the gaburedin **106** derived from L-glutamate showed an increase in its *m/z* = 276.9 to 281.9 and 286.0 for the $[M+H]^+$ ion, consistent with the incorporation of four ^{13}C labels into the conserved ureidyl-GABA moiety, in addition to five ^{13}C labels consistent with the addition of [U-^{13}C]-L-glutamic acid to the proposed AMP-carbamate intermediate (Figure 3.11). UHR-LC-MS analysis confirmed the molecular formulae of the ions for the unlabelled and labelled gaburedins containing ^{13}C-labels.

The relative abundance of the labelled vs unlabelled peaks in the mass spectra for gaburedins A-F **80-85**, gaburedin **106** and gaburedin **107** shown in Figure 3.11 may

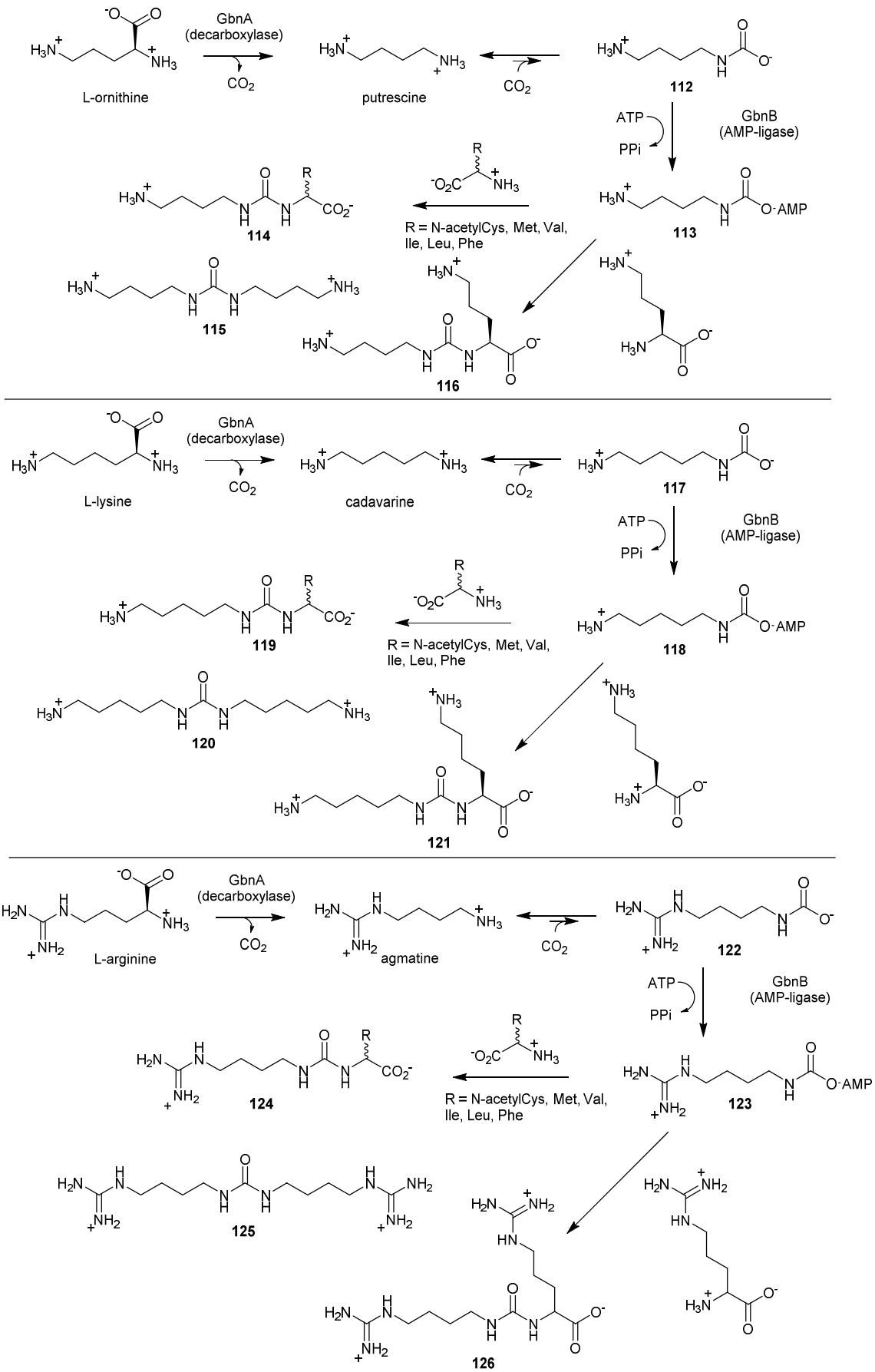
have been affected by the relative amounts of labelled vs unlabelled glutamate present in the media. It would be expected that the relative intensities of unlabelled:labelled peaks would be approximately 1:4 for each gaburedin, as there is 2.8 mM L-glutamic acid in the media from the casaminoacids, yet the media was supplemented with 10 mM [U-^{13}C]-L-glutamic acid, also. The complex fragmentation patterns observed may simply reflect the complicated role of L-glutamic acid in central metabolism, and supplementing the growth media with such a large quantity (final concentration of 12.8 mM L-glutamic acid in total compared to 2.8 mM in unmodified SMMS) may have diverted metabolic flux to other pathways rather than into gaburedin biosynthesis.

3.4 Probing substrate diversity of GbnA and GbnB

3.4.1 Substrate specificity of GbnA

Contrary to the bioinformatic analysis of GbnA, the experimental evidence presented thus far implies that L-glutamate, rather than L-ornithine, L-lysine or L-arginine is utilised by GbnA as a substrate. To determine whether any gaburedin analogues could be derived from decarboxylation of L-ornithine, L-lysine or L-arginine, the *gbnR::apra* mutant was grown on SMMS supplemented with these amino acids. Extracellular metabolites produced after 3 days were then extracted and submitted to LC-MS analyses.

As the enzyme GbnA and its orthologues belong to the Orn/Lys/Arg decarboxylase family, extracted ion chromatograms were generated for the *m/z* of species that would be expected if the decarboxylated form of the precursor molecules were incorporated to give gaburedin-like compounds derived from GABA analogues. For example, decarboxylation of L-ornithine would yield putrescine, L-lysine would yield



Scheme 3.4 – Biosynthetic route to gaburedin analogues that would be produced if GbnA decarboxylated L-ornithine, L-lysine or L-arginine, and if the resulting carbamoyl amines are activated by GbnB

cadaverine, decarboxylation of L-arginine would yield agmatine.[171,172] It is conceivable that each of these amines could spontaneously react with CO_2 to give analogous carbamates **112**, **117** or **122** shown in Scheme 3.4.

These carbamoyl compounds could then be activated by GbnB to yield the intermediates **113**, **118** and **123**, resulting from activation of carbomoyl derivatives of putrescine, cadaverine, and agmatine, respectively. These adenylated carbamoyl intermediates could then react with a range of nucleophiles; they could react with the amino acids that form gaburedins A-F (compounds **114**, **119**, and **124** in Scheme 3.4); they could react with the intermediate amines formed by action of GbnA to give compounds **115**, **120** and **125**; or they could react with the precursor amino acids to form **116**, **121** and **126** from L-ornthine, L-lysine and L-arginine, respectively. These potential biosynthetic processes are summarised in Scheme 3.4.

Extracted ion chromatograms were generated for the *m/z* of the ions expected if cadaverine, agmatine or putrescine were incorporated instead of GABA to yield the ureas **114-116**, **119-121** and **124-126**. No compounds corresponding to these *m/z* values were observed, implying that no putrescine-, cadaverine- or agmatine-derived gaburedin analogues are produced by the *gbnR::apra* strain on these supplemented media.

Extracted ion chromatograms were also generated for the $[M+H]^+$ ions of the gaburedins **127-129** that would be observed if the amino acids L-arginine, L-lysine and L-ornithine were incorporated into the aminoacyl moiety of gaburedins instead of the amino acids found in gaburedins A-F by reacting with intermediate **109**. Again, LC-MS analysis of the culture extracts revealed there to be no gaburedins with the expected *m/z*

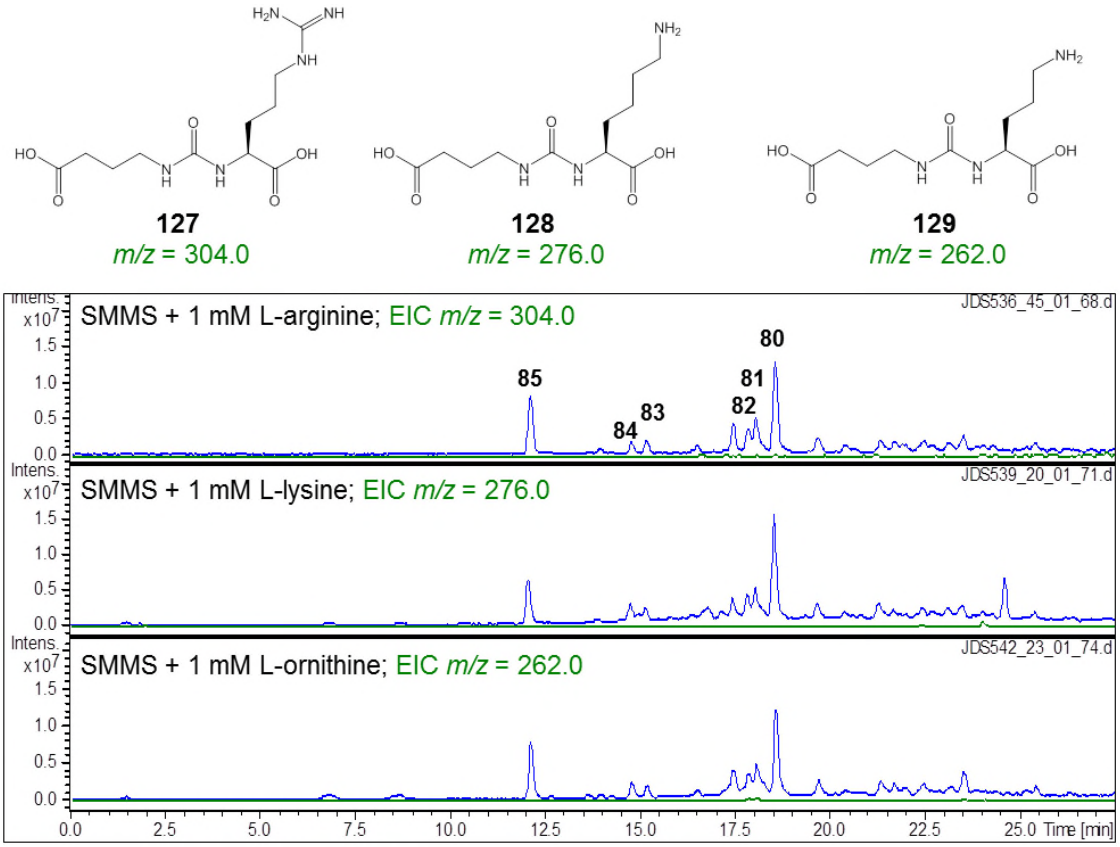


Figure 3.12 – Extracted ion chromatograms for *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue) for gaburedins A-F (**80**-**85**) extracted from the *S. venezuelae gbnR* mutant when grown on SMMS enriched with 1 mM L-arginine, 1 mM lysine and 1 mM L-ornithine. The structures and extracted ion chromatograms for the $[M+H]^+$ species of gaburedins **127**-**129** expected if these amino acids attack intermediate **109** are also shown

values of 304.0, 276.0 and 262.0 corresponding to **127**, **128** and **129** respectively. However, the gaburedins A-F could still be observed in the metabolites extracted from the SMMS enriched with each of these three amino acids (Figure 3.12), implying that the increased concentrations of L-ornithine, L-lysine and L-arginine in these media had not inhibited GbnA.

In the case of L-ornithine, there was no gaburedin found with *m/z* = 262.0 as would be expected for incorporation of L-ornithine into gaburedin biosynthesis (Figure 3.12). However, there was a gaburedin derivative **130** with *m/z* = 243.9 for its $[M+H]^+$ ion, 18 Da less than expected. The feeding experiment was repeated with 20 mM L-ornithine

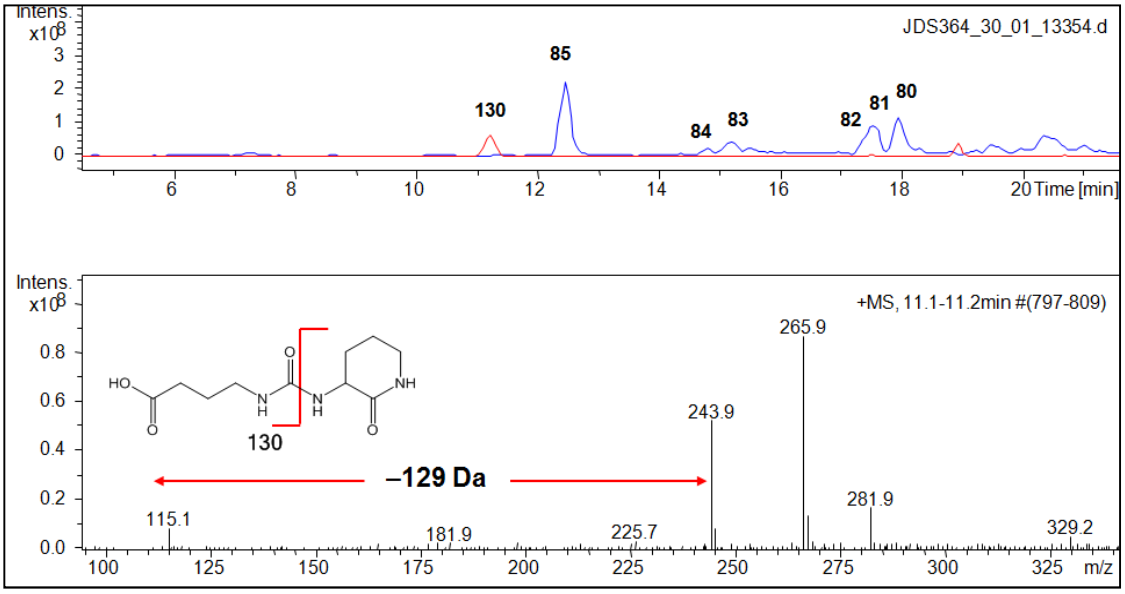


Figure 3.13 – Extracted ion chromatograms for *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue) for gaburedins A-F (**80**-**85**) and *m/z* = 244.0 (red trace) for new gaburedin-like compound **130** extracted from the *S. venezuelae gbnR* mutant when grown on SMMS supplemented with 20 mM L-ornithine (top panel) and mass spectrum and proposed structure of new gaburedin **130**

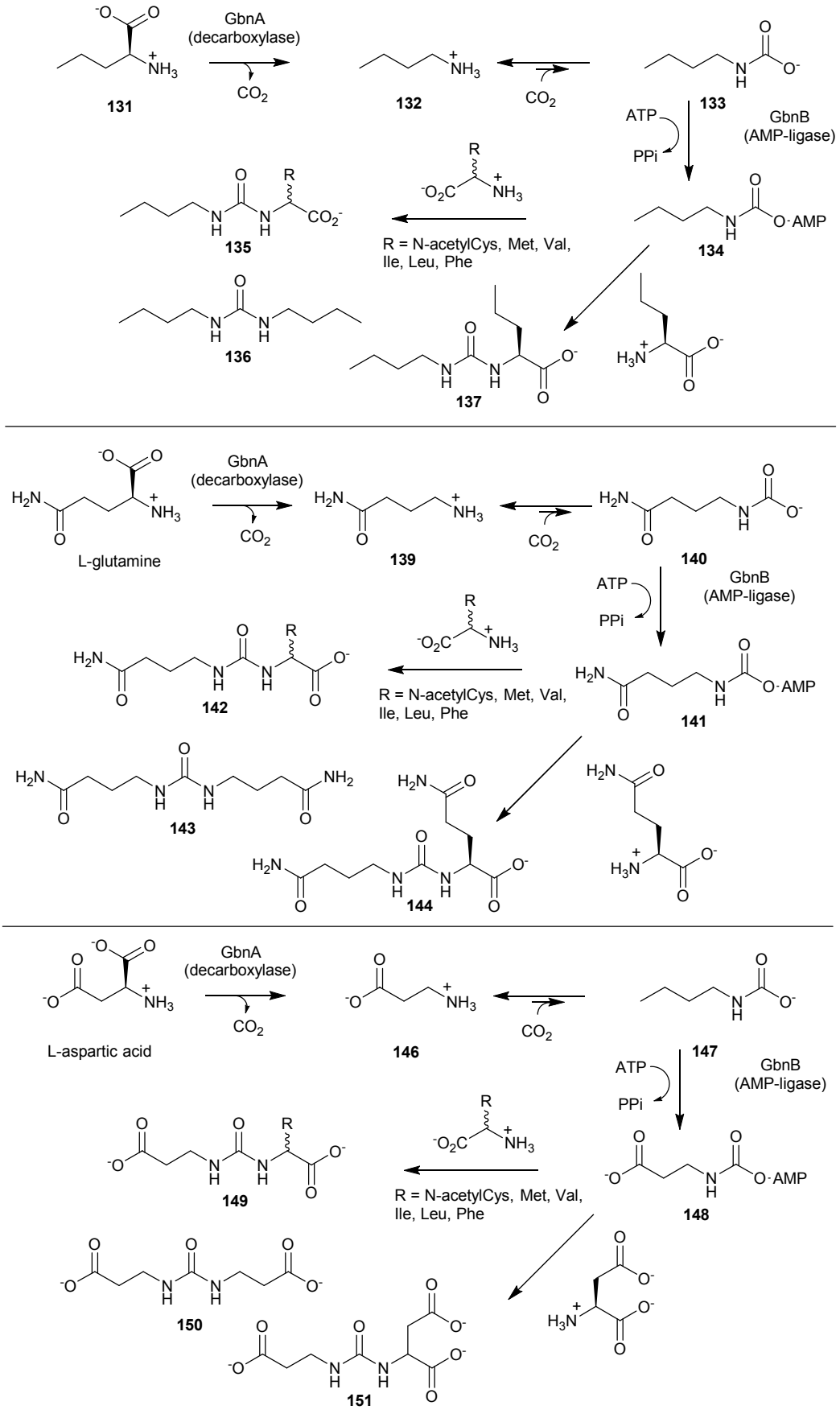
to verify the new compound **130** was present. This suggested that the L-ornithine fed to the *gbnR::apra* mutant was perhaps cyclised to 3-aminopiperidine-2-one before being incorporated into the gaburedin **130** shown in Figure 3.13.

UHR-LC-MS analysis confirmed the molecular formulae of each ion in the mass spectrum of this compound to be $C_{10}H_{18}N_3O_4$ for the $[M+H]^+$ ion (calculated *m/z* = 244.1292, found 244.1309), $C_{10}H_{17}N_3O_4Na$ for the $[M+Na]^+$ ion (calculated *m/z* = 266.1111, found 266.1116), consistent with a structure of 4-(3-(2-oxopiperidin-3-yl)ureido)butanoic acid **130**. The $C_5H_{11}N_2O$ for the $[M+H-C_5H_7NO_3]^+$ fragment ion (calculated *m/z* = 115.0866, found 115.0873) was also consistent with the molecular formula of 3-aminopiperidine-2-one. As L-ornithine was added to the SMMS before autoclaving, it is possible that the cyclisation occurred before being imported into the *gbnR::apra* cells.

These feeding experiments were repeated with media being supplemented with 20 mM each of L-arginine, L-lysine and L-ornithine. Indeed, as previously observed, in the metabolites extracted from the *gbnR::apra* mutant after being grown on these supplemented media, there were no new compounds observed with the expected *m/z* values for incorporation of L-arginine or L-lysine into new gaburedin analogues. It was observed however that the yield of gaburedins **80**-**85** was decreased in metabolites extract from the *gbnR::apra* mutant grown on SMMS enriched with 20 mM L-lysine. In the culture extract from the *gbnR::apra* strain grown on SMMS enriched with L-arginine, however, there was no significant reduction in the relative amount of gaburedins **80**-**85** produced compared to unmodified SMMS.

This series of feeding experiments was expanded on further by feeding L-norvaline (L-2-amino pentanoic acid) **131**, which is a potential substrate analogue for GbnA that has a propyl-side chain instead of a propionic acid side chain of L-glutamate. Therefore it was hypothesised that L-norvaline could be a substrate for GbnA, as the decarboxylation is occurring at the aminoacyl carboxylate compared with the side chain carboxylate. Decarboxylation of L-norvaline would produce *n*-butylamine **132**, which could be spontaneously activated by CO_2 to give **133**, which may be adenylated by GbnB to give **134**. Nucleophilc attack of this intermediate **134** would furnish a range of gaburedin analogues **135**-**137** with an *n*-butylamine moiety in place of the GABA moiety (Scheme 3.5).

Therefore, the *gbnR::apra* strain was grown on SMMS supplemented with 1 mM and 5 mM final concentration of L-norvaline, and the metabolites produced after 3 days were analysed by LC-MS. None of the *m/z* values for the species **135**-**137** were observed, implying that GbnA cannot catalyse decarboxylation of L-norvaline. However, an



Scheme 3.5 **–** Biosynthetic route to gaburedin analogues that would be produced if GbnA decarboxylated L-norvaline, L-glutamine or L-aspartate and the resulting carbamoyl amines are activated by GbnB

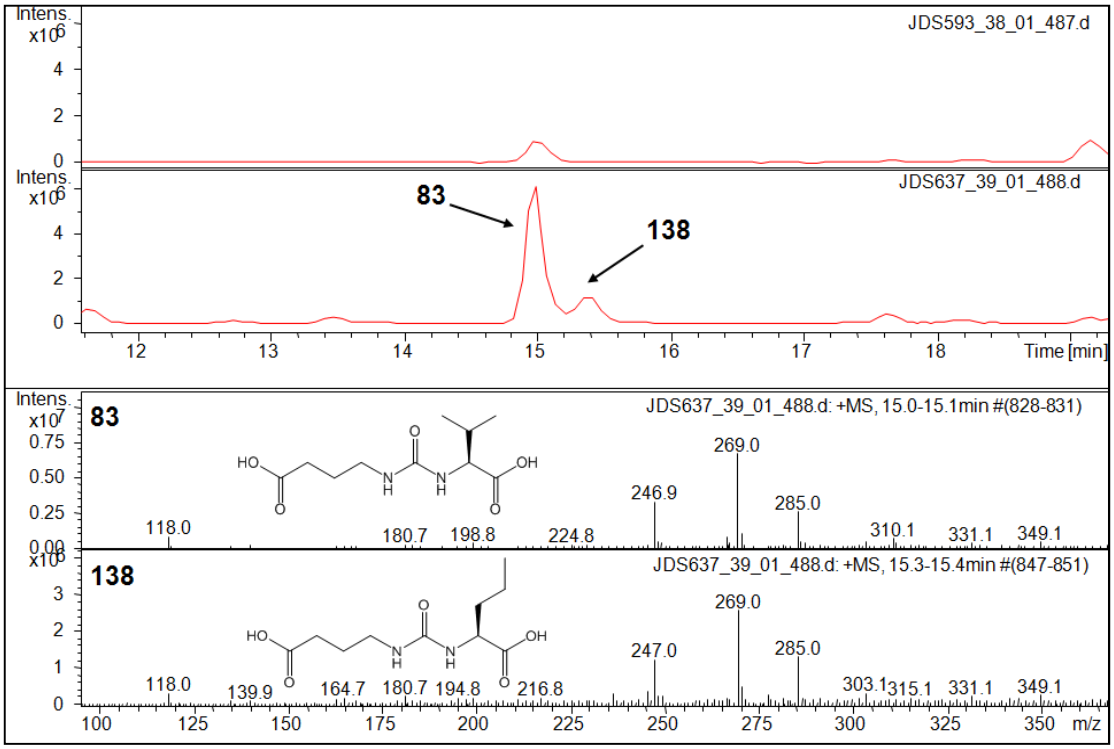


Figure 3.14 – Extracted ion chromatogram for *m/z* = 247.0 corresponding to gaburedin D (**83**) and new gaburedin **138** corresponding to incorporation of L-norvaline from the metabolites extracted from the *gbnR* mutant after growing on SMMS + 1 mM L-norvaline for 3 days

additional peak with *m/z* = 247.0 was observed, as would be expected if L-norvaline was acting as a nucleophile to attack the intermediate **109** to give a new gaburedin analogue **138** (Figure 3.14). This gaburedin has the same molecular formula as gaburedin D **83**, however, **138** elutes later. This is consistent with other molecules which only differ in the branching of their alkyl chains, as in the case of the germicidins and AHFCAs.[79,83]

To probe whether or not L-glutamine can be turned over by GbnA (L-glutamine has an amide instead of a carboxylic acid group at the end of its R substituent), SMMS was prepared with 20 mM final concentration of L-glutamine. If L-glutamine is decarboxylated by GbnA, 4-aminobutanmide **139** would result. This could then yield a series of gaburedin analogues **142-144** shown in Scheme 3.5. LC-MS analysis of metabolites extracted from the *gbnR::apra* strain grown on SMMS supplemented with

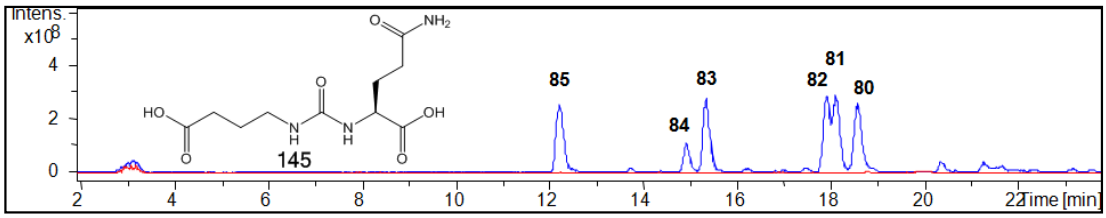


Figure 3.15 **–** Extracted ion chromatograms for *m/z* = 276.1 of gaburedin **145** which would be derived from L-glutamine (red trace), *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 of the gaburedins A-F (**80**-**85**, blue trace) in metabolites extracted from the *Streptomyces venezuelae gbnR* strain after being grown on SMMS containing 20 mM L-glutamine for 3 days

20 mM L-glutamine, however, revealed that no species with *m/z* values corresponding to the molecules **142-144** from Scheme 3.4 were observed.

There was also no observed compound with *m/z* = 276.1 as would be expected from incorporation of L-glutamine into gaburedin biosynthesis from attack of the intermediate **109**, to yield the gaburedin **145** (Figure 3.15**)**. However, the gaburedins A-F were still produced, therefore it would appear that L-glutamine is not inhibiting gaburedin biosynthesis, nor is it acting as a nucleophile to produce an addition gaburedin **145**.

A similar experiment was performed with L-aspartic acid in order to determine whether GbnA could tolerate a substrate with a shorter alkyl chain. Therefore extracted ion chromatograms for *m/z* = 278.9, 264.9, 231.0, 247.0 and 281.0 were generated for gaburedin-like ureas that would be observed if β-aminopropionate was inserted in place of GABA to furnish gaburedin analogues **149-151**. However, no compounds with these *m/z* values were identified in the culture extract of the *gbnR* mutant grown on SMMS supplemented with 10 mM L-aspartic acid. If β-aminopropionate is indeed produced by the action of GbnA on L-aspartate, however, it is conceivable that β-aminopropionate or its derivative would not be turned over by GbnB. Furthermore, there is no peak in

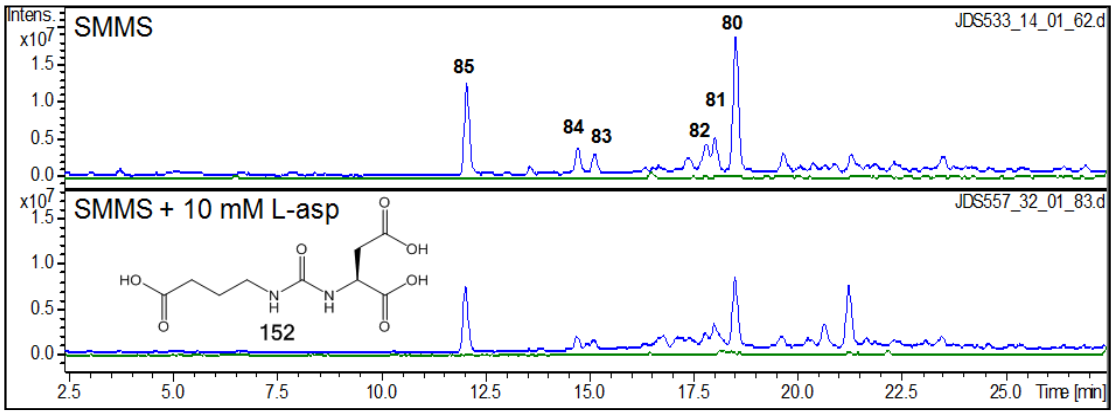
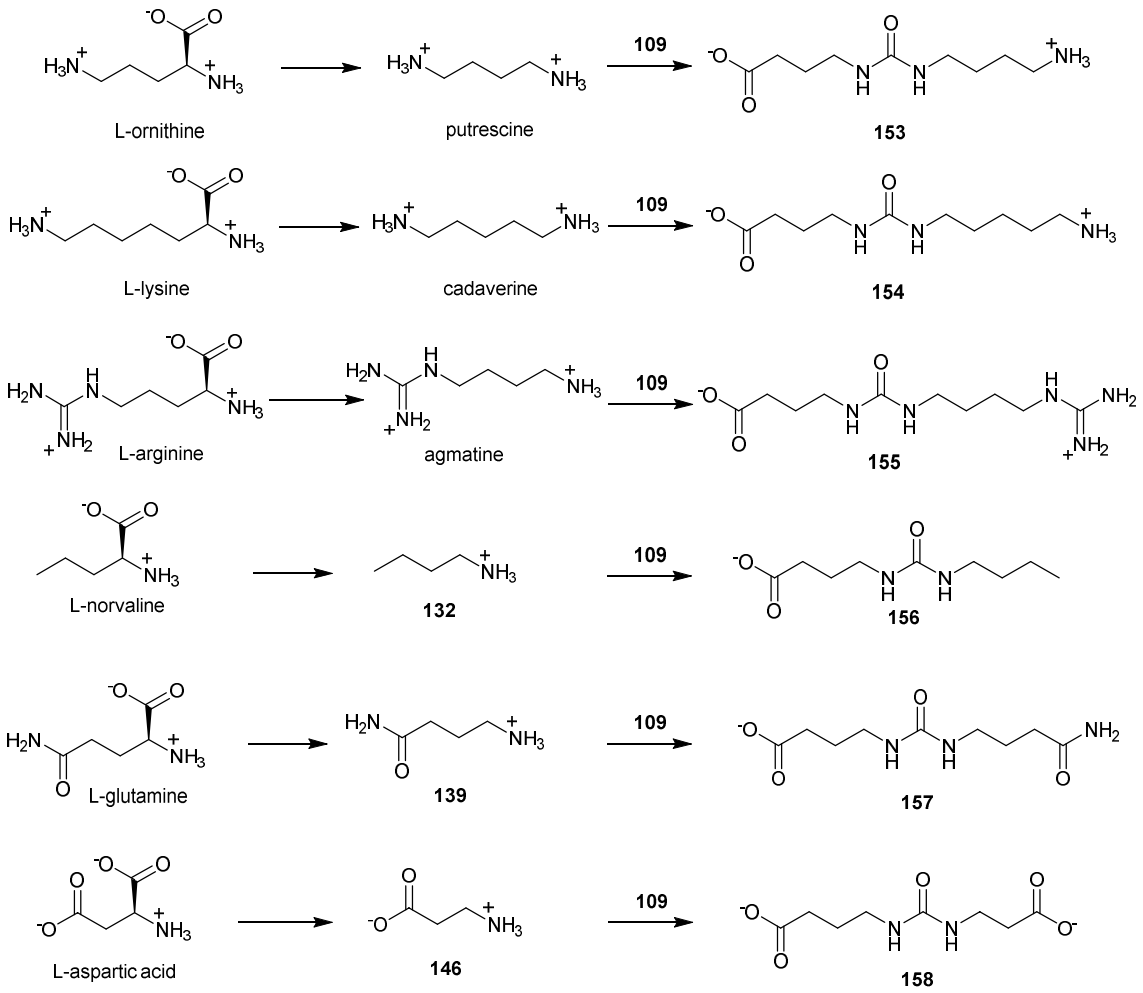


Figure 3.16 – Extracted ion chromatograms for metabolites extracted from the *S. venezuelae gbnR* mutant after growing on SMMS + 10 mM L-aspartate for 3 days. Gaburedins A-F (**80**-**85**) *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) and *m/z* = 263.0 (green trace) for new gaburedin **152** that would be expected for incorporation of L-aspartate into gaburedins via attack of intermediate **109**

the extracted ion chromatogram for *m/z* = 263.1 as would be expected for a gaburedin **152** containing L-aspartic acid, nor is there a peak with *m/z* = 219.0 corresponding to incorporation of β-aminopropionate attacking **109** to furnish **158**. Gaburedins A-F are still produced, however (Figure 3.16).

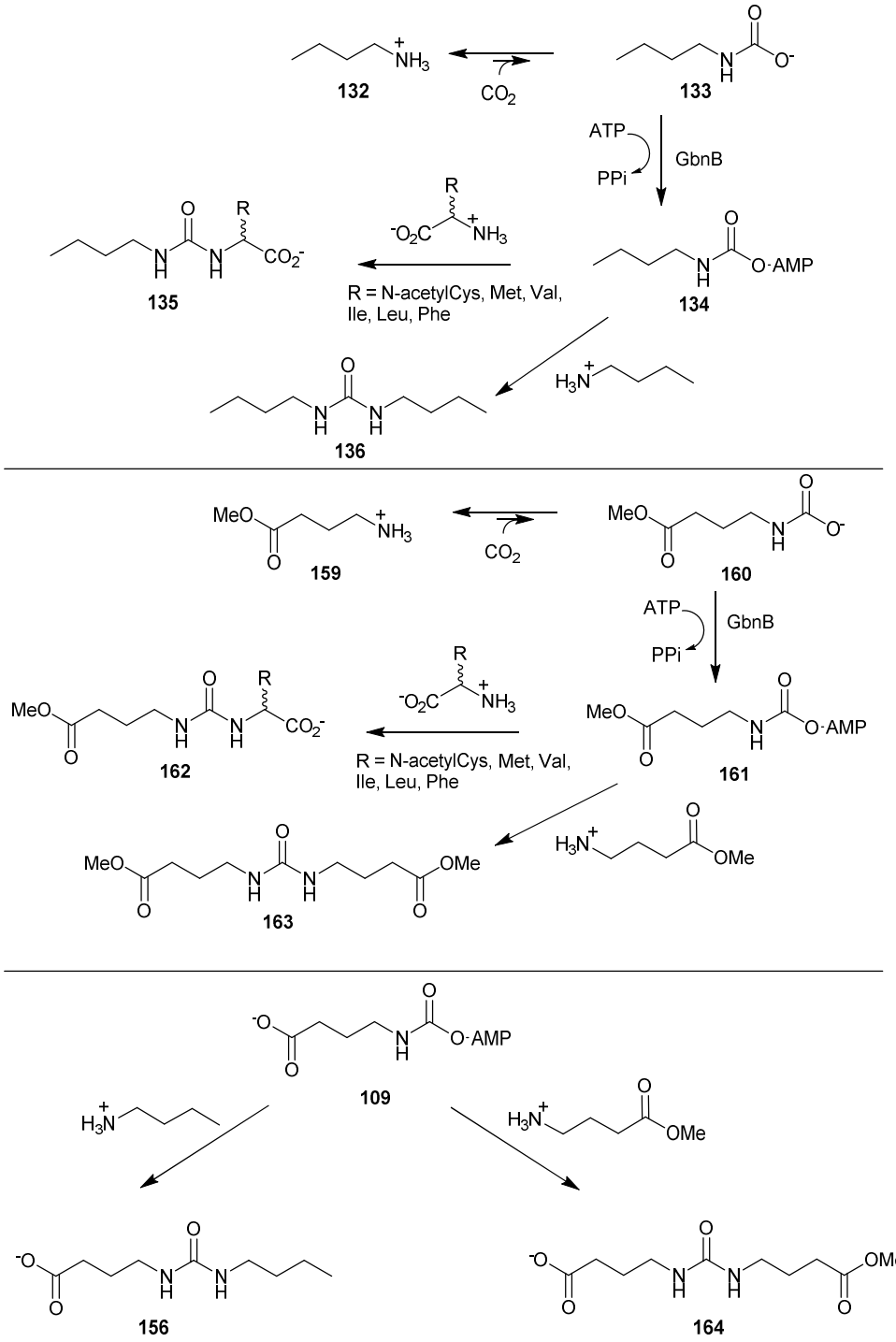
3.4.2 Probing GbnB substrate specificity

The feeding experiments in Section 3.4.1 suggest that the only substrate GbnA can tolerate out of the compounds tested is L-glutamate. However, these results do not confer much information regarding the substrate tolerance of GbnB. For example, it is possible that GbnA can tolerate other substrates (for example, L-norvaline could be decarboxylated to give *n*-butylamine), and the resulting amines/GABA analogues were simply not tolerated by GbnB and would therefore accumulate inside the cell. However, if there was an accumulation of these amines inside the cell, one might expect them to react with **109** to give a series of gaburedin analogues **153**-**158** (Scheme 3.6). No such species were observed, however.



Scheme 3.6 – Biosynthetic route to gaburedin analogues that would be produced if GbnA decarboxylated any of the putative precursors (left hand column) and the resulting amines attack **109** to furnish gaburedin analogues

In order to determine whether the carboxylic acid group of GABA is required for GABA (or a derivative of GABA) to be activated by GbnB, SMMS media supplemented with *n*-butylamine **132** and methyl 4-aminobutyrate **159** were prepared. The *gbnR::apra* mutant was grown on these media, metabolites extracted after 3 days, and analysed by LC-MS. If the GABA carboxylic acid group is not required for incorporation in gaburedin biosynthesis, it would perhaps be expected that *n*-butylamine could be activated in a similar way, leading to formation of a putative carbamoyl ester intermediate **134** as shown in Scheme 3.7. This intermediate **134** could



Scheme 3.7 – Biosynthetic route to gaburedin analogues that would be produced if carbamoylated *n*-butylamine or methyl 4-aminobutyrate were activated by GbnB (top panels) and structures of gaburedins containing *n*-butylamine or methyl 4-aminobutyrate via reaction with intermediate **109** (bottom panel)

then be available for attack by amino acid nucleophiles or *n*-butylamine itself to furnish gaburedin analogues **135** and **136**, respectively.

Similarly, if the carbamoyl derivative of methyl 4-aminobutyrate can be tolerated by GbnB as a substrate, it could form the carbamoyl adenylate **161**, which would be attacked by amino acids to furnish a series of methylated gaburedins **162**. Methyl 4-aminobutyrate could also attack intermediate **161** to give the new gaburedin **163** (Scheme 3.7). It is also possible that neither *n*-butylamine nor methyl 4-aminobutyrate would be activated by GbnB; they could both react with intermediate **109** to furnish new gaburedins **156** and **164** (Scheme 3.7).

Upon LC-MS analysis of the culture extract from the experiments with *n*-butylamine and methyl 4-aminobutyrate, it was found that no series of gaburedin analogues **135** or **162** were present (one would expect the $[M+H]^+$ ions corresponding to *n*-butylamine-derived analogues of gaburedins **135** to have *m/z* = 263.0, 249.0, 217.0, 231.0 and 265.0; the methylated gaburedin analogues **162** would have *m/z* = 307.0, 293.0, 247.0, 261.0 and 309.0). Furthermore, no peak with *m/z* = 173 was observed for compound **136** or peaks of *m/z* = 261.0 corresponding to species **163** which would arise from methyl 4-aminobutyrate attacking intermediate **161**.

As shown in Figure 3.17, the only new gaburedin analogue observed in the culture extract from SMMS enriched with *n*-butylamine had *m/z* = 202.9, as would have been expected for *n*-butylamine attacking the GABA-carbamoyl-AMP ester intermediate **109** to yield gaburedin **156**. Therefore, *n*-butylamine would appear to act only as a nucleophile attacking the intermediate **109** formed by GbnB, as opposed to acting as substrate for GbnB itself (Scheme 3.7). The fragment ion present at *m/z* = 104.0 is consistent with neutral loss of the amine and ureido moieties, leaving protonated GABA rather than neutral loss of the GABA moiety as was observed for all other gaburedins reported thus far.

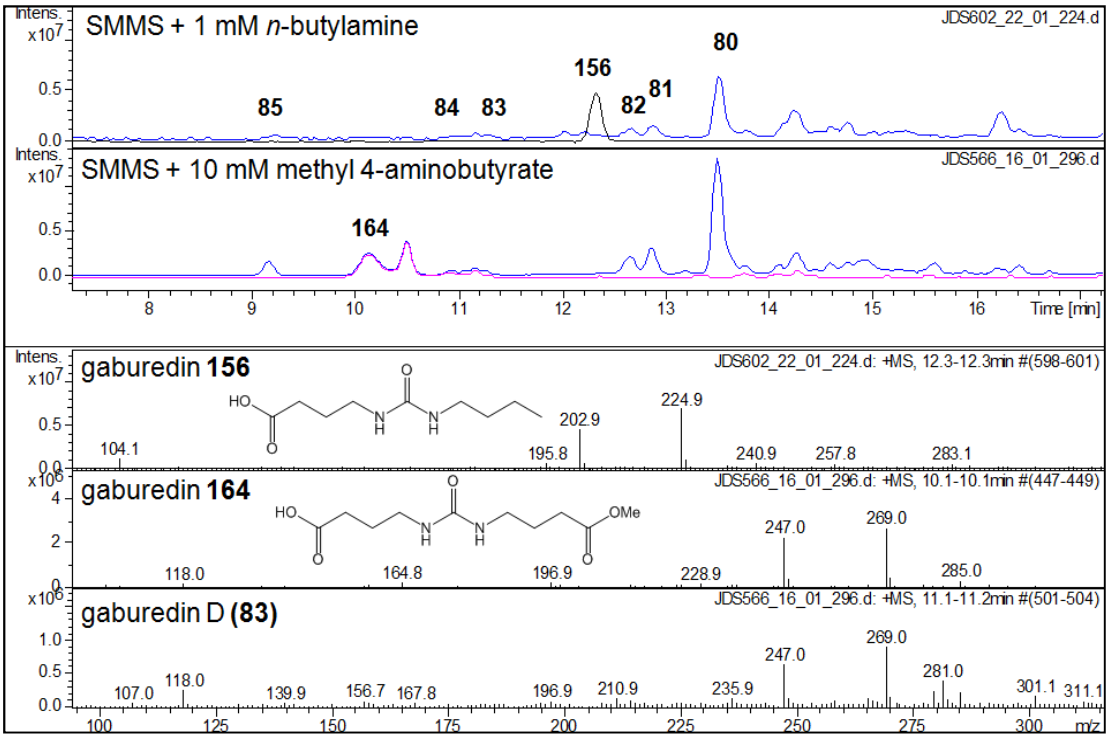


Figure 3.17 – Extracted ion chromatograms for gaburedins **80-85** *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) *m/z* = 203.0 (black trace) and *m/z* = 247.0 (pink trace) for new gaburedins **156** and **164** observed in the metabolites extracted from the *gbnR* mutant after growing on SMMS + 1 mM *n*-butylamine and SMMS + 10 mM methyl 4-aminobutyrate for 3 days

Interestingly, when the *gbnR::apra* strain was grown on SMMS supplemented with methyl 4-aminobutyrate, a new gaburedin **164** with the expected *m/z* of 247.0 for a gaburedin containing methyl 4-aminobutyrate was observed (Figure 3.17) and UHR-LC-MS analysis confirmed the molecular formulae to be $C_{10}H_{19}N_2O_5$ for the $[M+H]^+$ ion (calculated *m/z* = 247.1288, found 247.1290) and $C_{10}H_{18}N_2O_5Na$ for the $[M+Na]^+$ ion (calculated *m/z* = 269.1108, found 269.1110) however there was no observable peak at 118.1 as would be expected for the protonated GABA-*O*-methyl ester remaining after losing the conserved C_5H_7NO fragment.

It should be noted that the molecular formulae generated for the $[M+H]^+$ ion and $[M+Na]^+$ ions of this new compound **164** are identical to those generated for the molecular ions of gaburedin D **83** containing L-valine, however this new gaburedin **164**

also has a peak in its mass spectrum at *m/z* = 118.1, however this could correspond to a methyl 4-aminobutyrate fragment ion. Furthermore, this new gaburedin **164** elutes sooner (t_R = 10.2 min), compared with gaburedin D, (t_R = 11.3 min, Figure 3.17).

This compound corresponds to the *m/z* of the gaburedin that would be expected by attack of the methyl 4-aminobutyrate attacking the AMP-ester intermediate **109**. Therefore, it would appear that the methyl 4-aminobutyrate cannot act as a substrate for GbnB, as there would be a series of new methylated gaburedins expected to be observed if this was the case. This result implies that the carboxylic acid group of GABA is indeed required for turnover by GbnB.

These results of feeding *n*-butylamine and methyl 4-aminobutyrate do not change the proposed route to gaburedins as shown in Scheme 3.2, however. The carboxylic acid group of GABA may simply be required for substrate binding to GbnB (for example, favourable electrostatic interactions may arise from GbnB interacting with the carboxylate group of GABA). Even though it is not this carboxylic acid group which is being activated, changing the negatively charged carboxylate group of GABA for an uncharged methyl group of *n*-butylamine or the methyl ester group of methyl 4-aminobutyrate could lead to unfavourable interactions with the active site residues of GbnB that are involved in substrate binding; as a result these molecules are not activated by GbnB.

One other interpretation of these results could be that GbnA and GbnB operate as a complex within the cell, and the active site of GbnB is only accessible via the interface between GbnA and GbnB. This would imply that GbnB may be able to turn over *n*-butylamine or methyl 4-aminobutyrate, but the gaburedin analogues resulting from

activation of these substrates are not observed, because the GbnB substrate must arrive via GbnA rather than by diffusion into the active site from elsewhere in the cell.

3.5 Summary

From these *in vivo* experiments, it can be confirmed that GbnA does not catalyse the decarboxylation of lysine, ornithine or arginine, but instead catalyses the decarboxylation of glutamate into GABA. GABA (or a derivative of GABA) is then utilised by GbnB, which is likely involved in activation of **108**. This intermediate **109** then reacts with a range of nucleophiles (amino acids or amines) present in either the cell or extracellular milieu to yield the gaburedin natural products.

If the turnover of L-glutamate into GABA is indeed the reaction catalysed by GbnA, it would be expected that GABA would accumulate in a mutant lacking *gbnB*. Therefore, the *gbnB*/*gbnR* double mutant strain was grown on both SMMS and SMMS enriched with 1 mM, 5 mM and 10 mM L-glutamic acid, metabolites extracted and analysed by LC-MS. There was no observable accumulation of GABA, however.

The results from the feeding experiments presented thus far indicate that if the incorporation of amino acids into gaburedins is enzyme-catalysed, then the enzyme(s) involved must have quite broad substrate specificity (whilst the GABA-derived moiety is conserved, the amino acid moieties found on the other side of the ureido bridge in gaburedins A-F are structurally diverse). However, all new compounds identified thus far share the conserved $C_5H_7NO_3$ ureidyl-GABA fragment, implying that there is a much higher level of control of substrate tolerance by the enzymes GbnA and GbnB which are proposed to be involved in biosynthesis of the conserved GABA moiety,

compared with the final step of gaburedin biosynthesis which involves attack of the intermediate compound **109** by aminoacyl or amine nucleophiles. The structural diversity of different nucleophiles which can be incorporated into gaburedins is discussed in more detail in Chapter 4.

4. Exploring diversity of gaburedin aminoacyl moieties

In Chapter 3, feeding experiments with specific gaburedin precursors established that GbnA decarboxylates L-glutamate. It was hypothesised that GbnB adenylates the carbamoyl group of the GABA derivative **108** (Scheme 3.2). Nucleophilic attack of the amine onto the carbamoyl adenylate **109** results in ureido bridge formation and the gaburedin natural products.

Experiments presented thus far have indicated that the nucleophiles attacking this intermediate carbamoyl adenylate **109** do not require a carboxylic acid group to be incorporated into gaburedins (as illustrated by the observation that *n*-butylamine and methyl 4-aminobutyrate can attack intermediate **109** to form gaburedins **156** and **164**, respectively, Scheme 3.6). Furthermore – as noted in Chapter 2 – the side chains of amino acids that form gaburedins A-F are very diverse. Given the diversity of nucleophiles that can be incorporated into gaburedins, it was proposed that this step in gaburedin biosynthesis is not enzyme-catalysed. Incorporation experiments with a variety of other nucleophiles could therefore be performed to investigate the formation of other gaburedin analogues with even more diverse R and C-terminal groups (red and green in Figure 4.1).

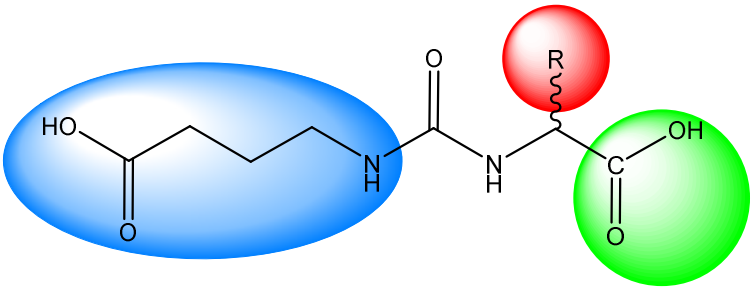


Figure 4.1 – General structure of gaburedins identified thus far. Chapter 4 will focus on the diversity of substrates with different amino acid side chains (red) and different groups in place of the aminoacyl carboxylic acid group (green) which can be incorporated into gaburedins

4.1 *In vivo* investigations of aminoacyl moiety tolerance

4.1.1 Incorporation experiments with small amino acid side chains

As the incorporation of amino acids into the aminoacyl moiety of gaburedins does not appear to be enzyme-catalysed, the level of incorporation of each of the non-polar amino acids phenylalanine, leucine, isoleucine, valine, methionine and cysteine (or *N*-acetylcysteine) may reflect their abundance. Figure 4.2 illustrates the concentration of amino acids present in the medium used (SMMS). The amino acids phenylalanine, leucine, isoleucine, valine and cysteine are not the most abundant amino acids in the media, but are the ones which are incorporated into gaburedins **80-85** in unmodified culture medium.

As described in Section 2.2.4, specific gaburedins can be overproduced by supplementing the culture medium with a particular amino acid. To gain further insight

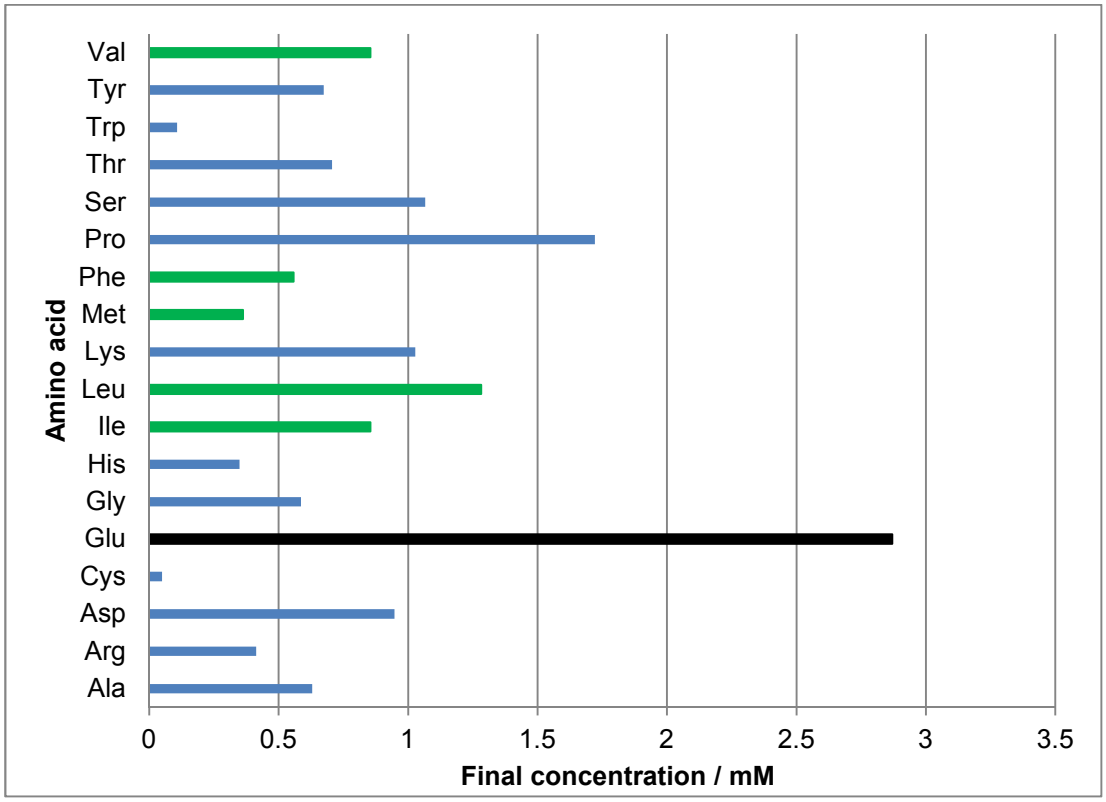


Figure 4.2 – Molar concentrations of proteinogenic amino acids in SMMS[162]

into the biosynthesis of gaburedins, SMMS modified to include greater concentrations of other amino acids was prepared. As described in the previous chapters, the *gbnR::apra* strain was grown on these media, the metabolites extracted and analysed by LC-MS to determine whether any additional gaburedin-like molecules were produced when additional amino acids were added. As in previous experiments, control plates of the modified media were prepared but not inoculated with bacteria. All plates were incubated at 30 °C for 3 days and ethyl acetate extracts performed. The samples were analysed by LC-MS and the presence of any new gaburedin-like molecules were confirmed by comparing the chromatograms of these samples with and without bacteria. These analyses excluded the possibility that the new metabolites observed were not simply products of media components reacting together during autoclaving or subsequent incubation at 30 °C.

The amino acids L-phenylalanine, L-leucine, L-isoleucine, L-valine, L-methionine and *N*-acetyl-L-cysteine all share similar properties (they all belong to the group of amino acids with uncharged side chains, whose overall charge is zero at pH 7). Therefore, growth media supplemented with other amino acids possessing uncharged side chain groups (namely L-alanine, glycine, L-serine, L-threonine) were prepared in order to determine whether the *gbnR::apra* strain could produce gaburedins containing these amino acids also.

Upon growing the *gbnR::apra* strain on SMMS supplemented with 20 mM L-alanine, LC-MS analysis revealed that in addition to the previously reported gaburedins A-F (**80**-**85**), there was also production of another new metabolite **165** with a retention time of 7.0 min, with peaks in its mass spectrum at *m/z* = 219.4 and *m/z* = 241.2, consistent with protonated and sodiated adducts of the same molecule (Figure 4.3). In addition, a

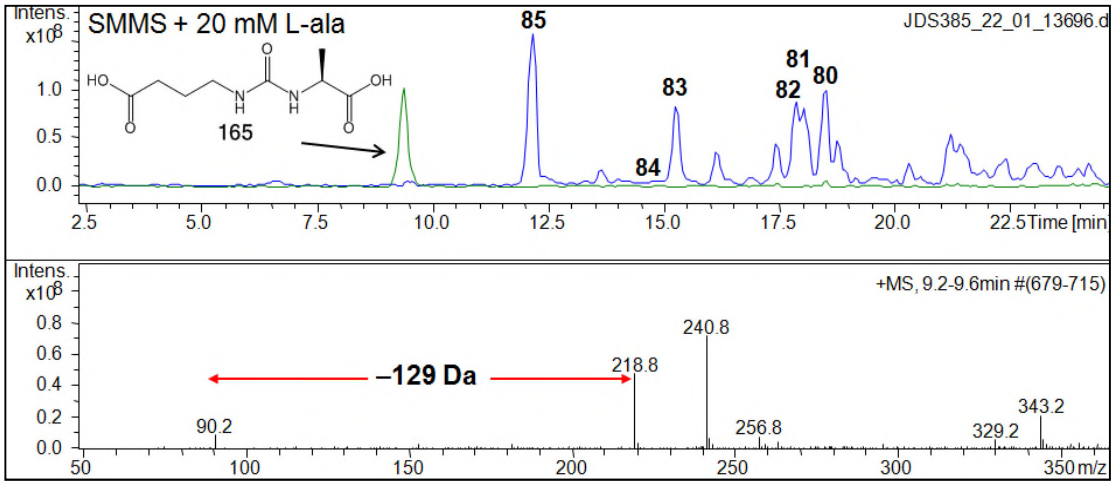


Figure 4.3 – Extracted ion chromatograms for gaburedins A-F **(80-85)** *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue trace) and new gaburedin **165** *m/z* = 219.0 (green trace) in metabolites extracted at pH 3 from the *Streptomyces venezuelae gbnR* strain after being grown on SMMS + 20 mM L-alanine for 3 days. The mass spectrum of new gaburedin **165** is also shown

m/z = 90.2 fragment ion expected for protonated L-alanine consistent with the characteristic gaburedin mass loss of 129 Da was observed.

Analogous experiments were performed on SMMS supplemented with L-proline, glycine, L-serine and L-threonine; LC-MS analyses revealed the presence of new gaburedin-like species **166-169** in these extracts corresponding to *m/z* = 244.0, 204.8, 234.8 and 248.9 for the $[M+H]^+$ ions of each new gaburedin containing L-proline, glycine, L-serine and L-threonine, respectively (Figure 4.4). All these new gaburedins exhibited the same characteristic gaburedin fragmentation pattern, furnishing ions corresponding to the protonated amino acids from each gaburedin. The incorporation of L-proline into gaburedin **166** is particularly interesting, as it is the first example of a secondary amine being incorporated into gaburedins.

This incorporation experiment was repeated with 1 mM, 5 mM and 10 mM final concentration of L-cysteine. In the extract of the *gbnR::apra* strain grown on SMMS supplemented with 1 mM final concentration of cysteine, there was very little change in

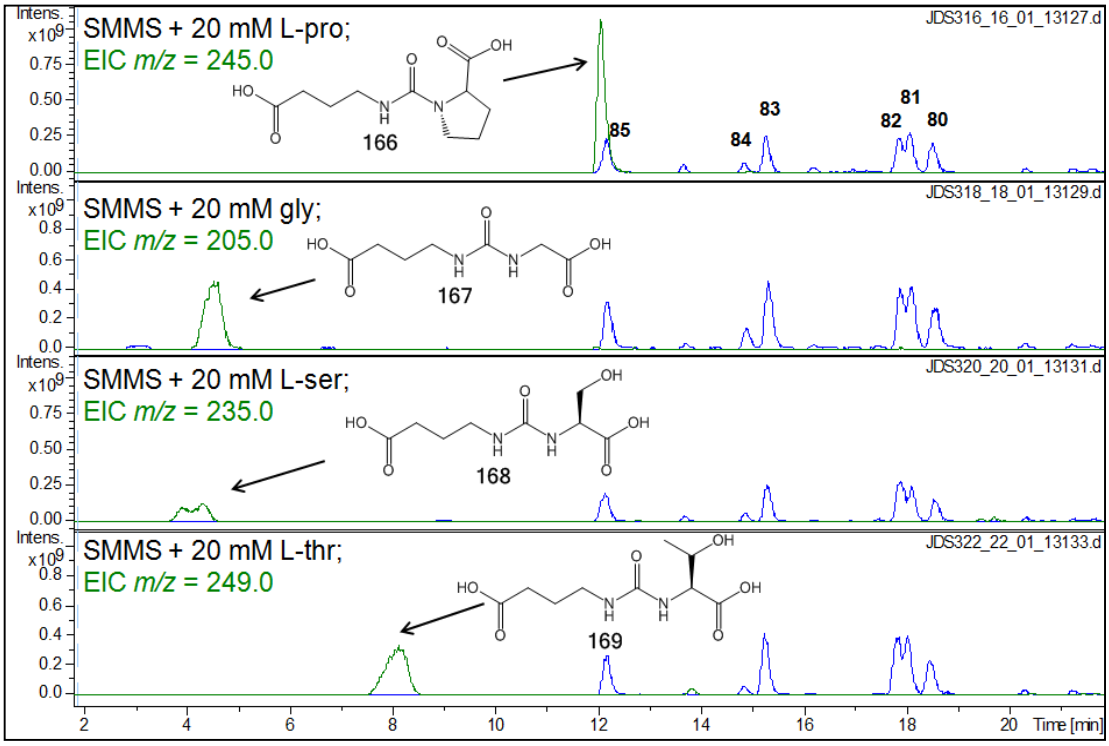


Figure 4.4 – Extracted ion chromatograms for gaburedins A-F (**80**-**85**) *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) and new gaburedins (green traces) of metabolites extracted at pH 3 from the *Streptomyces venezuelae gbnR* strain after being grown on SMMS + 20 mM L-proline, glycine, L-serine and L-threonine for 3 days

the yield of gaburedins, with 5 mM cysteine, there was a slight increase in the yield of the *N*-acetyl-cysteine derived gaburedin F (**85**). However, with 10 mM or 20 mM final concentration of L-cysteine, there was no production of gaburedins **80**-**85**.

In all cases, there was no new gaburedin **170** with the expected *m/z* = 251 observed in the culture extracts (Figure 4.5). There are fewer peaks in the base peak chromatogram of the culture extract from feeding with 10 mM L-cysteine, implying that generally there are fewer metabolites produced by the *gbnR::apra* strain in these conditions. This could perhaps be a stress response of the cells as a result of the dramatic increase in free thiols caused by the high concentration of L-cysteine in the media.

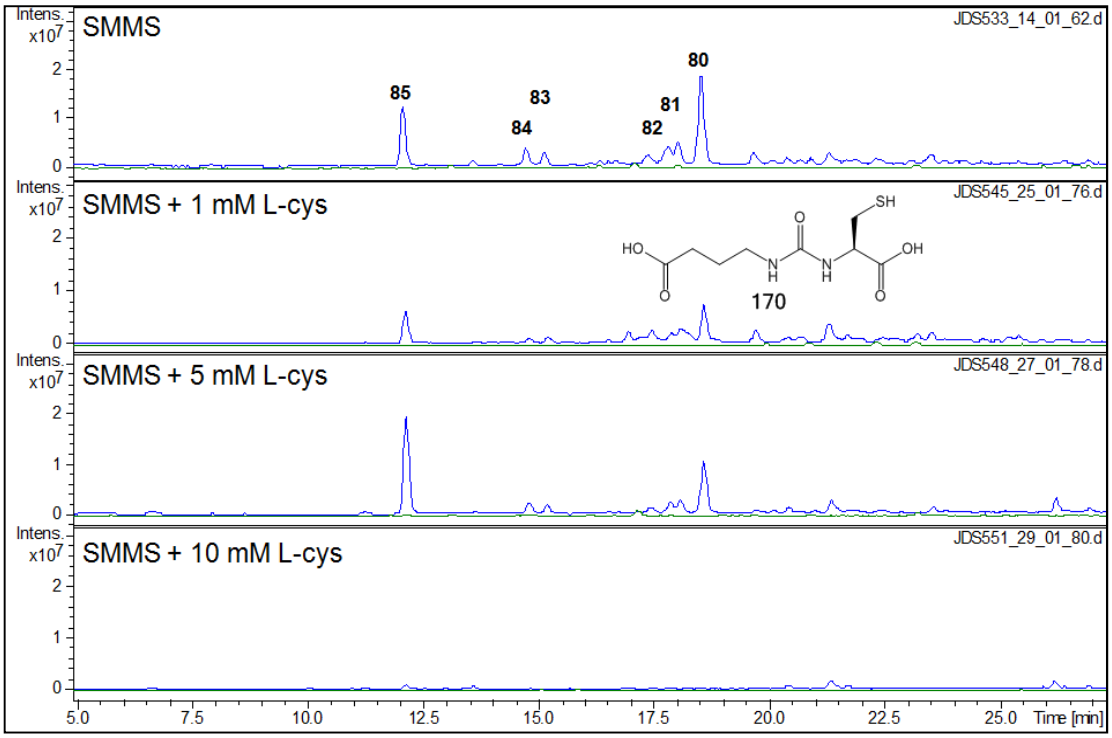


Figure 4.5 – Extracted ion chromatograms for gaburedins A-F (**80**-**85**) *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) and *m/z* = 251.0 for the expected new gaburedin **170** that would be expected from incorporation of L-cysteine (green traces) in metabolites extracted at pH 3 from the *Streptomyces venezuelae gbnR* strain after being grown on SMMS and SMMS supplemented with L-cysteine

4.1.2 Incorporation experiments with large amino acid side chains

So far, small polar and non-polar amino acids have been shown to be incorporated into the aminoacyl moiety of gaburedins. To further explore the range of polar amino acids which can be tolerated, feeding experiments were performed in which SMMS was supplemented with 20 mM L-asparagine, L-tyrosine and L-histidine.

In the experiments where SMMS was supplemented with 20 mM L-asparagine, a new metabolite with *m/z* = 262.1 was observed as would be expected for the gaburedin containing L-asparagine. UHR-MS confirmed that this new compound had both the molecular formulae and fragmentation pattern expected for an L-asparagine-derived gaburedin: $C_9H_{16}N_3O_6$ for the $[M+H]^+$ ion (calculated *m/z* = 262.1034, found 262.1035), $C_9H_{15}N_3O_6Na$ for the $[M+Na]^+$ ion (calculated *m/z* = 284.0853, found

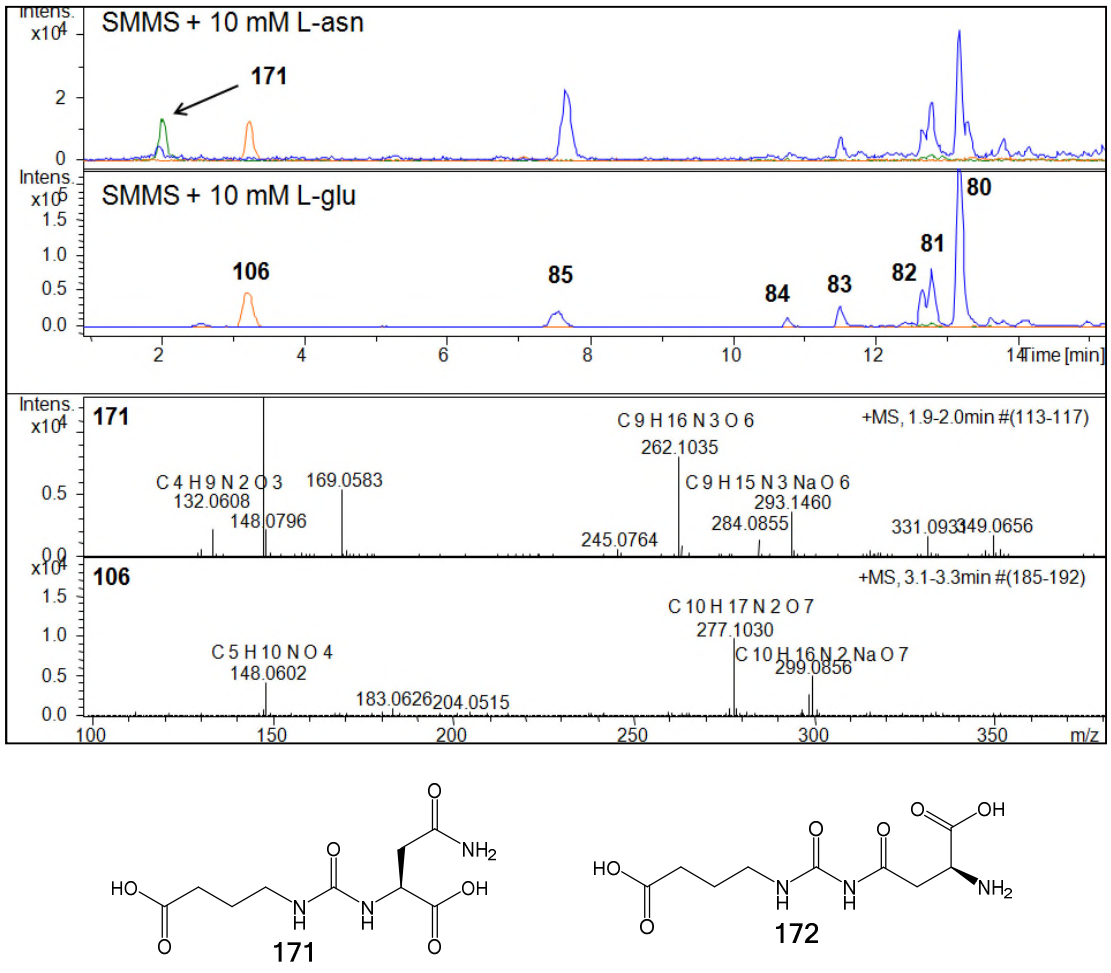


Figure 4.6 – Extracted ion chromatograms for *m/z* = 262.0 of gaburedin **171** (green trace) and *m/z* = 277.0 of the gaburedin **106** from incorporation of L-glutamate (orange trace) in metabolites extracted at pH 3 from the *Streptomyces venezuelae gbnR* strain after being grown on SMMS containing 20 mM L-asparagine and SMMS + 20 mM L-glutamate (top panels), and mass spectra of gaburedins **171** and **106**

284.0855), and $C_4H_9N_2O_3$ for the $[M+H-C_5H_7NO_3]^+$ fragment ion (calculated *m/z* = 133.0608, found 133.0608) corresponding to protonated L-asparagine.

L-asparagine has two nitrogens available to react (the N-terminus of its amino acid backbone, and the amide NH_2 of the side chain), it is conceivable that there might be two new gaburedins observed in the culture extract, resulting from the amino nitrogen attacking intermediate **109**, to give gaburedin **171**, and attack of intermediate **109** by the side chain amide, forming gaburedin **172** (these two compounds would be expected to have different retention times on a reverse phase liquid chromatography column).

However, only one is observed (Figure 4.6), and presumably has the structure of gaburedin **171**, as this would be the product of the more nucleophilic N-terminal nitrogen attacking intermediate **109**; the poor nucleophile amide nitrogen would not be expected to be nucleophilic enough to attack carbamoyl adenylate **109**.

Interestingly, there also appeared to be an L-glutamate derived gaburedin **106** present in this culture extract with molecular formulae of $C_{10}H_{17}N_2O_7$ for the $[M+H]^+$ ion (calculated *m/z* = 277.1030, found 277.1030), $C_{10}H_{16}N_2O_7Na$ for the $[M+Na]^+$ ion (calculated *m/z* = 299.0850, found 299.0858), and $C_5H_{10}NO_4$ for the $[M+H-C_5H_7NO_3]^+$ fragment ion (calculated *m/z* = 148.0604, found 148.0602) which were confirmed by UHR-LC-MS. The presence of a gaburedin containing glutamate may be due to adding such high concentrations of L-asparagine to the media resulting in the bacteria producing more L-glutamate *via* primary metabolism. Indeed, L-asparagine is readily converted into L-aspartate, which enters the Kreb's cycle as oxaloacetate, leading to production of 2-oxoglutarate, which is converted into L-glutamate.[176]

Upon feeding the *gbnR::apra* strain with 20 mM L-tyrosine, a new gaburedin **173** was identified in culture extracts of this sample. As expected for a more polar, less hydrophobic aminoacyl moiety (tyrosine in place of phenylalanine) gaburedin **173** has a t_R = 11.2 min, which is earlier than gaburedin A **80** (which contains phenylalanine and elutes at t_R = 12.6 min), and **173** has the same characteristic gaburedin fragmentation pattern (Figure 4.7).

As L-tyrosine has a free hydroxyl group in its side chain, it is possible it might attack the carbamoyl adenylate **109**. However – at physiological pH, the side chain hydroxyl group would not be deprotonated, therefore would not be as available to act as a good

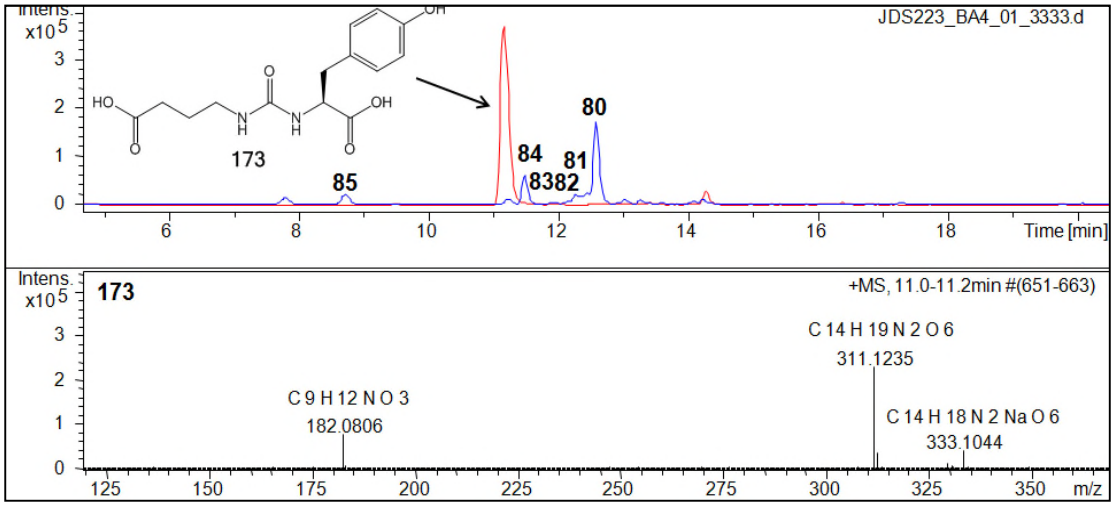


Figure 4.7 – Extracted ion chromatograms for gaburedins A-F (**80-85**) *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) and new gaburedin **173** with *m/z* = 311.0 (red trace) expected for incorporation of L-tyrosine, and UHR mass spectrum of this new gaburedin **173**

nucleophile. Therefore, the only gaburedin expected to be observed is the gaburedin **173** that would arise from attack of the carbamoyl adenylate **109** by the amino NH_2 group.

UHR-LC-MS analysis confirmed the molecular formulae of the gaburedin containing L-tyrosine to be $C_{14}H_{19}N_2O_6$ for the $[M+H]^+$ ion (calculated *m/z* = 311.1238, found 311.1235), $C_{14}H_{18}N_2O_6Na$ for the $[M+Na]^+$ ion (calculated *m/z* = 333.1057, found 333.1043), and $C_9H_{12}NO_3$ for the $[M+H-C_5H_7NO_3]^+$ fragment ion (calculated *m/z* = 182.0812, found 182.0806) corresponding to protonated L-tyrosine. (Figure 4.7).

A similar incorporation experiment was also performed using L-histidine at final concentrations of 1 mM and 10 mM. Analysis of the culture extract revealed that there was no new gaburedin analogue **174** with the *m/z* = 285.1 expected for incorporation of L-histidine, however, the gaburedins **80-85** were still produced. These results are consistent with the results of feeding experiments with other positively charged amino acids L-lysine and L-arginine reported in Section 3.4.1.

4.2 Incorporation experiments with amino acid analogues

4.2.1 Incorporation of amines

In Section 3.4.2, it was reported that feeding with *n*-butylamine led to production of a new gaburedin **156** (shown in Figure 4.8). This observation implies that amines can be incorporated into gaburedins in place of amino acids, suggesting that the carboxylic acid group of amino acids is not required for incorporation into the aminoacyl moiety of gaburedins. This hypothesis was tested further by preparation of SMMS enriched with the valine analogue isobutylamine and leucine analogue isopentylamine at 1 mM final concentrations. The *gbnR::apra* strain was grown on these modified media and the metabolites were extracted. New compounds with *m/z* of 202.9/224.9 and 216.9/238.9 were observed, as would be anticipated for $[M+H]^+$ and $[M+Na]^+$ adducts of gaburedins **175** and **176** containing isobutylamine and isopentylamine. The mass

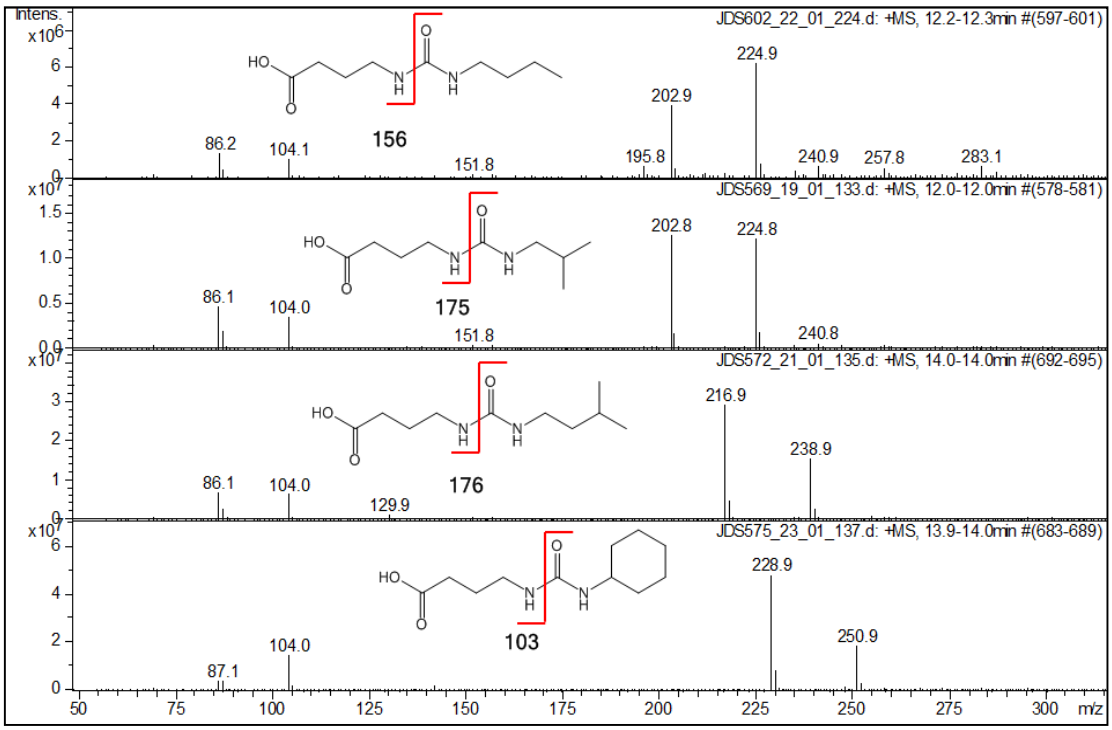


Figure 4.8 – Mass spectra new gaburedins **156**, **175**, **176** and **103** containing *n*-butylamine, isobutylamine, isopentylamine and cyclohexylamine displaying fragments ions at *m/z* = 104.0 – corresponding to GABA – that are common to all four amine-derived gaburedins

spectra of both these new gaburedins both exhibited fragment ions corresponding to protonated GABA at *m/z* = 104.0 (Figure 4.8).

The above observations demonstrate that amines can be incorporated into gaburedins. This is very interesting as GbnB may have the potential to be used as a bio-catalyst to quickly create libraries of such molecules if supplemented with a range of suitable precursor nucleophiles. Therefore, it was hypothesised that a gaburedin analogue **103**, with the same structure as the human epoxide hydrolase inhibitor shown in Figure 2.17 may be produced by the *gbnR::apra* strain grown on media supplemented with cyclohexylamine.[167]

Upon growing the *gbnR::apra* strain on SMMS enriched with 1 mM cyclohexylamine, it was observed that the ethyl acetate extract of this culture contained a new compound **103**, with *m/z* = 229.1 and 251.1, consistent with those expected for the $[M+H]^+$ and $[M+Na]^+$ ions of 4-(3-cyclohexylureido)butanoic acid (Figure 4.8). Molecular formulae were generated for these ions and confirmed to be $C_{11}H_{21}N_2O_3$ for the $[M+H]^+$ ion (calculated *m/z* = 229.1547, found 229.1547) and $C_{11}H_{20}N_2O_3Na$ for the $[M+Na]^+$ ion (calculated *m/z* = 251.1366, found 251.1368). As with the gaburedins derived from amines *n*-butylamine, isobutylamine and isopentylamine, a fragment ion was observed at *m/z* = 104.0, corresponding to protonated GABA (Figure 4.8). The intensity of this peak was too low for molecular formula to be generated in UHR-MS analysis.

When metabolites extracted from the *gbnR::apra* cultures grown on media containing isobutylamine, isopentylamine and cyclohexylamine were resuspended in methanol instead of acetonitrile, it was also observed that there were new gaburedin-like species present in the base peak chromatograms corresponding to *m/z* 217.0 for an

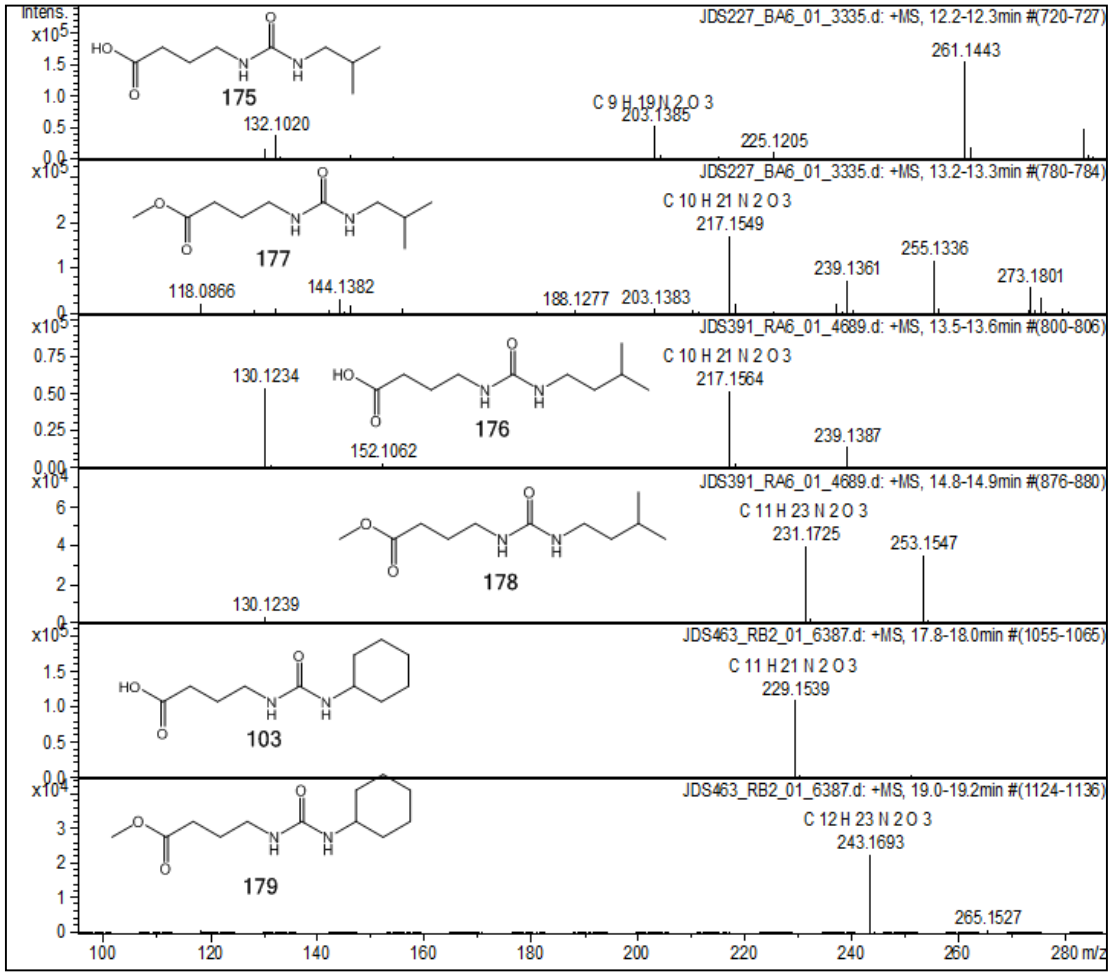


Figure 4.9 – Mass spectra of new gaburedins containing isobutylamine, isopentylamine and cyclohexylamine (panels a, c and e) and the species corresponding to $[M+14]^+$ ions for additional metabolites present in the extracts upon feeding with these precursor molecules (panels b, d and f)

isobutylamine derivative **177**, *m/z* = 231.0 for an isopentylamine derivative **178** and *m/z* = 243.0 for a cyclohexylamine derivative **179**, respectively (Figure 4.9). These metabolites correspond to methylated gaburedins **175**, **176** and **103**, respectively. As these compounds were only observed upon resuspension with acidic methanol, it was concluded that they are artefacts created as a result of using methanol to redissolve metabolites and acidic methanol as the eluent in LC-MS. These molecules are therefore not produced during the biosynthesis of gaburedins. They likely arise due to attack of the carboxylic acid terminus of the GABA moiety by methanol under acidic conditions.

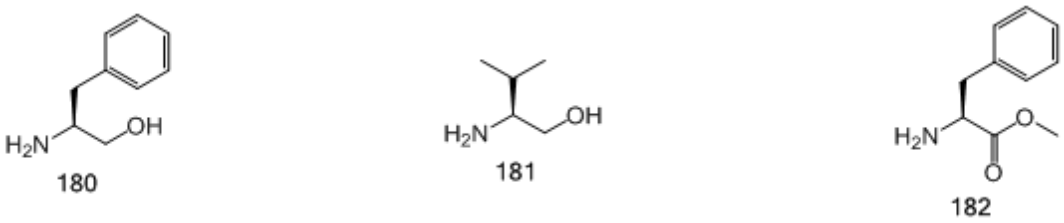


Figure 4.10 – Structures of amino acid analogues fed to the *gbnR::apra* mutant to probe diversity of aminoacyl compounds with different C-terminal groups which can be incorporated into gaburedins

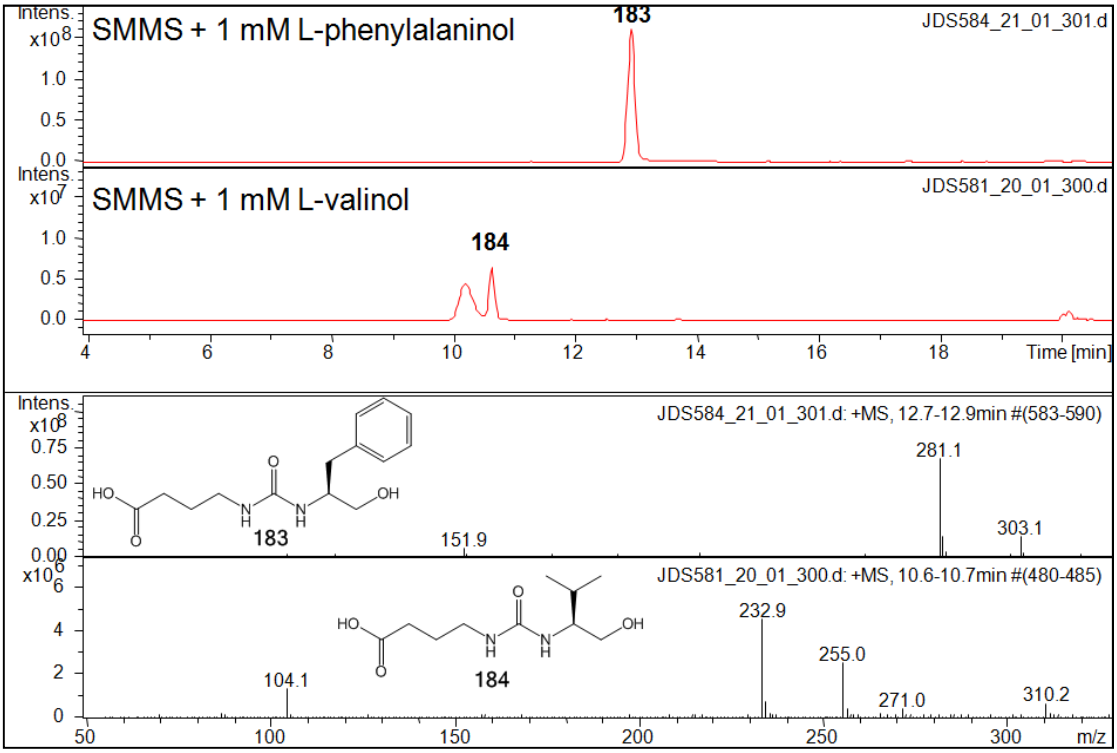


Figure 4.11 – Extracted ion chromatograms for *m/z* = 281.0 (upper panel) and 233.0 (lower panel) and mass spectra of the new gaburedins **183** and **184** that are produced by the *gbnR* mutant when grown on SMMS + 1 mM L-phenylalaninol (upper panel) and 1 mM L-valinol (lower panel)

4.2.2 Incorporation experiments with aminoalcohols and aminoacyl esters

To further investigate the diversity of amine derivatives which can be incorporated as amine/aminoacyl moieties in gaburedin biosynthesis (green in Figure 4.1), the *gbnR::apra* strain was grown on SMMS supplemented with aminoalcohols L-phenylalaninol **180**, L-valinol **181** and the aminoacyl ester L-phenylalanine-*O*-methyl ester **182** (Figure 4.10). Upon growing the *gbnR::apra* strain on SMMS supplemented with 1 mM L-phenylalaninol and L-valinol, new gaburedins **183** and **184** were

observed, which have the *m/z* values expected for $[M+H]^+$ ions of 281.1 and 232.9 as would be expected for incorporation of L-phenylalaninol and L-valinol, respectively. Also present were the fragment ions *m/z* = 151.9 and 104.1 that would be expected for protonated L-phenylalaninol and L-valinol (Figure 4.11).

When the *gbnR::apra* strain was grown on SMMS supplemented with L-phenylalanine-*O*-methyl ester.HCl (10 mM final concentration), two new gaburedin-like molecules **185** and **186** were observed (Figure 4.12). Compound **185** has *m/z* of 309.1 for its $[M+H]^+$ ion, with a mass spectrum containing the expected mass loss of 129 Da to leave a fragment at *m/z* = 180.1, as would be anticipated for a gaburedin containing protonated L-phenylalanine-*O*-methyl ester. A second molecule **186** with *m/z* of 309.1 was also observed, however the mass spectrum of this peak featured a

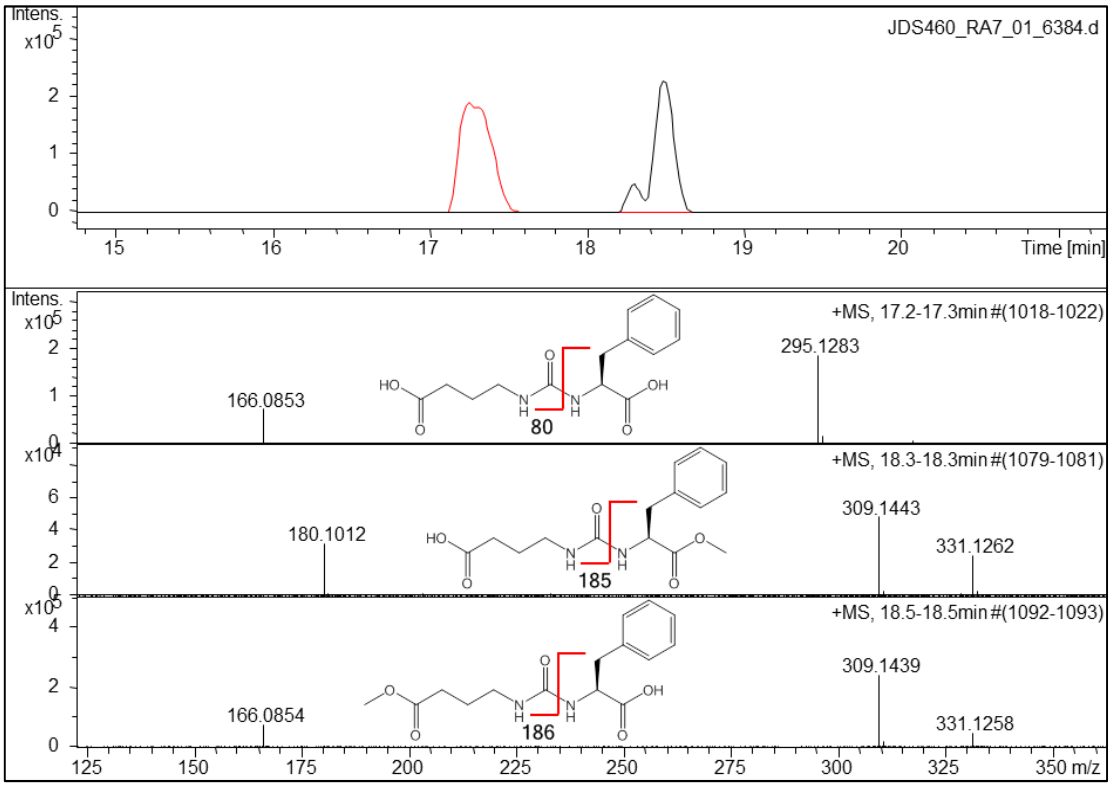


Figure 4.12 – Extracted ion chromatograms for gaburedin A *m/z* = 294.9 (red trace) and *m/z* = 309.0 (black trace) expected for incorporation of L-phenylalanine-*O*-Me ester in the metabolites extracted after growing the *gbnR* mutant on SMMS + 10 mM L-phenylalanine-*O*-methyl ester. Mass spectra of gaburedins **80**, **185** and **186** are also shown

fragment ion at *m/z* = 166.1, corresponding to protonated phenylalanine, implying that the methyl group lost is part of the ureidyl-GABA fragment.

Therefore this experiment was repeated with acetonitrile, there were no metabolites with *m/z* = 309.0 present in this culture extract, suggesting that the methyl group of L-phenylalanine-*O*-methyl ester is not preserved throughout the biosynthesis of gaburedins. This result implies that both gaburedins **185** and **186** shown in Figure 4.12 are artefacts of the extraction/analysis process. This would result also account for overproduction of gaburedin A **80**, as there may be an increased intracellular concentration of L-phenylalanine due to hydrolysis of the L-phenylalanine-*O*-methyl ester.[178]

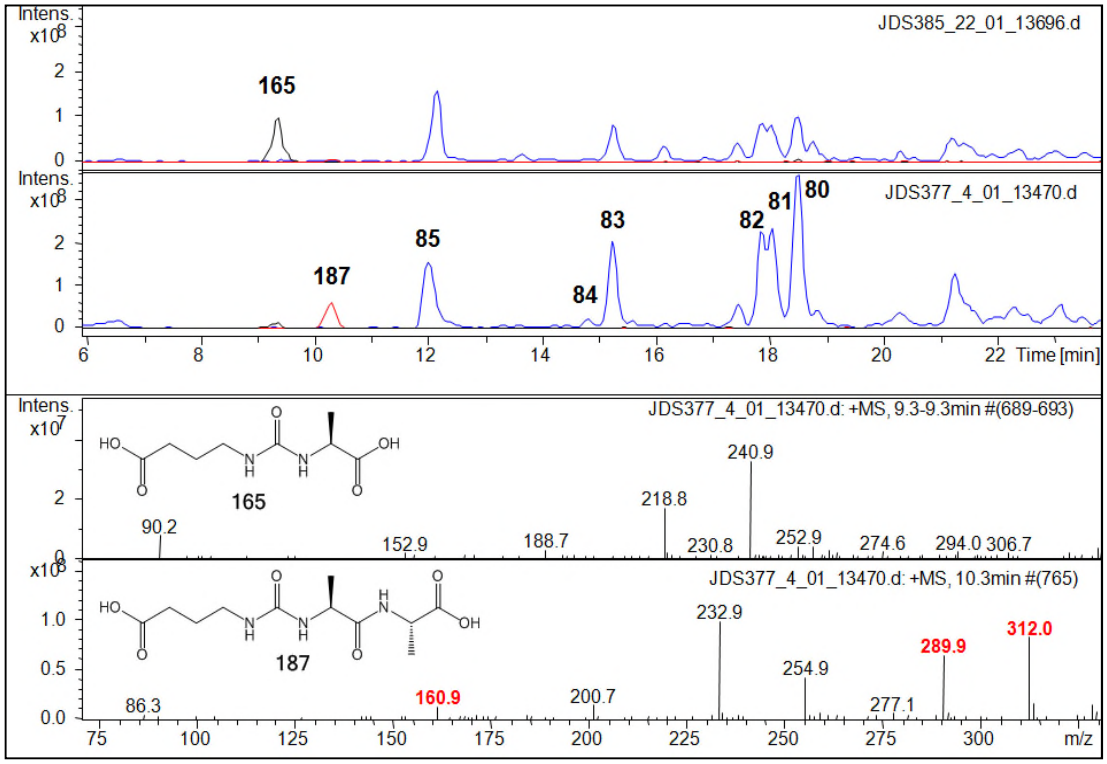


Figure 4.13 – Extracted ion chromatograms gaburedins A-F (**80**-**85**) *m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9 (blue traces) and *m/z* = 289.8 (red traces) for L-ala-ala incorporation into a new gaburedin **187** and *m/z* = 218.8 (black traces) for gaburedin **165** arising from incorporation of L-alanine. Mass spectra of **165** and **187** also shown, with the molecular and fragment ions for **187** highlighted in red

4.2.3 Incorporation of a dipeptide

After having established that a variety of amino acids amines can be incorporated into gaburedins, SMMS enriched with the dipeptide L-ala-ala was prepared. LC-MS analysis of culture extracts from the *gbnR::apra* strain grown on 10 mM L-ala-ala revealed the presence of a new gaburedin-like compound **187** with the expected *m/z* = 289.9 and a fragment ion at 160.9, consistent with the expected *m/z* values for protonated L-ala-ala and also having a mass difference corresponding to the conserved $C_5H_7NO_3$ ureidyl-GABA fragment (Figure 4.13). The gaburedin **165** containing an aminoacyl L-alanine moiety was also observed in this culture extract, implying that some L-ala-ala was hydrolysed either during autoclaving or by the bacteria.

4.3 Discussion of gaburedin biosynthesis

4.3.1 Effect of altering SMMS composition on production of gaburedins A-F

Thus far, there has been discussion of the new gaburedin-like compounds produced by the *gbnR::apra* strain when grown on supplemented SMMS, however details of the effect of the relative amounts of gaburedins A-F (**80**-**85**) produced under these growth conditions has not been addressed in every case. When the *gbnR::apra* mutant is grown on media supplemented with amino acids L-glutamic acid, L-alanine, glycine, L-serine, L-threonine, L-tyrosine, L-proline, L-asparagine, L-tyrosine, L-ornithine there were new gaburedins produced in addition to **80**-**85** produced on unmodified SMMS, with each of these amino acids being incorporated into the aminoacyl moiety of gaburedins. New gaburedins produced are summarised in Table 4.1 and Figure 4.14. In other growth conditions (on SMMS supplemented with 1 mM L-lysine, L-arginine, L-histidine, L-aspartic acid and L-glutamine), there was no dramatic reduction in the yields of **80**-**85**, however there were no new gaburedins produced corresponding to

Table 4.1 – Gaburedins identified and overproduced throughout this study

aminoacyl precursor	*compound(s) overproduced*	m/z $[M+H]^+$	m/z $[M+Na]^+$	m/z *[fragment ion]*$^+$
L-phenylalanine	**80**	295.1298	317.1114	166.0872
D-phenylalanine	**95**	295.0	317.0	166.1
L-leucine	**81**	261.1141	283.1260	132.1019
d_{10}-L-leucine	d_{10}-**81**	271.1	293.1	142.0
L-isoleucine	**82**	261.1440	283.1260	132.1018
d_{10}-L-isoleucine	d_{10}-**82**	271.1	293.1	142.0
L-valine	**83**	247.1295	269.1114	118.0874
d_8-DL-valine	**d_8-DL-83**	255.1802	277.1617	126.1372
D-valine	**95**	247.0	269.0	117.9
L-methionine	**84**	279.1015	301.0834	150.0591
N-acetyl-L-cysteine	**85**	293.0809	315.0622	164.0382, 122.0281
L-glutamic acid	**106, 107**	277.1030	299.0858	148.0602
GABA	**107**	233.1134	255.0949	104.0714
L-ornithine	**130**	244.1309	266.1116	115.0873
L-lysine	none	-	-	-
L-glutamine	none	-	-	-
L-aspartic acid	none	-	-	-
L-norvaline	**138**			
n-butylamine	**156**	202.9	224.9	104.1
methyl 4-aminobutyrate	**164**	246.9	269.9	117.9
L-alanine	**165**	219.4	241.2	90.2
L-proline	**166**	244.9	266.9	-
glycine	**167**	204.8	226.8	76.2
L-serine	**168**	235.0928	257.0747	106.0506
L-threonine	**169**	249.1072	271.0890	120.0651
L-cysteine	**85**	-	-	-
L-asparagine	**171**	262.1030	284.0852	133.0607
L-histidine	none	-	-	-
L-tyrosine	**173**	311.1235	333.1043	182.0806
isobutylamine	**175**	203.1385	225.1205	104.0
isopentylamine	**176**	217.1564	239.1387	104.0
cyclohexylamine	**103**	229.1539	250.9	104.0
L-phenylalaninol	**183**	281.1	303.1	151.9
L-valinol	**184**	232.9	255.0	104.1
L-phenylalanine-*O*-methyl ester	**80**	-	-	-
L-ala-ala	**187, 165**	289.8	311.8	160.9

these amino acids being incorporated. However, it was noted that at 20 mM final concentration of L-lysine, production of gaburedins **80-85** was dramatically reduced.

In the incorporation experiments with isobutylamine, isopentylamine and cyclohexylamine, there was reduced production of gaburedins **80-85**, even at 1 mM final concentration of each of these amines. This may be because these amines are more

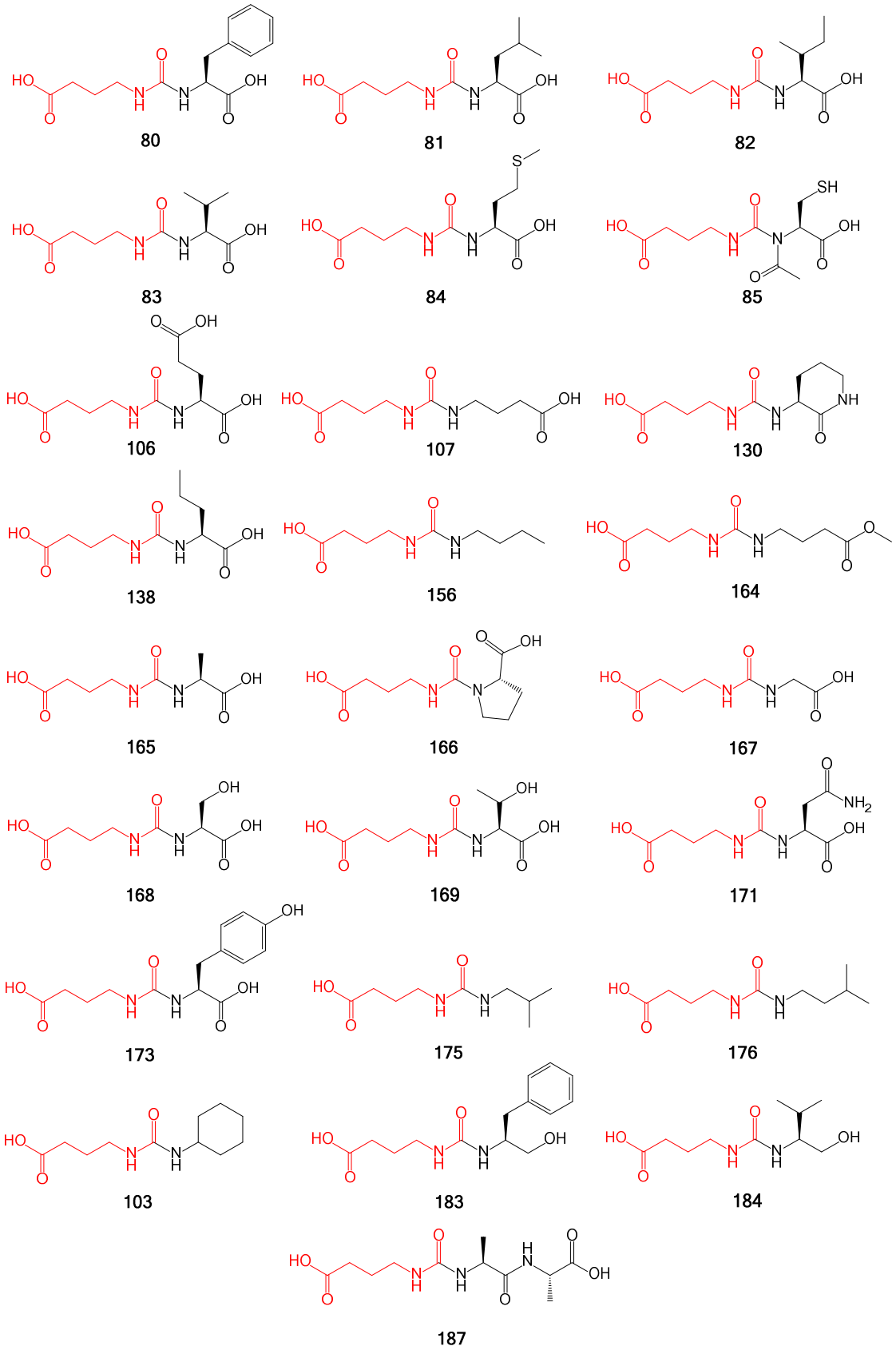


Figure 4.14 – Library of gaburedins produced during this study

nucleophilic than amino acids and will attack carbamoyl adenylate **109** at a faster rate than the amino acids that attack this intermediate to furnish gaburedins A-F **80-85**.

Therefore gaburedins are still being biosynthesised, the native aminoacyl moieties of gaburedins **80-85** are simply substituted for the amines added to the growth media.

As discussed in Section 2.2.5, these feeding experiments appear to suggest that the incorporation of phenylalanine, leucine, isoleucine, valine and methionine into the natural gaburedins **80-84** is perhaps mostly due to the intracellular availability of these amino acids at the time when the carbamoyl adenylate **109** is produced. Therefore, upon addition of a large excess of other nucleophiles into the growth media, there is incorporation of these nucleophiles into gaburedin biosynthesis (assuming that these nucleophiles are taken up by the cell and the nucleophilic attack of carbamoyl adenylate **109** occurs intracellularly).

In the cases of L-cysteine, L-phenylalaninol and L-valinol, it would appear that gaburedin production is in fact abolished as the concentrations of these compounds in the media are increased to 10 mM. However, addition of 1 mM L-cysteine does not yield new gaburedins, whereas L-phenyalaninol and L-valinol are both incorporated at concentrations of 1 mM. Therefore, the observation that production is abolished at high concentrations could be due to addition of reactive species such as cysteine and aminoalcohols exerting stress on the cells, resulting in gaburedin biosynthesis being down-regulated (so none of the amines, aminoalcohols or potential derivatives can be incorporated into gaburedin production). This hypothesis could be tested by RT-PCR to monitor gene expression of *gbnB* in these different growth conditions.

The production of such a diverse range of gaburedins (the six gaburedins produced by the *gbnR::apra* mutant on unmodified SMMS, and the nineteen additional ones produced on SMMS supplemented with different amino acids and amines) provide

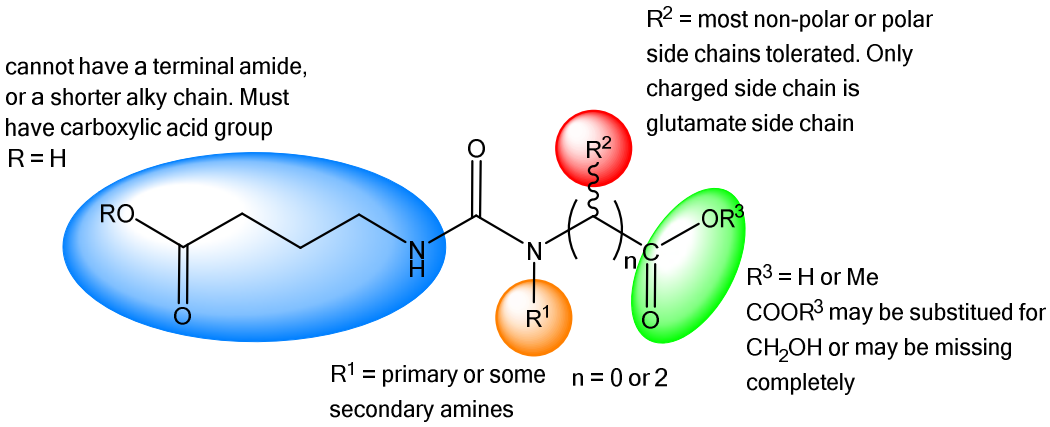


Figure 4.15 – Schematic to illustrate the substrate tolerance of GbnAB leading to conserved GABA moiety (blue) and diversity of different amino acids and amino acid analogues that can be incorporated into gaburedins *in vivo* (orange, red and green)

more evidence for the hypothesis that the incorporation of the aminoacyl/amine moiety of gaburedins is not enzyme-catalysed, as this would require unprecedented substrate specificity. The presence of such a reactive intermediate (carbamoyl adenylate **109**) could account for the diverse range of nucleophiles being able to be incorporated into gaburedin biosynthesis, as the *O*-AMP of **109** is a very good leaving group. Figure 4.15 illustrates a summary of the feeding experiments discussed throughout Chapters 3 and 4 with respect to the substrates that can react with intermediate **109** to furnish the gaburedin natural products.

4.3.2 The acetylation observation

In the presence of amines isobutylamine, isopentylamine, cyclohexylamine and aminoalcohols L-phenylalaninol and L-valinol, acetylation of the precursor molecules fed to the *gbnR::apra* mutant was observed. This is a well documented phenomenon in bacteria[177] however in this case, the observation that acetylation of amine at concentrations >10 mM appears to be preferential to the incorporation of these amines into gaburedins. Conversely, the presence of acetylated phenylalanine is observed in culture extracts of *S. venezuelae* in which no gaburedin A (**80**) is observed (this

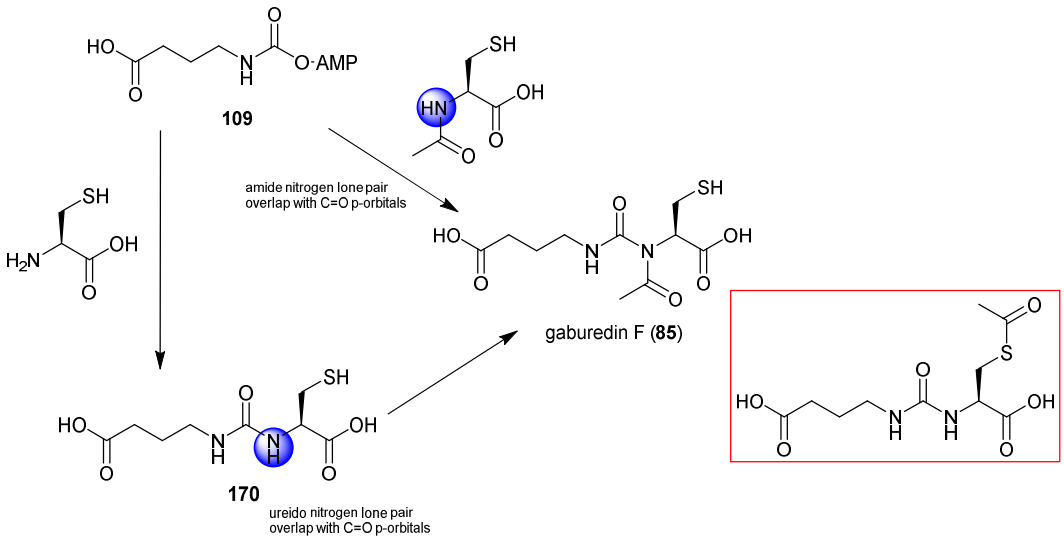


Figure 4.16 – Schematic to show incorporation of *N*-acetyl-L-cysteine into the proposed structure for gaburedin F (**85**) either directly or via the cysteine derived gaburedin **170**. Both routes would require very poor nucleophiles attacking either the intermediate **109**, or acetylation occurring at the poorly nucleophilic ureido nitrogen of **170**. The possible structure of gaburedin F if derived from *S*-acetylcysteine is highlighted

compound with *m/z* = 208.0 is observed in wild-type culture extracts in addition to *gbnR::apra* culture extracts in which gaburedin production has been abolished, such as when feeding high concentrations of cysteine or aminoalcohols).

The incorporation of *N*-acetylcysteine into gaburedins to overproduce gaburedin F (**85**) would appear to indicate that *N*-acetylated amino acids can be incorporated into gaburedins. However, it is unlikely that the amide nitrogen in *N*-acetylcysteine would be nucleophilic enough to attack intermediate **109** to give **85**. Alternatively, it is unlikely that a gaburedin containing cysteine would be acetylated, as this would require acetylation of a very poorly nucleophilic ureido nitrogen atom (Figure 4.16). One could therefore argue that the gaburedin **85** requires further analysis in order to confirm its structure. Indeed, attempts to purify **85** by HPLC have proved unsuccessful so far. Once purified, a sample of gaburedin F could be treated with Ellman's reagent which would react with free thiols, staining yellow.[179] If no reaction occurs, this would indicate that the cysteine in gaburedin **85** is in fact *S*-acetylated.

5. Reconstitution of *gbnAB* system from *Streptomyces venezuelae*

The *in vivo* feeding experiments discussed in Chapter 3 and 4 only permits an indirect observation of the biosynthetic pathway to gaburedins in the native host *S. venezuelae*. In addition, specific (putative) precursors added to the extracellular medium might not enter the cell, or might have unexpected effects on global metabolism.

By contrast, *in vitro* assays studies of the proposed biosynthetic enzymes GbnA and GbnB would allow more refined structure-activity relationships to be established. In particular, addition of GABA or GABA derivatives to GbnB could be used to probe whether *N*-activated GABA is the true substrate for this enzyme. Such assays would also allow the diversity of gaburedin analogues to be extended to substrates which could be toxic to – or metabolised by – the native *S. venezuelae* host strain.

Furthermore, expression of *gbnABC* in a heterologous host would allow further insight into whether *gbnABC* alone are required for gaburedin biosynthesis. As demonstrated in Section 3.2.2, *gbnB* is essential for gaburedin biosynthesis. However, whether or not *gbnA* and *gbnC* are also essential for gaburedin biosynthesis remains to be demonstrated. By comparison of the metabolites produced by strains carrying *gbnA* and *gbnB*, or *gbnA*, *gbnB* and *gbnC*, the role of *gbnC* can be deduced. Cloning of individual – or combinations of – genes into an inducible expression vector in *E. coli* offers a fast, convenient method to achieve this, providing that the enzymes are active in this heterologous host.

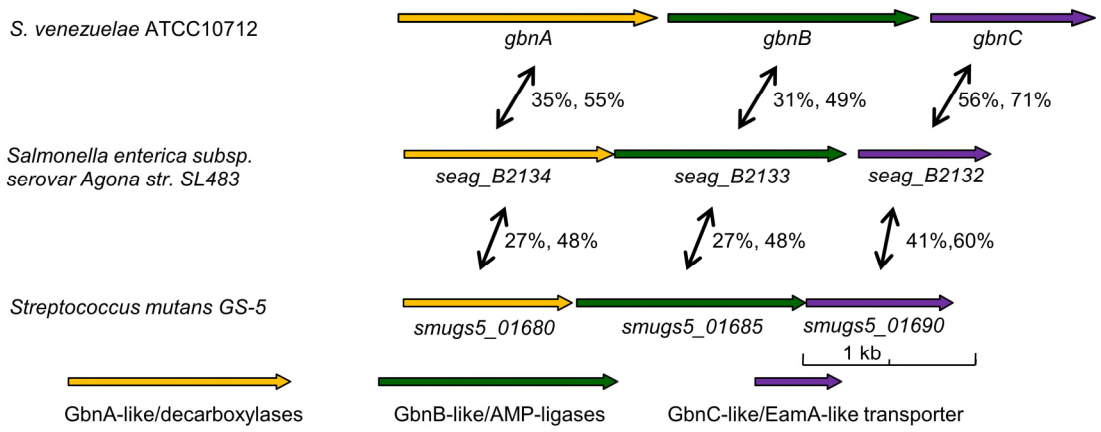


Figure 5.1 – *gbnABC* clusters from *Streptomyces venezuelae*, *Salmonella enterica* and *Streptococcus mutans* GS-5. Only the metabolic products of the *gbnABC* system in *S. venezuelae* have been reported so far

Once it has been determined that heterologous expression of *S. venezuelae gbnABC* in *E. coli* confers the gaburedin-producing chemotype, this approach can be extended to include some of the other *gbnABC* systems identified in Section 3.1.1, to determine the nature of the gaburedin-like natural products produced by these systems. Two of these *gbnABC* clusters – one from the Gram-positive *Streptococcus mutans* GS-5 and one from the Gram-negative *Salmonella enterica* – have been selected for this purpose (Figure 5.1). As a proof-of-concept study, it is important to first reconstitute the *gbnAB(C)* system from *S. venezuelae* to establish if gaburedins can be produced when *gbnAB* or *gbnABC* are overexpressed in *E. coli*, before extending this approach to orthologous *gbnABC* systems.

5.1 Cloning *gbnAB (sven_4179 and sven_4180)* from *S. venezuelae*

Initially, primers were designed to amplify the genes *gbnA* (*sven_4179*) and *gbnB* (*sven_4180*) separately for cloning into an expression vector, allowing each individual biosynthetic enzyme to be studied *in vitro* in order to gain greater insight into the precise function of each of the putative biosynthetic enzymes GbnA and GbnB.

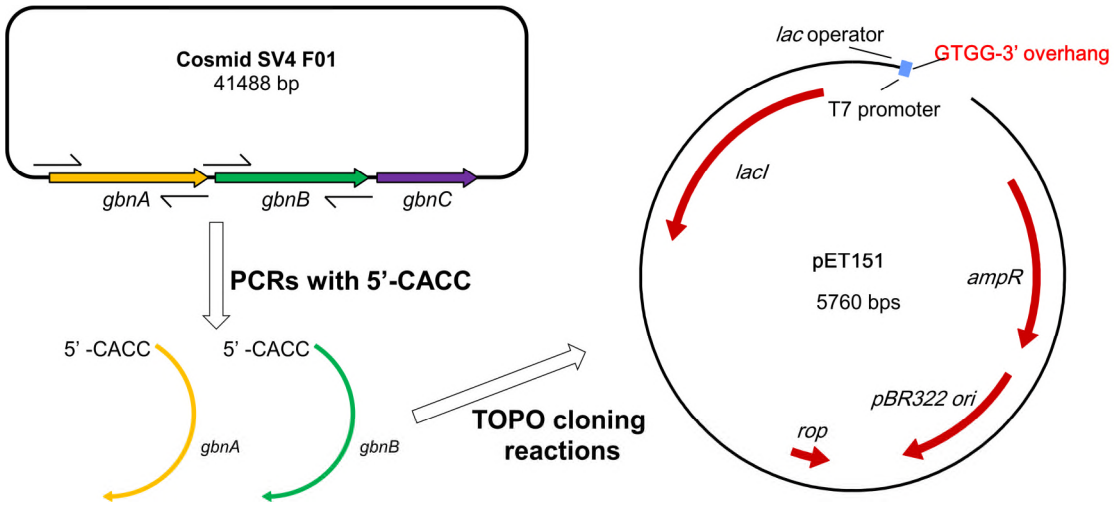


Figure 5.2 – Schematic to show the steps in creating pET151-derived plasmids using the *gbnA* and *gbnB* genes from *S. venezuelae*. Primers were designed to include a 5'-CACC overhang to allow topoisomerase-directed cloning of *gbnA* and *gbnB* into pET151

Different primer pairs were designed in order to amplify the genes *gbnA* and *gbnB* for cloning into pET151, a plasmid in which the inserted gene is under the control of the T7 promoter, and expression is inducible using IPTG via the *lac* operon. The T7 promoter in pET151 lies upstream of the *lac* operator, to which the *lac* repressor LacI binds, therefore repressing transcription of the downstream gene(s). The expression of the downstream gene of interest can be induced by addition of the allolactose analogue isopropyl β-D-1-thiogalactopyranoside (ITPG), which binds to LacI, causing it to be released from its DNA target. This results in expression of the downstream gene of interest, as the *lac* operator is no longer being repressed.

Therefore, the gene of interest contained within pET151-derived plasmids can be overexpressed – leading to overproduction of the desired protein – when the plasmid is introduced into *E. coli* BL21star, and IPTG added to the culture broth. *E. coli* BL21star is a strain which overproduces the specific T7 polymerase when exposed to IPTG,

therefore allowing overproduction of the T7 polymerase to be coordinated with derepression of expression of the *lac* operator in the pET151-derived construct.

The addition of a 5'CACC overhang is required for this cloning, so that the topoisomerase in the TOPO cloning kit can direct the insertion of the gene into the correct site in pET151, which has a single stranded overhang corresponding to GTGG-3' (Figure 5.2). This sequence is complementary to the 5'–CACC overhang included upstream of the desired insert, therefore allowing the insert to be incorporated into pET151 when it is recircularised.

Initial attempts to amplify *gbnA* using the primers A/A' and to amplify *gbnB* using primers D/D' (sequences of primers used for TOPO cloning can be found in Section 8.5) were unsuccessful, as no PCR products were formed. The template DNA was varied (different preparations of SV4 F01 cosmid and a *ΔgbnR* derivative of SV4 F01 were both used). Some non-specific binding was observed, therefore primer pairs B/B' and E/E' were ordered in which the regions overlapping with the 5' and 3' ends of each gene were extended.

The SV4 F01 templates used were therefore re-purified – both by use of a mini-prep kit and by phenol-chloroform extraction. It was found that efficient recovery of cosmid could only be achieved if the elution buffer was pre-heated to 60 °C and the DNA pre-incubated on the centrifuge filter prior to elution. The purified SV4 F01 was then tested by restriction digest with *Sac*I, and by amplification of the *gbnR* gene with specific screening primers F/F' in Table 8.1.

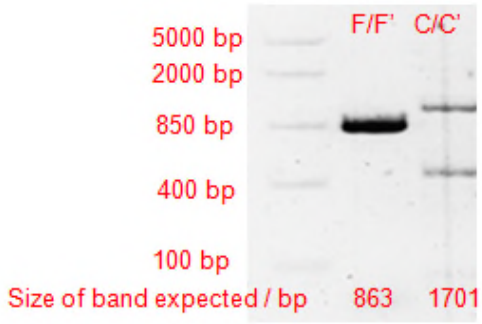


Figure 5.3 – 1% agarose gel of PCR products obtained for DNA regions amplified using control screening primers F/F' specific for *gbnR*, and TOPO cloning primers C/C' specific for *gbnA*. The bands for the PCR products obtained with primers C/C' were shorter than expected. Colours have been inverted for clarity

More SV4 F01 DNA was prepared, checked by restriction digest for the correct pattern with *Sac*I, and used as the templates for amplification by PCR using the screening primers I/I' for *gbnA*, and H/H' for *gbnB*. Control reactions with the *gbnR* screening primers F/F' were also performed. This time, PCR products were obtained for the reactions using the F/F' primer pair for *gbnR*, and H/H' pair of *gbnB* screening primers, yet no PCR products were obtained for the I/I' *gbnA* screening primers.

Shorter *gbnA* primers were also designed (C/C') and used for PCRs with the same SV4 F01 template DNA. This time, there were two PCR products (Figure 5.3), one at approximately 500 bp, and one at approximately 1200 bp. The 1200 bp product was purified and sent for sequencing with the primers C/C'. The results showed that *gbnA* had been amplified, although it would appear that primer C binds approximately 400 bp downstream of the start codon of *gbnA*. Attempts were also made to amplify *gbnB* with the all combinations of D/D' and E/E' primer pairs, with only a weak band at approximately 1600 bp being amplified. This band was too weak for any DNA to be recovered after gel extraction.

5.2 Cloning of synthetic *S. venezuelae gbnB* gene and orthologues

5.2.1 Synthetic genes for heterologous expression in *E. coli*

To get optimal chance of gene expression in *E. coli*, synthetic genes were purchased from Life Technologies. These genes were based on the native *S. venezuelae* GbnB, as well as the GbnB homologues *Salmonella enterica* SeAg_B2133 and *Streptococcus mutans* GS-5 SMUGS5_01685 amino acid sequences. However, these synthetic genes were codon optimised to reflect *E. coli* codon usage to improve gene expression and protein production levels in this heterologous host (sequences of the synthetic genes can be found in Appendix 3). The genes were synthesised and cloned into the pUC-derived pMK-RQ vector carrying a kanamycin resistance gene (Figure 5.4).

Specific primers were designed to amplify these *E. coli* codon optimised versions of *S. venezuelae gbnB* (primers J/J'), as well as the *gbnB* homologues *Salmonella enterica b2133* (primers K/K') and *Streptococcus mutans* GS-5 *smugs5_01685* (primers L/L') from their respective pMK-RQ vectors. As discussed in the previous section, the 5' CACC overhang required for TOPO cloning were also included in the primers designed. The PCR products obtained using a high fidelity polymerase were purified and cloned into pET151. Full experimental details can be found in Section 8.5.

Positive colonies containing the new plasmids pJDS06 (carrying synthetic *sven_gbnB*), pJDS02 (carrying synthetic *seag_gbnB*) and pJDS04 (carrying synthetic *smug_gbnB*) after TOPO cloning reaction were identified by selection on ampicillin plates, followed by restriction digests with *Nde*I for pJDS06 and pJDS02 and digestion with *Pst*I for pJDS004 (Figure 5.4). These enzymes were selected as they cut once in the insert and once in the backbone. Clearly different restriction patterns were therefore expected for

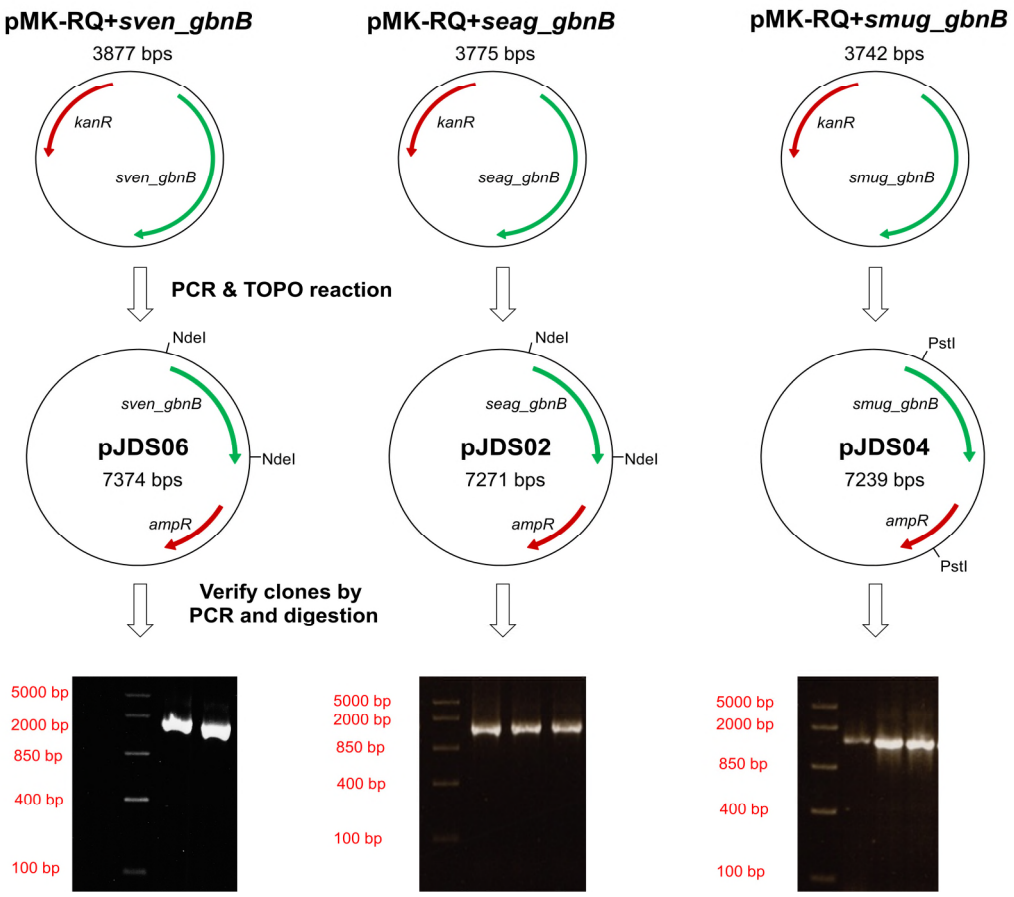


Figure 5.4 – Diagram to illustrate TOPO cloning of synthetic *S. venezuelae gbnB, Salmonella enterica seag_b2133* and *Streptococcus mutans* GS-5 *smugs5_01685* into pET151. Restriction digests using the enzymes *Nde*I and *Pst*I were used to confirm the nature of plasmids pJDS06 and pJDS02 and pJDS04, respectively. These plasmids were also analysed by PCR using primer pairs J/J', K/K' and L/L' and the inserted genes were sequenced using the T7 primer pair

the empty vector compared that the vectors in which *sven_gbnB*, *seag_gbnB* or *smug_gbnB* had been inserted. Insertion of the correct DNA sequence into each plasmid was also checked by PCR amplification of the gene of interest using the primer pairs J/J', K/K' and L/L' for pJDS06, pJDS02 and pJDS04, respectively.

Plasmids that gave both PCR products of the expected sizes and restriction digest patterns consistent with the insertion of the genes of interest were sent for sequencing by GATC Biotech, with amplification using the T7 primer pair (see Section 8.5). Sequencing data confirmed that pJDS06, pJDS02 and pJDS04 contained the intact sequences of synthetic *sven_gbnB*, *seag_gbnB* and *smug_gbnB*, respectively.

5.2.2 Metabolic profiling of engineered *E. coli* strains carrying *gbnB* and orthologues

The effect of introduction of *gbnB* and *gbnB*-related genes on the metabolite profiles of *E. coli* were investigated. All three *E. coli* BL21star strains carrying the plasmids pJDS02, pJDS04 and pJDS06 containing synthetic *gbnB* gene and orthologues were grown in 50 mL LB containing 10 mM GABA (the proposed precursor for GbnB) and gene expression was induced with IPTG. Cells were pelleted and ethyl acetate extracts of the culture supernatants were prepared.

If the biosynthetic pathway to gaburedins presented in Scheme 3.2 is correct, and if GbnB is functional in the heterologous host *E. coli*, it is possible that GbnB could turn over GABA (or its derivative) and produce gaburedins, yet they may not be exported because GbnC is not present in *E. coli*. Therefore, cells were lysed and ethyl acetate

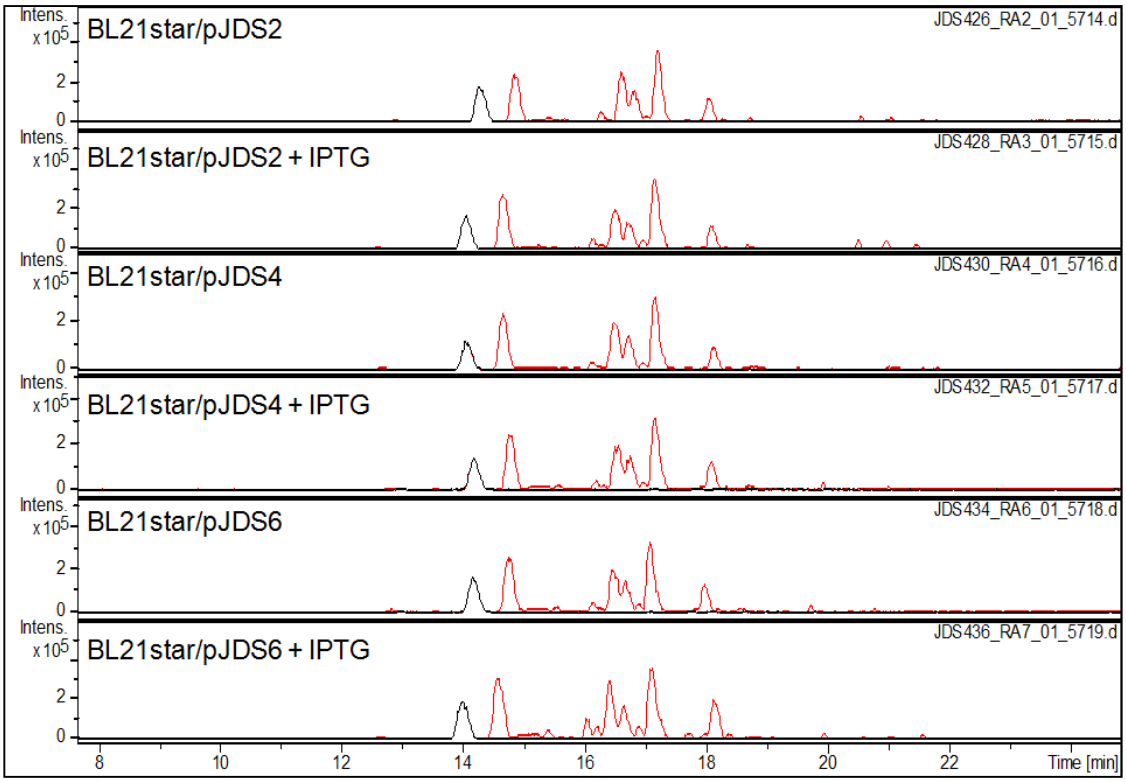


Figure 5.5 – Metabolites extracted from culture supernatants of *E. coli*/pJDS02, pJDS04 and pJDS06 grown in LB supplemented with 10 mM GABA. Base peak chromatograms (red traces) and extracted ion chromatograms for gaburedins A-F (**80-85**) (*m/z* = 247.0, 261.0, 279.0, 292.9 and 294.9) are also shown (black traces). The large peak at t = 14.0 min has *m/z* = 261.0, but is unrelated to gaburedins

extracts of the cell lysates were also performed.

As shown in Figure 5.5, no new metabolites were produced by *E. coli* carrying pJDS02, pJDS04 or pJDS06 upon induction with IPTG, indicating that if GbnB (and orthologues) are functional, no gaburedins are produced upon feeding *E. coli* BL21star carrying plasmids pJDS02, pJDS04 or pJDS06 with GABA. Cell permeability to GABA may be a problem, also the activation of the amino group of GABA that is predicted to produce the substrate for GbnB may not be occur without GbnA, or spontaneous reaction of GABA with CO_2 may not occur in *E. coli*. It is therefore possible that both GbnA and GbnB are required to form a soluble/functional enzyme complex.

5.2.3 Overproduction of recombinant enzymes in *E. coli* BL21star

LB cultures of *E. coli* BL21star carrying each of the plasmids pJDS002, pJDS04 and pJDS06 were prepared and gene expression induced with IPTG as described in Section 8.3.13. In parallel, control experiments without IPTG induction were performed with all three strains.

After growing for 16 h at 15 °C, cells were pelleted and lysed. SDS-PAGE gels were run of soluble and insoluble protein fractions, suggesting that all three GbnB and GbnB-like proteins were present in the insoluble fractions. For example, a band was expected at 60.2 kDa for the protein SeAg_B2133 overproduced by overexpression of *seag_b2133* in pJDS02. Indeed, a band at approximately 60 kDa was observed in the insoluble fraction, but not in the soluble fraction (Figure 5.6).

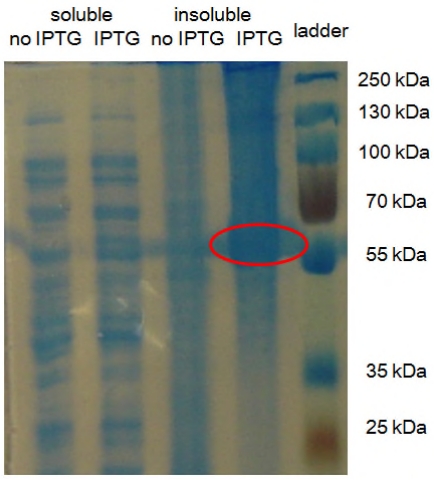


Figure 5.6 – 10% SDS-PAGE gels to show comparison of soluble and insoluble proteins extracted from *E. coli*/pJDS02 carrying synthetic *seag_gbnB*. The only lane with an intense band for a protein of the expected mass of 60.2 kDa is the lane loaded with the insoluble protein fraction

The observation that all three proteins were insoluble therefore reduces the possibility of developing an *in vitro* assay to prove their substrate specificity. However, GbnB being insoluble in *E. coli* does not necessarily mean that it is not functional in this host. The lack of gaburedins being produced may be due to the other enzymes GbnA and GbnC being absent in this host. Indeed, as shown in Section 3.1.1, several of the GbnB and GbnC orthologues appear to be encoded for by a single gene, suggesting that GbnB and GbnC may function as a complex in the organisms in which GbnB and GbnC are encoded by discreet genes. Therefore, future attempts at reconstituting the GbnABC biosynthetic pathway focussed on construction of expression vectors to try and incorporate both biosynthetic genes *gbnA* and *gbnB*, and on attempts to incorporate all three genes *gbnA*, *gbnB* and *gbnC* into a single expression system.

5.3 Alternative cloning strategies for reconstitution of *gbnABC* in *E. coli*

5.3.1 Sub-cloning synthetic *gbnA* into pJDS06 and synthetic *gbnB* into pJDS05

As GbnB enzymes from *Streptomyces venezuelae*, *Salmonella enterica* and *Streptococcus mutans* all appear to be insoluble, attempts were made to create a vector which could be used for protein overproduction of synthetic *S. venezuelae* GbnA and GbnB together. This was attempted with the *S. venezuelae* system in order to first verify whether or not overexpression of both GbnA and GbnB together would lead to the gaburedin-producing chemotype. If this proved to be the case, then this approach could be expanded to include the GbnAB systems from *Salmonella* and *Streptococcus*, in order to determine the metabolic products of the orthologous GbnAB enzymes in those organisms.

Initially, a pET151 derived plasmid containing synthetic *S. venezuelae gbnA* (*sven_4179*) was constructed. Primers M/M' were used to amplify *synthetic gbnA* from a pMK-RQ template (as in Section 5.2.1, the sequence of this synthetic gene was based on the protein sequence of *S. venezuelae* GbnA but with codons altered to reflect codon usage in *E. coli*, see Appendix 3). The plasmid pJDS05 purified from ampicillin resistant colonies was verified as having the correct gene sequence inserted by restriction digests with *Sph*I, amplification of the gene of interest by PCR using primers M/M' and by sequencing of the DNA region of interest with the T7 primer pair.

After sequence verification, the plasmids pJDS05 and pJDS06 were each digested using *Sac*I in order to create linearised vectors into which synthetic *gbnA* synthetic *gbnB* could be sub-cloned. Therefore, new primer pairs N/N' and O/O' were designed

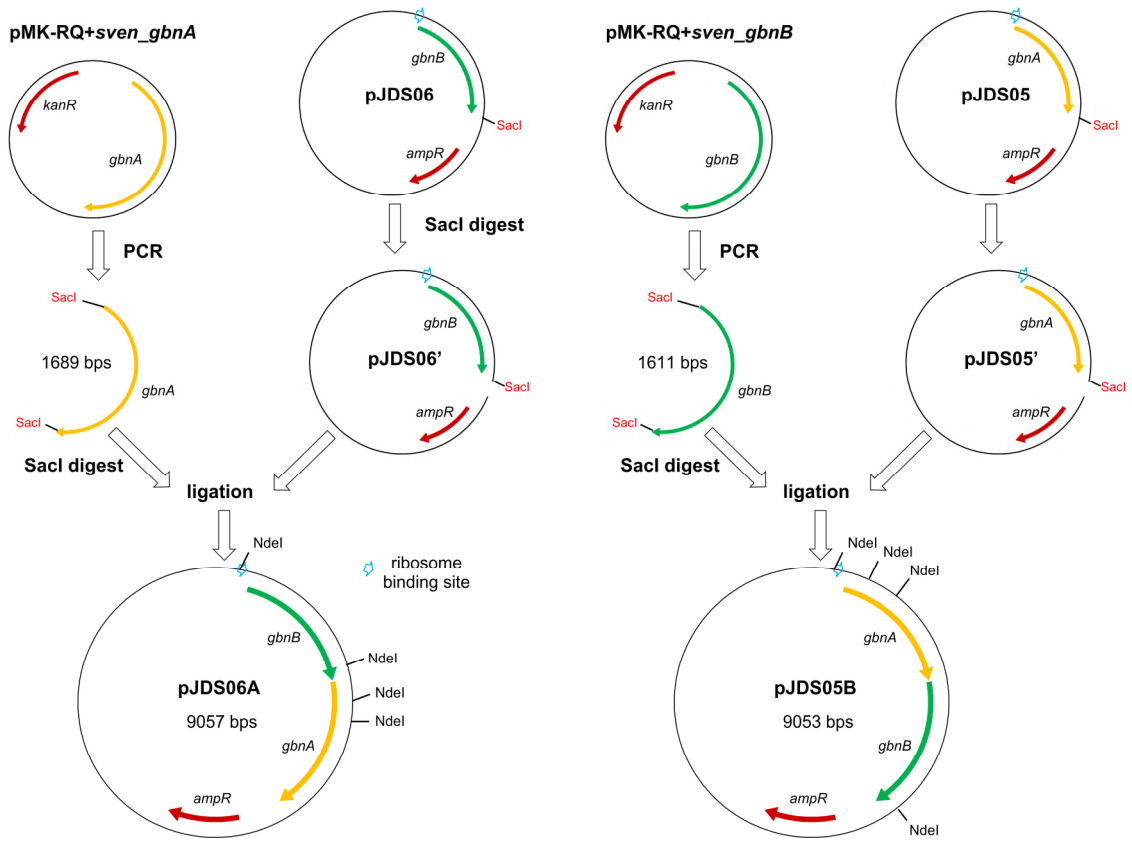


Figure 5.7 – Schematic of the strategy used to sub-clone synthetic *gbnA* into pJDS06 carrying synthetic *gbnB* and sub-cloning of synthetic *gbnB* into pJDS05 carrying synthetic *gbnA*

in which the synthetic versions of *gbnA* and *gbnB* would be flanked by the *Sac*I restriction site (Figure 5.7).

PCR products were obtained using these primer pairs and purified. Single stranded overhangs were then produced by digestion with *Sac*I. Digested synthetic *gbnA*, synthetic *gbnB*, pJDS05 and pJDS06 were all purified by gel extraction, and then attempts were made to ligate synthetic *gbnA* with digested pJDS06 (carrying synthetic *gbnB*) and also to ligate synthetic *gbnB* with digested pJDS05 (carrying synthetic *gbnA*). Table 8.3 illustrates the various ligation conditions attempted.

Transformation of *E. coli* TOP10 with the ligation mixtures did result in colonies on ampicillin selection plates. However, upon digestion of the plasmids purified form

these colonies with *Nde*I, and analysis of the resulting DNA fragments on an agarose gel, the pattern of bands observed was consistent with empty pJDS05 and pJDS06 vectors. In other cases, the plasmid appears to have rearranged into a much shorter one, but retained the ampicillin resistance gene (Figure 5.8).

In two of the purified plasmids for *gbnB* ligated with pJDS05, there was a band at approximately 2700 bp upon digestion with *Nde*I (if *gbnB* had been successfully inserted in pJDS05, then digestion with *Nde*I would be expected to give four bands at 5789 bp, 2699 bp, 295 bp and 270 bp) but only very faint bands around 300 bp.

Therefore, sequencing data for the DNA region of interest was obtained using both the T7 primer pair, and the primer pair O/O'. However, the resulting data showed that the primers bound non-specifically to the plasmid, implying the insert was not correct. The primer pair O/O' were also used in PCRs with the same plasmids, but no PCR products were obtained. Together, both these results implied that the insertion of *gbnB* into pJDS05 was unsuccessful.

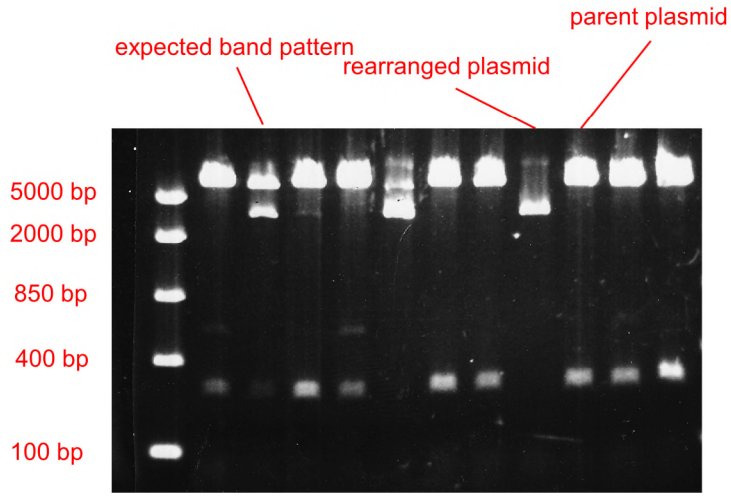


Figure 5.8 – Agarose gel of *Nde*I restriction digests of plasmids after ligations to introduce *gbnB* into pJDS05. Lanes giving examples of the different outcomes of the ligations are annotated

5.3.2 Cloning with pET-Duet and pACYC-Duet vectors

One alternative set of vectors which could be used for co-expression of *gbnA*, *gbnB* and *gbnC* are the Duet cloning vectors available from Novagen, for example pET-Duet and pACYC-Duet, which can each be used to co-express two genes.[180] Both pET-Duet and pACYC-Duet contain two multiple cloning sites (MCSs), one of which can include an N-terminal polyhistidine tag, and the other which can incorporate an S-tag into the N-terminus of the other protein of interest (Figure 5.9 and Appendix 4).

These tags allow purification of the overproduced proteins either by nickel affinity chromatography in the case of the polyhistidine tag, or by recognition with specific antibodies in the case of the S-tag. Each vector has a different origin of replication (pBR322 in pET-Duet and P15A in pACYC-Duet) making them compatible; these plasmids will be maintained in *E. coli* co-transformed with both of them. The vectors pET-Duet and pACYC-Duet can therefore be used to co-express up to four target genes of interest. As with pET151-derived constructs, gene expression in pET-Duet and pACYC-Duet can be induced upon addition of IPTG.

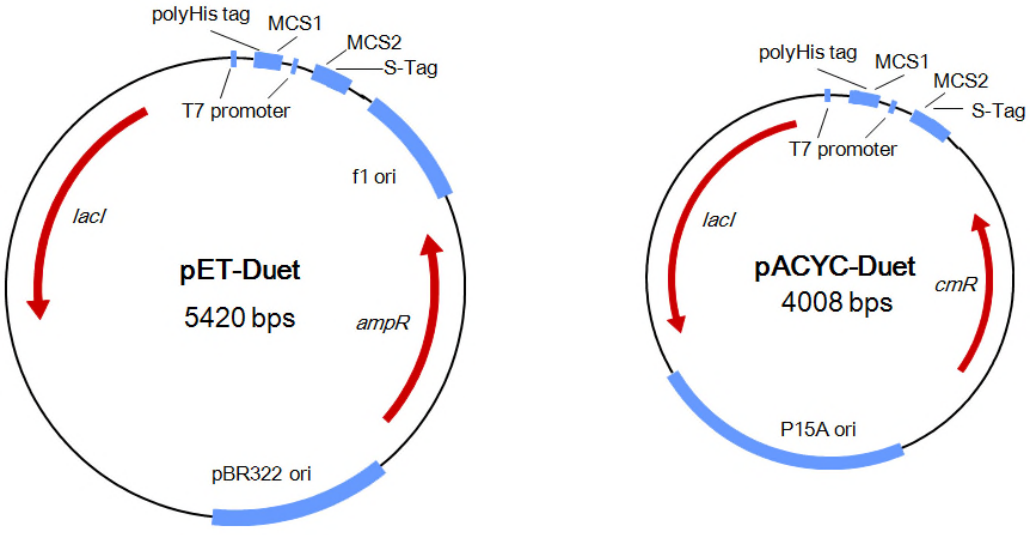


Figure 5.9 – Vectors pET-Duet and pACYC-Duet used for cloning of synthetic *S. venezuelae gbnA*, *gbnB* and *gbnC*

Specific primers P/P' were designed to include restriction sites *Bam*HI and *Hind*III for cloning of synthetic *gbnA* into MCS 1 of pET-Duet. This primer pair was used to amplify *gbnA* by PCR using the pMK-RQ+*sven_gbnA* plasmid (Figure 5.10). As the multiple cloning sites lie downstream of the ATG start codon in pET-Duet, extra nucleotides were added to ensure that the insert was in frame in the destination vector.

The purified *gbnA* PCR product and pET-Duet were both digested with *Bam*HI and *Hind*III, and ligated together using T4 ligase. The ligation products were used to transform *E. coli* TOP10. Plasmids were extracted from ampicillin-resistant colonies and verified as having the correct insert by digestion with *Pst*I and PCR using the primer pair P/P'. Plasmids that had the correct PCR product size and restriction digest pattern had their multiple cloning sites sequenced using the primers 'pETup'/'T7_rev' to confirm that pJDS09 had been produced (Figure 5.10 and Section 8.5).

For the subsequent cloning of synthetic *gbnB* into MCS 2 of pET-Duet, synthetic *gbnB* was amplified from pMK-RQ using specific primers Q/Q' to introduce *Fse*I and *Kpn*I restriction sites at the 5' and 3' ends of synthetic *gbnB*, respectively. This PCR product and pJDS09 were both digested with *Fse*I and *Kpn*I prior to ligation, and the ligated products were used to transform *E. coli* TOP10.

Plasmids were extracted from positive colonies and verified as having the correct insert by digestion with *Hind*III and PCRs were attempted with primers 'UP2'/'T7term'. No PCR products were obtained using this primer pair; however some plasmids did exhibit the expected DNA fragment band pattern upon digestion with *Hind*III. These MCS2 of these plasmids were sequenced using the primer pair 'UP2'/'T7term', confirming that pJDS10 had been created successfully.

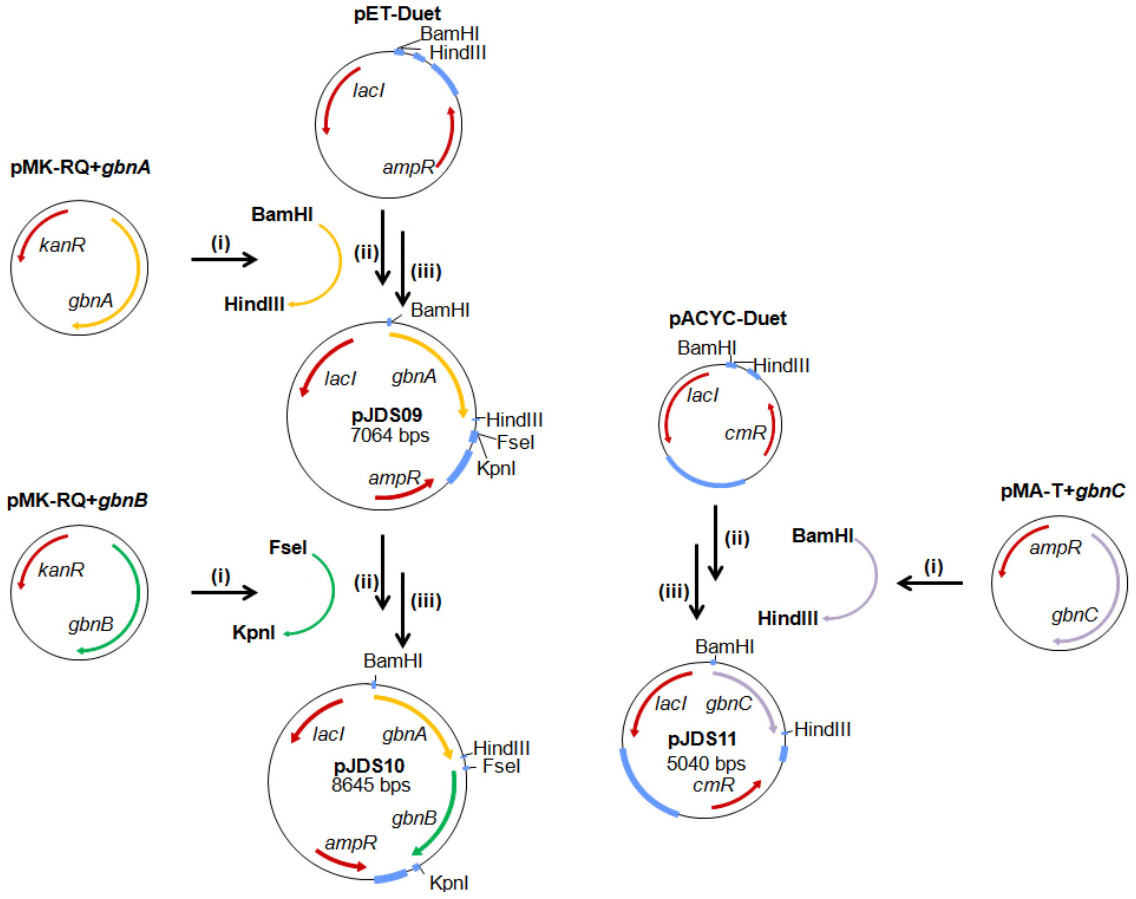


Figure 5.10 – Cloning strategy used to generate plasmids pJDS09-pJDS11 containing synthetic *S. venezuelae gbnA*, *gbnB* and *gbnC*. Step (i) – amplification of each gene of interest by PCR; step (ii) – digestion using specifically selected restriction enzymes to leave appropriate overhangs; step (iii) – ligation of DNA fragments with complementary overhangs

Synthetic *gbnC* was amplified from a pMA-T template carrying the *E. coli* codon optimised *S. venezuelae gbnC* (*sven_4181*) using primers R/R' to introduce *Bam*HI and *Hind*III sites into the PCR product. This product and pACYC-Duet were then digested using *Bam*HI and *Hind*III and were ligated together to create pJDS11. Plasmids from positive colonies were digested with *Acc*I to identify plasmids with the correct insert. These plasmids were verified by sequencing of the MCSs using primer pair 'pACYCup'/'T7_term'.

5.3.3 Metabolic profiling of *E.coli* carrying pET/pACYC-Duet derived plasmids

Attempts were made to co-transform *E. coli* BL21star using pJDS10 (carrying synthetic *S. venezuelae gbnA* and *gbnB*) and pJDS11 (carrying synthetic *S. venezuelae gbnC*) in order to create a strain which could overexpress *gbnA*, *gbnB* and *gbnC*. Therefore, in this strain, all three proteins GbnA, GbnB and GbnC would be overproduced. It was hypothesised that there may be gaburedin-like metabolites produced by this strain, as overproduction of GbnC would circumvent the potential problem of gaburedins being unable to cross the cell membrane without GbnC. However, neither chemical transformation nor transformation of *E. coli* BL21star by electroporation using both plasmids pJDS10 and pJDS11 were successful during the timeframe of this project. Attempts to transform *E. coli* BL21star using just pJDS10 were successful, however.

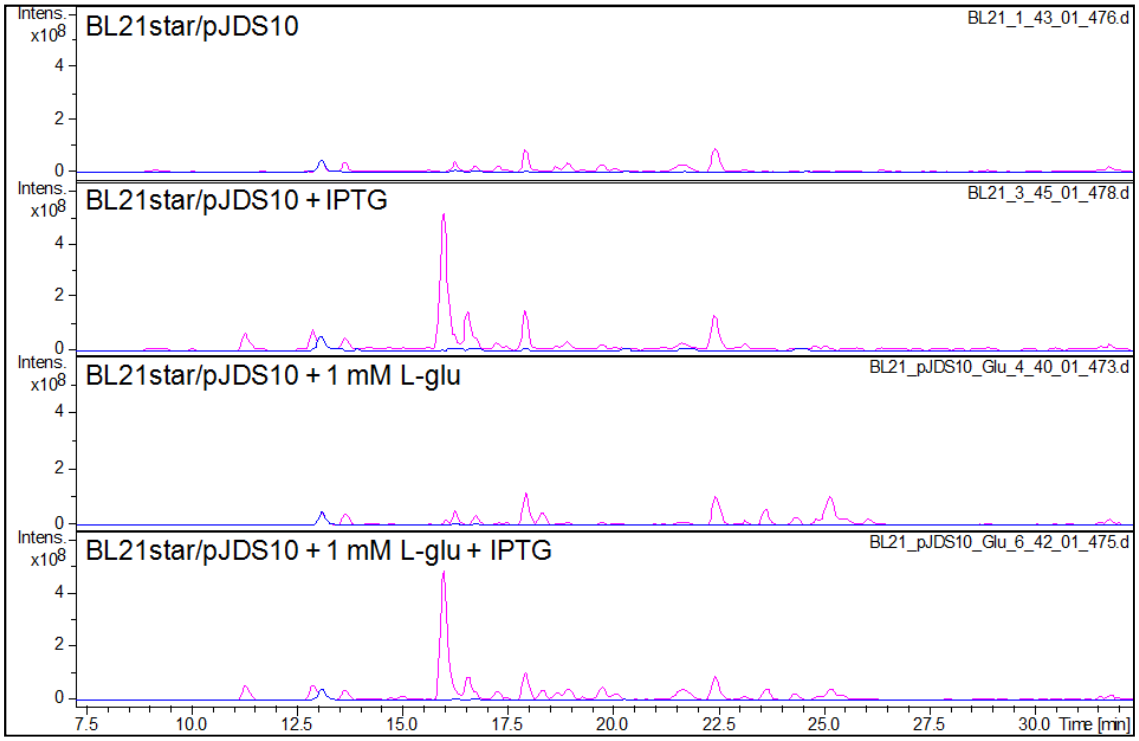


Figure 5.11 – Base peak chromatograms (pink traces) and extracted ion chromatograms for gaburedins A-F (**80-85**) *m/z* = 247.0, 261.0, 279.0, 292.9, 294.9 (blue traces) of metabolites extracted from culture supernatants of *E. coli* BL21star/pJDS10 grown without IPTG and with 1 mM IPTG in LB with and without addition of 1 mM L-glutamic acid

Plasmid pJDS10 was used to transform *E. coli* BL21star, and this culture was grown in LB to an OD_{600nm} of 0.6, at which time protein expression was induced with IPTG. In parallel, cultures without IPTG added were also grown. Cells were pelleted and ethyl acetate extracts of the culture supernatants were prepared. As shown in Figure 5.11, there were some new metabolites produced by *E. coli* BL21star/pJDS10 induced with IPTG compared to *E. coli* BL21star/pJDS10 grown without any induction using IPTG. However, these compounds exhibit a predominant ion with *m/z* = 303 in their mass spectra, corresponding to acetylated derivative of IPTG, which has been previously reported.[181] Indeed, there were no gaburedins or gaburedin-like compounds produced. Therefore, cells were lysed and ethyl acetate extracts of the cell lysates were also performed, however there were still no gaburedin-like metabolites observed in the LC-MS chromatograms of the extracts from the cell lysates.

To account for concentration of L-glutamic acid in the LB media being a limiting factor in gaburedin biosynthesis (for example, if there was an insufficient quantity available to be utilised by GbnA), this experiment was repeated. The *E. coli* BL21star/pJDS10 strain was grown in media enriched with 1 mM L-glutamic acid, as this would increase the availability of the proposed GbnA substrate L-glutamate. Again, LC-MS analysis of the extracted metabolites revealed that the only new metabolites produced were IPTG-derived compounds (Figure 5.11).

It is possible that the lack of gaburedin-like metabolites present in these samples is due to poor production levels of the GbnA and GbnB proteins. Therefore, protein samples were run to compare the soluble and insoluble fractions of *E. coli* BL21star/pJDS10 with and without induction using IPTG. However, there were no obvious differences between the protein fractions extracted from the *E. coli* BL21star/pJDS10 which were

induced with IPTG compared to the metabolites extracted from the sample which was not induced with IPTG (it would be expected that bands at 60.1 kDa and 58.7 kDa would be present for recombinant GbnA and GbnB, respectively), indicating that the protocol for co-expression of *gbnA* and *gbnB* requires optimisation.

5.4 – Summary of molecular biology work

Analysis of the SV4 F01 cosmid sequence showed there to be an unusually large intergenic region of 1024 bp between the *gbnA* gene (*sven_4179*) and the adjacent, divergent gene *sven_4178*. This could imply that there were mistakes in the sequencing of this part of the *S. venezuelae* genome; therefore the primers may have been designed using incorrect sequences based upon this sequencing data. Indeed, this is true for many examples of sequencing data being used to order synthetic versions of native genes. Therefore, it would perhaps be prudent in future to re-sequence the entire SV4 F01 cosmid. However, the remaining genes in the *gbn-sgn* cluster have all been amplified by PCR from the SV4 F01 template.

It would appear that – despite development of several distinct molecular biology strategies for producing vectors which carry multiple synthetic genes from the *gbnABC* system – none of the plasmids constructed confer the gaburedin-producing chemotype. The sequences of all genes inserted into the plasmids pJDS02, pJDS04, pJDS06 and pJDS09-pJDS11 have been verified; therefore future work should be focussed on optimisation of co-expression of the genes in the plasmids pJDS10 and pJDS11 in *E. coli* BL21star.

It may be possible that *E. coli* BL21star can confer the gaburedin-producing chemotype when transformed using combination of plasmids constructed during this study. For example, both plasmids pJDS06 (carrying synthetic *sven_gbnB*) and pJDS11 (carrying synthetic *sven_gbnC*) could be used to co-transform the same *E. coli* BL21star, allowing *gbnB* and *gbnC* to be overexpressed in the same cells. The feeding experiments with GABA in Section 5.2.2 with *E. coli*/pJDS06 only could be then repeated, in order to definitively confirm whether or not gaburedins are produced by this strain when the export protein GbnC is present. This would allow the role of GbnC in the export of gaburedins to be ascertained.

6. Regulation of gaburedin biosynthesis in *S. venezuelae*

In *S. coelicolor*, deletion of the transcriptional repressor gene *mmyR* leads to overproduction of not only methylenomycin, but AHFCA signalling molecules as well.[100] This led to the hypothesis that identification and subsequent deletion of *mmyR* homologues in other bacteria (such as *gbnR* in *S. venezuelae*) where the *mmfLHP*, *mmfR* and *mmyR* set of genes is conserved would turn on the production of both new natural products (as in the case of gaburedins) and turn on production of the AHFCA signalling molecules predicted to be produced by the MmfLHP analogues SgnLPH (Figure 6.1).

An induction assay has already been developed for the methylenomycin system, whereby specific AHFCA signalling molecules were shown to restore the antibiotic activity of a strain lacking *mmfLHP*. An analogous induction assay could potentially be developed to assess if AHFCAs can induce gaburedin production in *S. venezuelae*. As

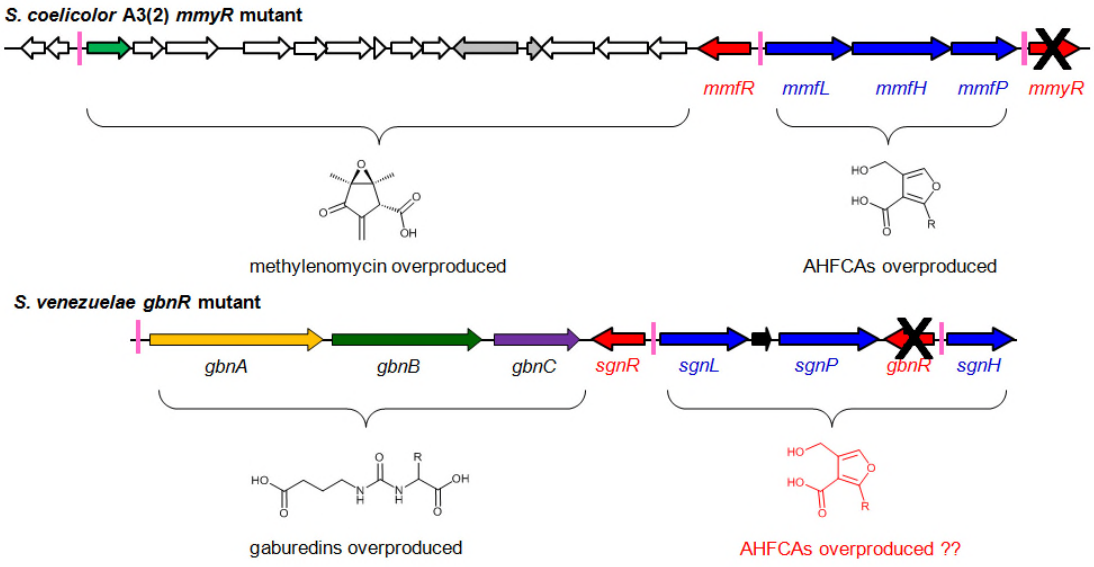
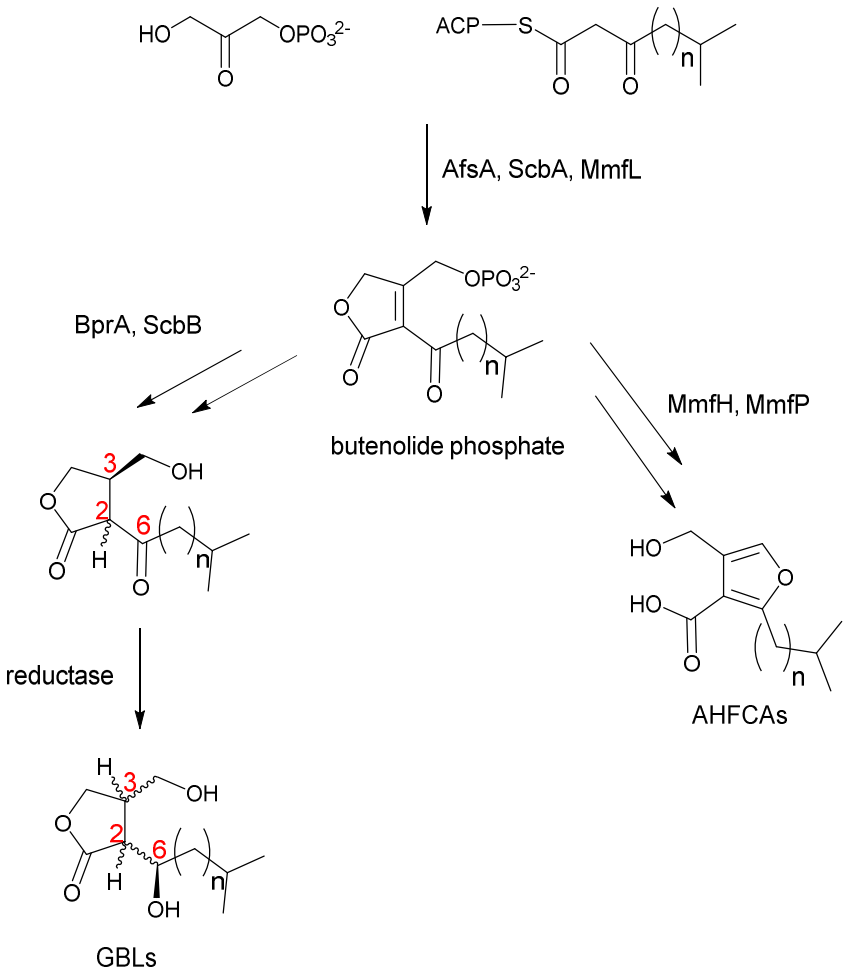


Figure 6.1 – Schematic to demonstrate phenotype of *S. coelicolor mmyR* deletion mutant and proposed phenotype of *S. venezuelae gbnR* deletion mutant. Repressor binding sites (AREs) shown in pink. Sizes of genes not to scale



Scheme 6.1– GBL and AHFCA biosynthesis proceed via a shared butenolide phosphate intermediate. The downstream enzymes (for example MmfH, MmfP) determine the nature of the final signalling molecules produced

the function of gaburedins has not yet been established, an LC-MS assay rather than an activity-based assay could be developed.

The key biosynthetic enzyme (the butenolide synthase) involved in AHFCA biosynthesis in *S. coelicolor*, MmfL, has homologues not only in homologous gene clusters that contain *mmfP* and *mmfH*, but also in gene clusters involved in the biosynthesis of GBL signalling molecules (there is a shared intermediate in the proposed biosynthesis of furan and butyrolactone signalling molecules, Scheme 6.1).[79,109]

In *Streptomyces venezuelae*, there are three homologues of *mmfL*; *sven_4183* (*sgnL* in the gaburedin cluster, 33 % identity, 44 % similarity at amino acid level), *sven_5093* (29 % identity, 38 % similarity at amino acid level) and *sven_5969* (*jadW1* in the jadomycin gene cluster, 30 % identity, 43 % similarity at amino acid level). The *jadW1* gene is involved in GBL biosynthesis in the regulation of jadomycin (See Section 1.4).[95] Therefore, if *sgnL* is overexpressed (as expected if the transcriptional repressor gene *gbnR* is deleted), it is possible that an accumulation of the intermediate butenolides produced by SgnL could lead to production of both AHFCAs and GBLs (Scheme 6.1). Indeed, some crosstalk between pathways including *afsA* homologues has been implied in jadomycin biosynthesis (in order to abolish jadomycin production in a *jadW1* mutant, it was necessary to also delete *sven_5093*, yet deletion of *sgnL* did not affect jadomycin production). Therefore, crosstalk between the *afsA*-like pathways cannot be excluded.[95]

GBLs and AHFCAs have both been extensively studied by LC-MS. The base peak in electrospray MS of GBLs is the $[M+Na]^+$; the base peak for AHFCAs is the $[M-H_2O+H]^+$.[140,182] Furthermore, mass spectra of GBLs also exhibit peaks for $[M-H_2O+H]^+$ and $[M-2H_2O+H]^+$ ions, whereas the next most intense peak for AHFCAs corresponds

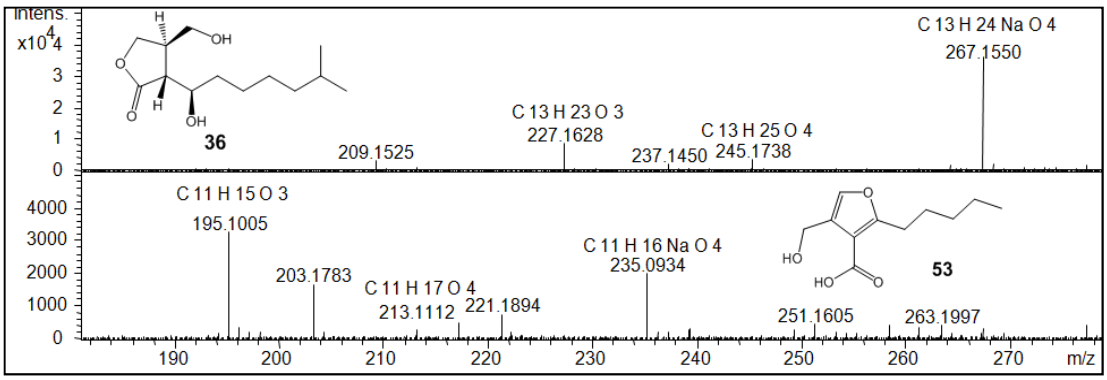


Figure 6.2 – UHR-MS data for SCB1 (**36**, top panel) and AHFCA5 (**53**, bottom panel) illustrating the typical mass spectra of 6-hydroxy GBLs and AHFCAs observed in electrospray MS in positive ion mode

to the $[M+Na]^+$ ion, only a very small peak for the $[M+H]^+$ ion is observed. Figure 6.2 shows example mass spectra of SCB1 (**36**) and AHFCA5 (**53**).

6.1 Identification and characterisation of *Streptomyces venezuelae* signalling molecules

6.1.1 Metabolic profiling of *S. venezuelae* mutants

It was noted during initial LC-MS analysis of the culture extracts of the *S. venezuelae* wild type, *sgnL::apra*, *gbnR::apra* and *gbnB::apra*,*gbnR::scar* mutants, that there was another series of compounds produced by the *gbnR::apra* mutant strains that were not produced by the wild type and *sgnL::apra* strains (Figure 2.1). These metabolites have $[M+Na]^+$ ions in their mass spectra with *m/z* = 253.0, 267.0 and 281.0.

Extracted ion chromatograms were produced for the $[M+Na]^+$ species expected for the known GBLs of both the 6-keto and 6-hydroxy types for the metabolites produced by the *gbnR* mutant strain. These analyses revealed that five molecules **188-192** share a similar fragmentation pattern to that observed for 6-hydroxy type GBLs.[183] The mass spectra of each of these molecules consists of a series of peaks corresponding to the $[M+Na]^+$ and $[M+H]^+$ molecular ions, accompanied by $[M-H_2O+H]^+$ and $[M-2H_2O+H]^+$ fragment ions (Figure 6.3). The molecular formulae of these ions were confirmed by UHR-MS analyses, confirming there to be one GBL-like molecule **188** with formula $C_{11}H_{20}O_4$, two with formulae of $C_{12}H_{22}O_4$ (**189** and **190**) one with formula of $C_{13}H_{24}O_4$ (**191**) and one with formula $C_{14}H_{26}O_4$ (**192**, Figure 6.3).

6.1.2 Comparison of putative signalling molecules with *Streptomyces* M1152

In order to characterise *S. venezuelae* signalling molecules, the LC-MS chromatograms of the *gbnB/R* strain were compared with culture extract from *Streptomyces* M1152. *Streptomyces* M1152 is a 'super-host' which has been engineered to have the actinorhodin (*act*), prodiginine (*red*), calcium-dependent antibiotic (*cda*) and coelimycin (*cpk*) gene clusters removed.[81] This results in a very clean base-peak chromatogram upon LC-MS analysis culture extracts, due to there being far fewer secondary metabolite compounds in the chromatogram background.

The gene *scbR2 (sco_6286)* – encoding for a transcriptional regulator which is thought to regulate SCB biosynthesis – is not present in the M1152 host, whereas the

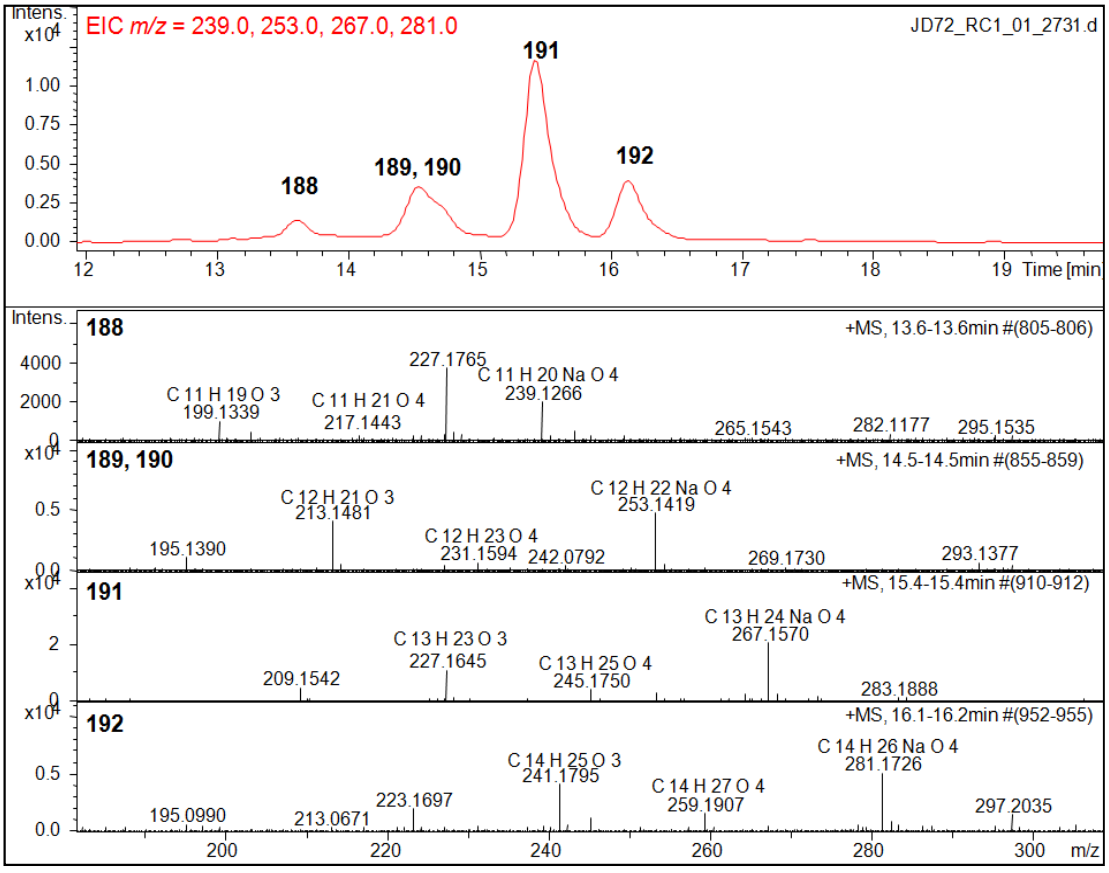


Figure 6.3 – Extracted ion chromatograms for $[M+Na]^+$ ions of 6-hydroxy type GBLs *m/z* = 239.0, 253.0, 267.0 and 281.0 observed in metabolites extracted from *S. venezuelae gbnR* mutants. Mass spectra of these GBL-like molecules are also shown with the molecular formulae generated for the $[M+Na]^+$, $[M+H]^+$ and $[M-H_2O+H]^+$ ions

biosynthetic genes involved in SCB biosynthesis *scbA* (*sco_6266*), *scbB* (*sco_6267*) and *sco_6264* remain.[81] This was expected to lead to overproduction of SCBs in M1152, as an *scbR2* deletion mutant of the parent strain *S. coelicolor* M145 was previously shown to overexpress *scbA*, and to overproduce SCB1 (**36**).[149]

Hence, extracted ion chromatograms were generated for *m/z* = 239.0, 253.0, 267.0, and 281.0, corresponding to $[M+Na]^+$ ions for SCB-like GBLs in *Streptomyces coelicolor* M1152 culture extract, and metabolites **193-200** were identified (Figure 6.4). As expected, several of these metabolites **197-200** elute at the approximate retention times and have fragmentation patterns that would be consistent with SCBs 1-3. The butenolide synthase *scbA* is the only *afsA* homologue in M1152 and is presumably responsible for the assembly of the GBL-like molecules **193-200** observed in M1152. As M1152 also has intact *scbB* and *sco_6264* genes (which are proposed to catalyse the downstream steps to furnish SCBs from the butenolide phosphate intermediate, see Figure 1.18), therefore the stereochemistry of these GBLs is presumably the same as the stereochemistry of the SCBs 1-3 (2*R*, 3*R*, 6*S*).[81,132,138] However, the stereochemistry could be confirmed by purification of the compounds and subsequent analysis by CD and NMR spectroscopy.[95,183] Alternatively, chiral HPLC of the culture extract could be performed, with co-injection of an authentic standard of SCBs.[137]

Comparison of culture extracts of the *S. venezuelae gbnB/R* strain and *Streptomyces* M1152 strains revealed that several of the compounds **193-200** were also present at the same retention times in the culture extract of the *S. venezuelae gbnB/R* strain; compounds **188** and **193** elute at the same time (19.9 min), **189** and **195** both elute at 21.8 min, **190** and **196** elute at 22.2 min, **191** and **197** elute at 23.7 min and compounds **192** and **199** both elute at 25.0 min (Figure 6.4). If the stereochemistry of the GBLs

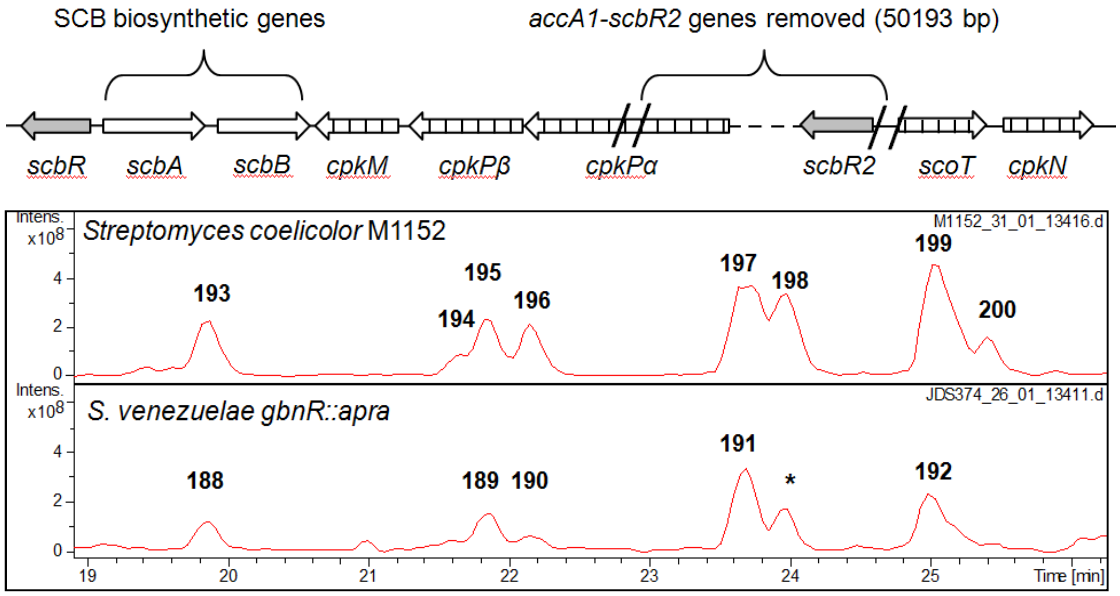
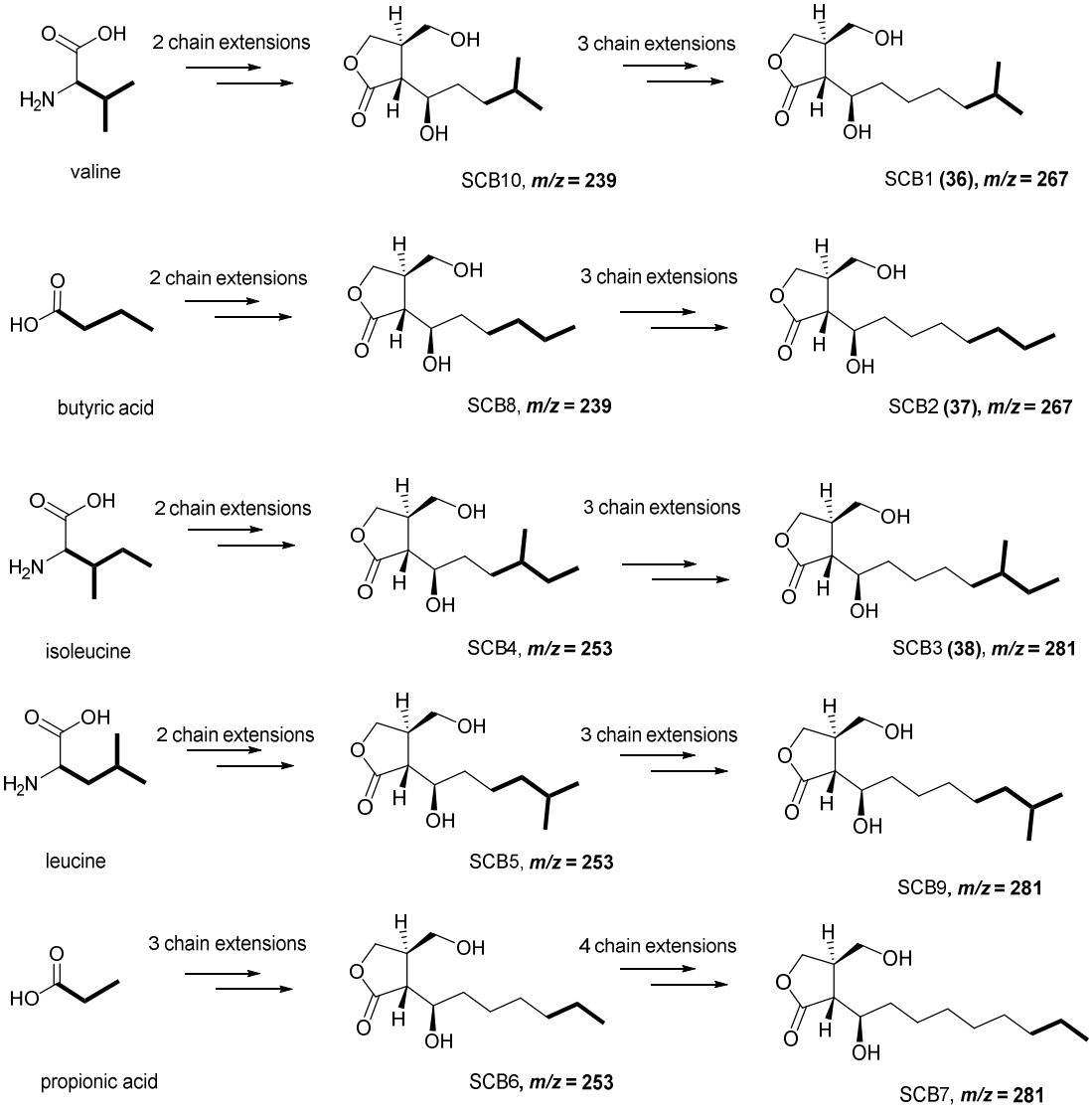


Figure 6.4 – Top: genetic organisation of the *scb-cpk* gene cluster as modified in *S. coelicolor* M1152. The gene encoding for the transcriptional repressor protein *scbR2* has been removed. Bottom: Extracted ion chromatograms for $[M+Na]^+$ ions *m/z* = 239.0, 253.0, 267.0 and 281.0 of 6-hydroxy type butyrolactones in the metabolites extracted from *S. coelicolor* M1152 and *S. venezuelae gbnB/R*. Compound denoted by * has *m/z* = 267.0 but is not related to GBLs

produced by *S. venezuelae* was different to that of the GBLs extracted from *S. coelicolor* M1152, one would expect the GBLs derived from the same precursor molecules to have different retention times in an achiral column – if the stereochemistry was different at one or two out of the three stereocentres then the compounds would be diastereoisomers and would elute at different retention times.[137]

6.1.3 Characterisation of novel SCBs in *S. coelicolor* M1152

After initial identification of the eight GBL-like molecules in *Streptomyces* M1152, experiments were devised in order to solve their structures. GBLs comprise a conserved butyrolactone moiety with varying length and branching in the alkyl chain (see Section 1.5). As these are derived from fatty acid biosynthesis, there are specific precursor molecules (amino acids and carboxylic acids) which are used as starter units in fatty acid biosynthesis, which are subsequently incorporated into the butenolide intermediates and ultimately found in SCBs. Scheme 6.2 illustrates ten possible



Scheme 6.2 – Biosynthetic route to SCBs which would give rise to $[M+Na]^+$ ions with *m/z* = 239.0, 253.0, 267.0 and 281.0. The number of chain extensions in fatty acid biosynthesis is also shown

structures of SCBs that would derive from five different starter units in fatty acid metabolism and would give rise to the $[M+Na]^+$ ions with *m/z* = 239, 252, 267 and 281 shown in Figure 6.4.

As discussed in Section 1.5, the amino acids valine, leucine and isoleucine lose their α-hydrogens as they are converted into their cognate α-keto acids, which are then transformed into starter units isobutyryl-CoA, 3-methylbutyryl-CoA and 2-methylbutyryl-CoA for fatty acid biosynthesis (Scheme 1.4).[135,136] Therefore, it is

expected that any SCBs derived from these amino acids would retain 7 deuterium labels for those derived from d_8-valine, and 9 deuterium labels from d_{10}-leucine or d_{10}-isoleucine. SCBs that were derived from propionic and butyric acid would be expected to retain 5 and 7 deuterium labels, respectively. As the deuterium labels added form part of the alkyl chain of the SCBs, it was expected that labelling would be observed in the $[M\text{-}H_2O+H]^+$ and $[M\text{-}2H_2O+H]^+$ fragment ions as well as the $[M+H]^+$ and $[M+Na]^+$ molecular ions for each labelled SCB.

Therefore, *Streptomyces* M1152 was grown in AlaMM media supplemented with 1 mM final concentration deuterium-labelled fatty acid precursor molecules d_8-DL-valine, d_{10}-L-leucine, d_{10}-L-isoleucine, d_5-propionic acid and d_7-butyric acid. This method was used previously to probe incorporation of specific deuterium-labelled precursor molecules into the AHFCAs.[79,140] The metabolites were extracted and analysed by UHR-LC-MS. As expected, feeding with each different precursor molecule altered the mass spectra of specific SCBs.

For instance by feeding with d_8-DL-valine, a proportion of the compound **197** (SCB1) with retention time 16.4 min and $[M+Na]^+$ *m/z* = 267.2 was specifically labelled with 7 deuterium labels (Figure 6.5), and the *m/z* for all peaks for the $[M+H]^+$ molecular ion and $[M\text{-}H_2O+H]^+$ and $[M\text{-}2H_2O+H]^+$ fragment ions also increased by 7. This confirms that this SCB is derived from valine, therefore it was proposed that **197** is SCB1 (compound **36**). No labelling was observed for any of the other SCBs in the *Streptomyces* M1152 culture extract, implying that SCB1 is the only SCB directly derived from valine.

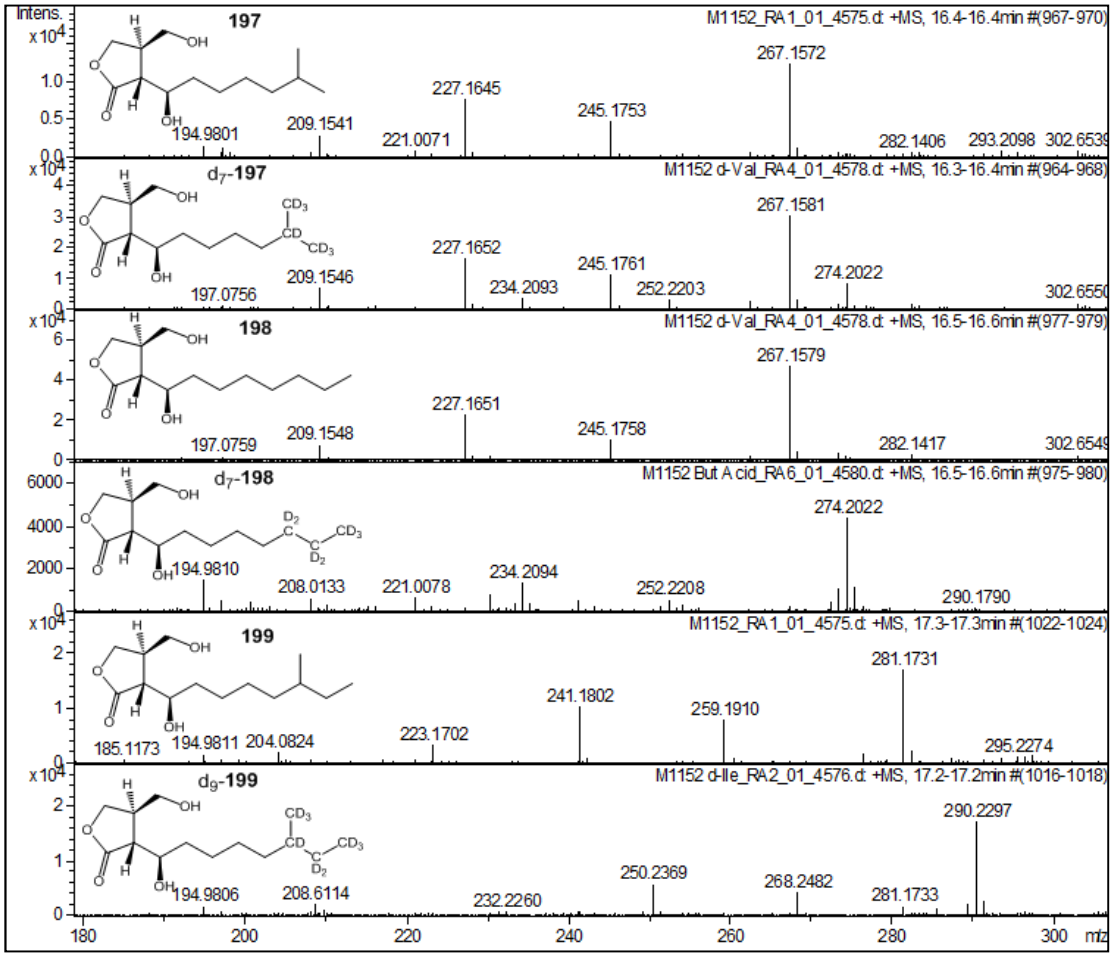


Figure 6.5 – UHR-MS of **197-199** extracted from *S. coelicolor* M1152 grown on different AlaMM media. From top to bottom: **197** without any labelling, **197** labelled upon feeding with d_8-DL-valine, **198** without any labelling, **198** labelled upon feeding with d_7-butyric acid, **199** without any labelling, and **199** labelled upon feeding with d_{10}-isoleucine

Extracted ion chromatograms were generated for the $[M+Na]^+$ ions expected for the remaining SCBs identified, and for the *m/z* of ions that would be expected if any of these species were labelled with deuterium from each precursor fed. As shown in Figure 6.5, compound **198** (t_R = 16.5 min) incorporated 7 deuterium labels upon feeding with d_7-butyric acid, suggesting that **198** has the same structure as SCB2 (**37**). Compound **199** (t_R = 17.3 min) was enriched with 9 deuterium labels upon feeding with d_{10}-isoleucine, consistent with the known structure of SCB3 (**38**). Molecular formulae generated from UHR-LC-MS analyses of these labelled and unlabelled compounds **197-199** are shown in Table 6.1.

Table 6.1 – UHR-MS of $[M+Na]^+$ ions of SCBs 1-8 identified in culture extract of *S. coelicolor* M1152 grown on AlaMM media supplemented with different stable isotope precursor molecules. Full UHR-MS characterisation shown in Appendix 5

retention time / min	*compound*	*precursor fed*	*Molecular formula*	*Observed* m/z	*Calculated* m/z
14.3	SCB8 (**193**)	none	$C_{11}H_{20}NaO_4$	239.1258	239.1254
14.3	d_7-SCB8 (**193**)	d_7-butyric acid	$C_{11}H_{13}D_7NaO_4$	246.1705	246.1693
15.2	SCB4 (**194**)	none	$C_{12}H_{22}NaO_4$	253.1421	253.1410
15.2	d_9-SCB4 (**194**)	d_{10}-L-isoleucine	$C_{12}H_{13}D_9NaO_4$	262.1991	262.1975
15.3	SCB5 (**195**)	none	$C_{12}H_{22}NaO_4$	253.1421	253.1410
15.3	d_9-SCB5 (**195**)	d_{10}-L-leucine	$C_{12}H_{13}D_9NaO_4$	262.1982	262.1975
15.5	SCB6 (**196**)	none	$C_{12}H_{22}NaO_4$	253.1418	253.1410
15.5	d_5-SCB6 (**196**)	d_5-propionic acid	$C_{12}H_{17}D_5NaO_4$	258.1726	258.1724
16.4	SCB1 (**197**)	none	$C_{13}H_{24}NaO_4$	267.1571	267.1567
16.4	d_7-SCB1 (**197**)	d_8-DL-val	$C_{13}H_{17}D_7NaO_4$	274.2022	274.2006
16.5	SCB2 (**198**)	none	$C_{13}H_{24}NaO_4$	267.1579	267.1567
16.5	d_7-SCB2 (**198**)	d_7-butyric acid	$C_{13}H_{17}D_7NaO_4$	274.2021	274.2006
17.3	SCB3 (**199**)	none	$C_{14}H_{26}NaO_4$	281.1729	281.1723
17.3	d_9-SCB3 (**199**)	d_{10}-L-isoleucine	$C_{14}H_{17}D_9NaO_4$	290.2298	290.2288
17.5	SCB7 (**200**)	none	$C_{14}H_{26}NaO_4$	281.1731	281.1723
17.5	d_5-SCB7 (**200**)	d_5-propionic acid	$C_{14}H_{21}D_5NaO_4$	286.2045	286.2037

Furthermore, feeding with d_{10}-isolecuine resulted in specific labelling of compound **194** (t_R = 15.2 min). Metabolites **195** (t_R = 15.3 min) and **196** (t_R = 15.5 min) were found to specifically incorporate nine deuterium atoms from d_{10}-leucine and five deuterium atoms from d_5-propionic acid, respectively. Metabolite **200** (t_R = 17.5 min) was labelled specifically upon addition of d_5-propionic acid and **193** (t_R = 14.3 min) was labelled specifically upon addition of d_7-butyric acid to the media. It was therefore proposed that compounds **193**, **194**, **195**, **196** and **200** have structures corresponding to the novel SCBs 8, 4, 5, 6 and 7 respectively (Scheme 6.2). The results of the labelling experiments are summarised in Figure 6.6 and Table 6.1. By comparison of the retention times of the GBLs produced by the *S. venezuelae gbnB/R* strains with those

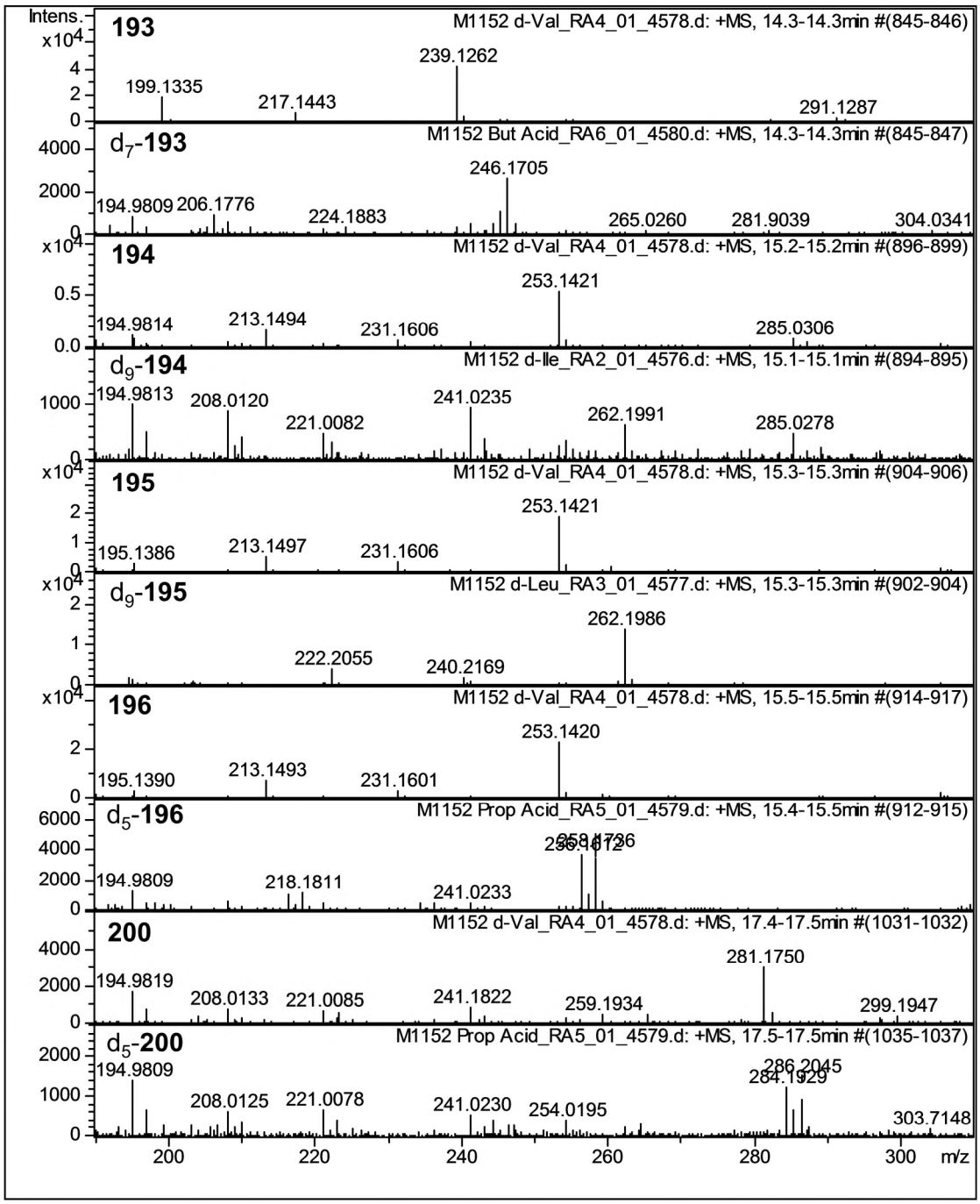


Figure 6.6 – UHR-MS of metabolites **193**, **194**, **195**, **196** and **200** extracted from *S. coelicolor* M1152 grown on different AlaMM media. **193** was labelled with 7 deuteriums from incorporation of d_7-butyric acid; **194** was labelled with 9 deuteriums from incorporation of d_{10}-isoleucine; **195** was labelled with 9 labels from incorporation of d_{10}-leucine; **196** was labelled with 5 labels from incorporation of d_5-propionic acid; and **200** was labelled with 7 deuterium labels from incorporation of d_7-butyric acid

produced by M1152, it was deduced that SCB1 (**191**), SCB3 (**192**), SCB5 (**189**), SCB6 (**190**) and SCB8 (**188**) are also produced by *S. venezuelae* (Figure 6.4 and Scheme 6.2).

6.1.4 AHFCA signalling molecules in *Streptomyces venezuelae*

In addition to the observation of SCBs being present in the *S. venezuelae gbnR* and *S. venezuelae gbnB/R* mutant extracts, closer inspection of the LC-MS data for revealed the presence of new metabolites **201**–**204**, corresponding to the *m/z* values consistent with $[M\text{-}H_2O+H]^+$ ions of AHFCA or AHFCA-like molecules. Mass spectra of each of these metabolites illustrated that the two most intense peaks correspond to $[M+Na]^+$ and $[M\text{-}H_2O+H]^+$ ions.

Therefore, extracted ion chromatograms were generated for the *m/z* values corresponding to the $[M\text{-}H_2O+H]^+$ peaks for the previously characterised AHFCA2 (*m/z* = 167.0), AHFCAs 1 and 4 (*m/z* = 181.0) and AHFCAs 3 and 5 (*m/z* = 195.0).[79] Extracted ion chromatograms were also generated for ions with *m/z* = 209.0, 223.0 and

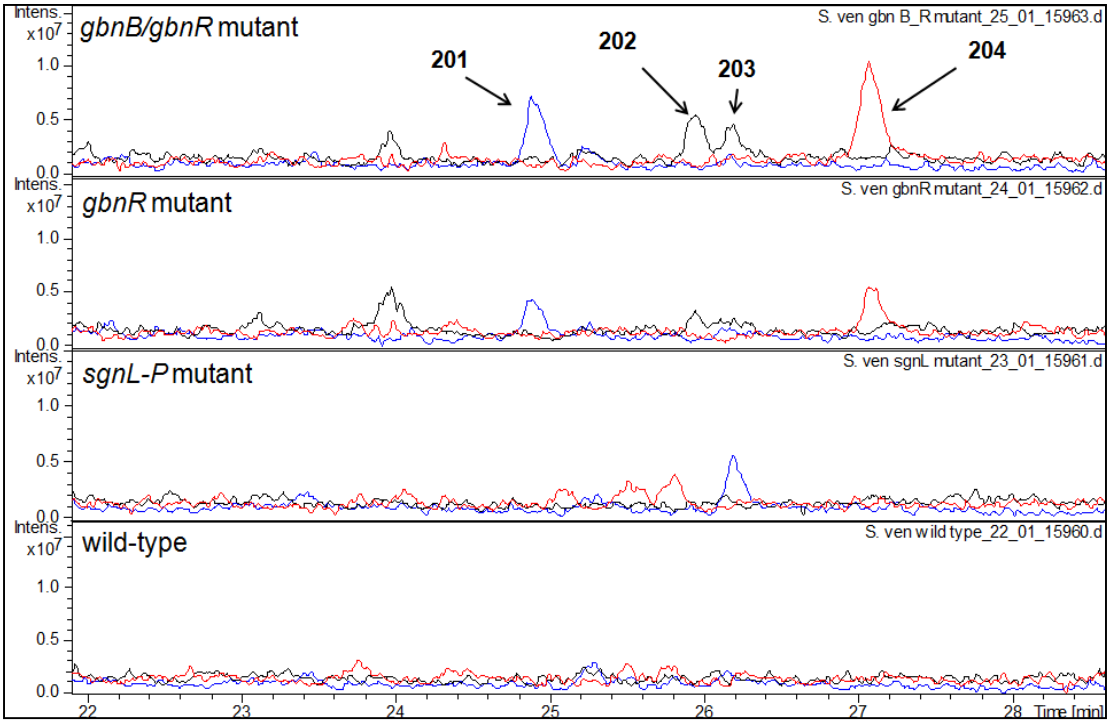


Figure 6.7 – Extracted ion chromatograms for $[M\text{-}H_2O+H]^+$ ions of AHFCA species present in the metabolites extracted from the *S. venezuelae gbnB/gbnR*, *gbnR*, *sgnL* and wild type. Chromatograms correspond to *m/z* = 195.0 (blue traces), 209.0 (black traces) and 223.0 (red traces). Peaks denoted by * are unrelated to AHFCA compounds. EICs for *m/z* = 167.0, 181.0 and 237.0 have been omitted for clarity

237.0 that would correspond to $[M\text{-}H_2O+H]^+$ ions of AHFCAs that contain longer alkyl chains. As demonstrated by Figure 6.7, there is one metabolite **201** with $[M\text{-}H_2O+H]^+$ ion with *m/z* = 195.0 (corresponding to either AHFCA3 or 5), two metabolites **202** and **203** with $[M\text{-}H_2O+H]^+$ ion with *m/z* = 209.0, and compound **204** with an $[M\text{-}H_2O+H]^+$ ion with *m/z* = 223.0.

AHFCAs are derived from the same precursors as SCBs, and they exhibit a variety of alkyl chains in the 4-position of the furan ring.[79,109,132] Therefore, to solve the structures of these AHFCA compounds, a similar set of feeding experiments analogous to the above feeding experiments were devised. The *S. venezuelae gbnB/R* strain was grown on SMMS supplemented with 10 mM d_8-DL-valine, 10 mM d_{10}-L-leucine, 10 mM d_{10}-L-isoleucine, 1 mM d_5-propionic acid and 1 mM d_7-butyric acid in order to determine the nature of the side chains of each of the AHFCAs identified (d_5-propionic acid and d_7-butyric acid were found to inhibit the growth of the *S. venezuelae gbnB/R* strain at 10 mM concentration). The *S. venezuelae gbnB/R* strain was selected for these experiments instead of the *gbnR* strain as the *gbnB/R* strain does not produce gaburedins, therefore has a cleaner background for LC-MS analyses.

Figure 6.8 illustrates that the *m/z* of AHFCA-like compound **201** observed at t_R = 24.9 min (*m/z* = 195.0 for its $[M\text{-}H_2O+H]^+$ ion) does indeed increase by 7 upon feeding with d_7-butyric acid, confirming that this is the same structure as that of AHFCA**5** (**53**) previously isolated from *Streptomyces coelicolor*.[79] Furthermore, the AHFCA-like metabolites **202** (t_R = 25.9 min) and **203** (t_R = 26.1 min) were shown to be specifically labelled upon supplementing the growth media with d_{10}-leucine and d_5-propionic acid, respectively. These new metabolites **202** and **203** were named AHFCA6 and AHFCA7, respectively.

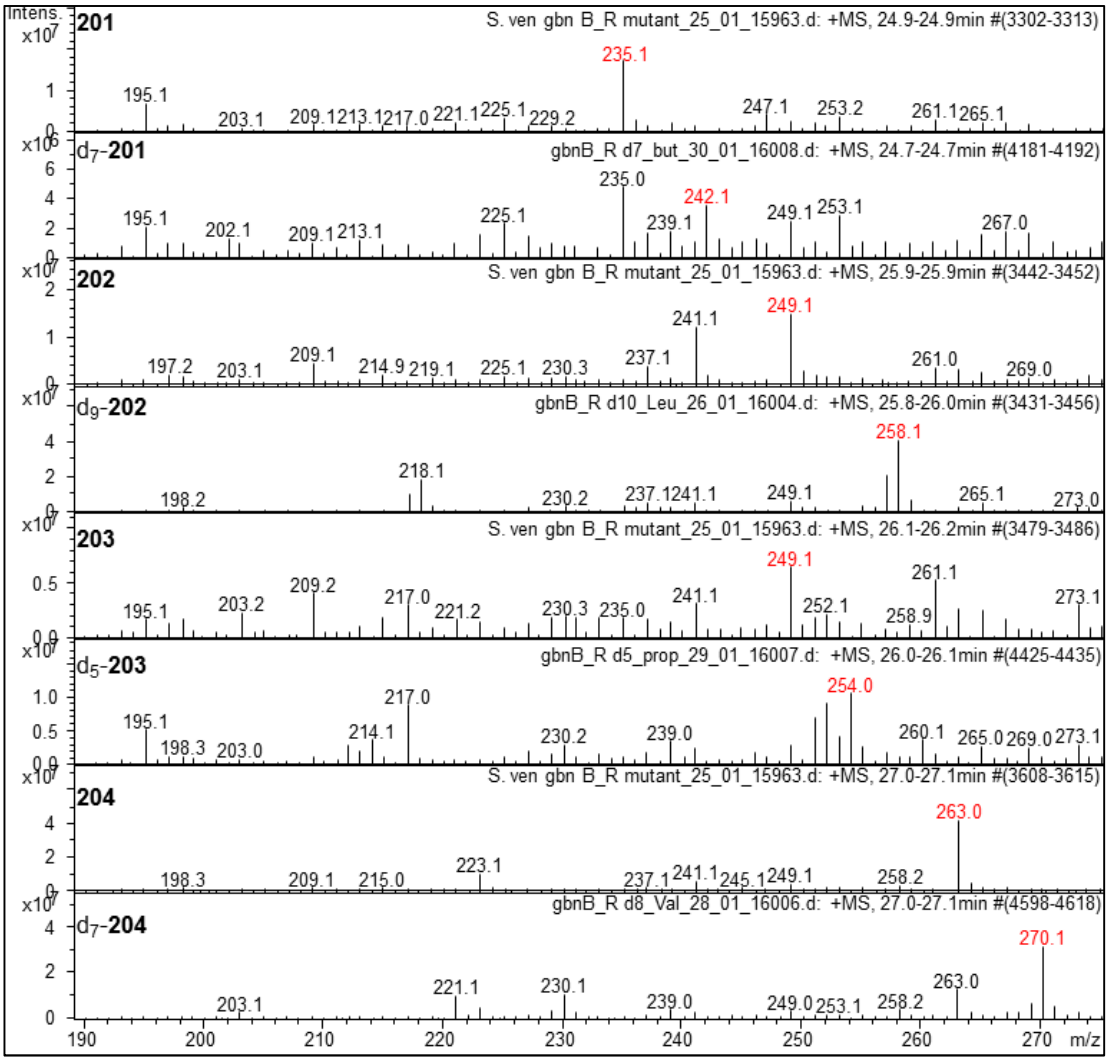


Figure 6.8 – Mass spectra of the AHFCAs 5-8 (**201-204**) identified in the *S. venezuelae gbnR* deletion mutants when grown on modified SMMS. AHFCA5 is labelled with 7 deuteriums from d_7-butyric acid; AHFCA6 is labelled with 9 deuteriums from d_{10}-leucine; AHFCA7 is labelled with 5 labels from d_5-propionic acid; and AHFCA8 is labelled with 7 deuteriums from d_8-valine. $[M+Na]^+$ ions are highlighted in red

The compound **204** with t_R = 27.0 min was named AHFCA8 and shown to be specifically labelled by d_8-valine (Figure 6.8). Molecular formulae for the deuterated species were all confirmed by UHR-MS analyses (Table 6.2 and Appendix 6), as exemplified by AHFCA8 in Figure 6.9. All four AHFCAs have an absorption maximum at 254 nm, as previously reported for AHFCAs.[79]

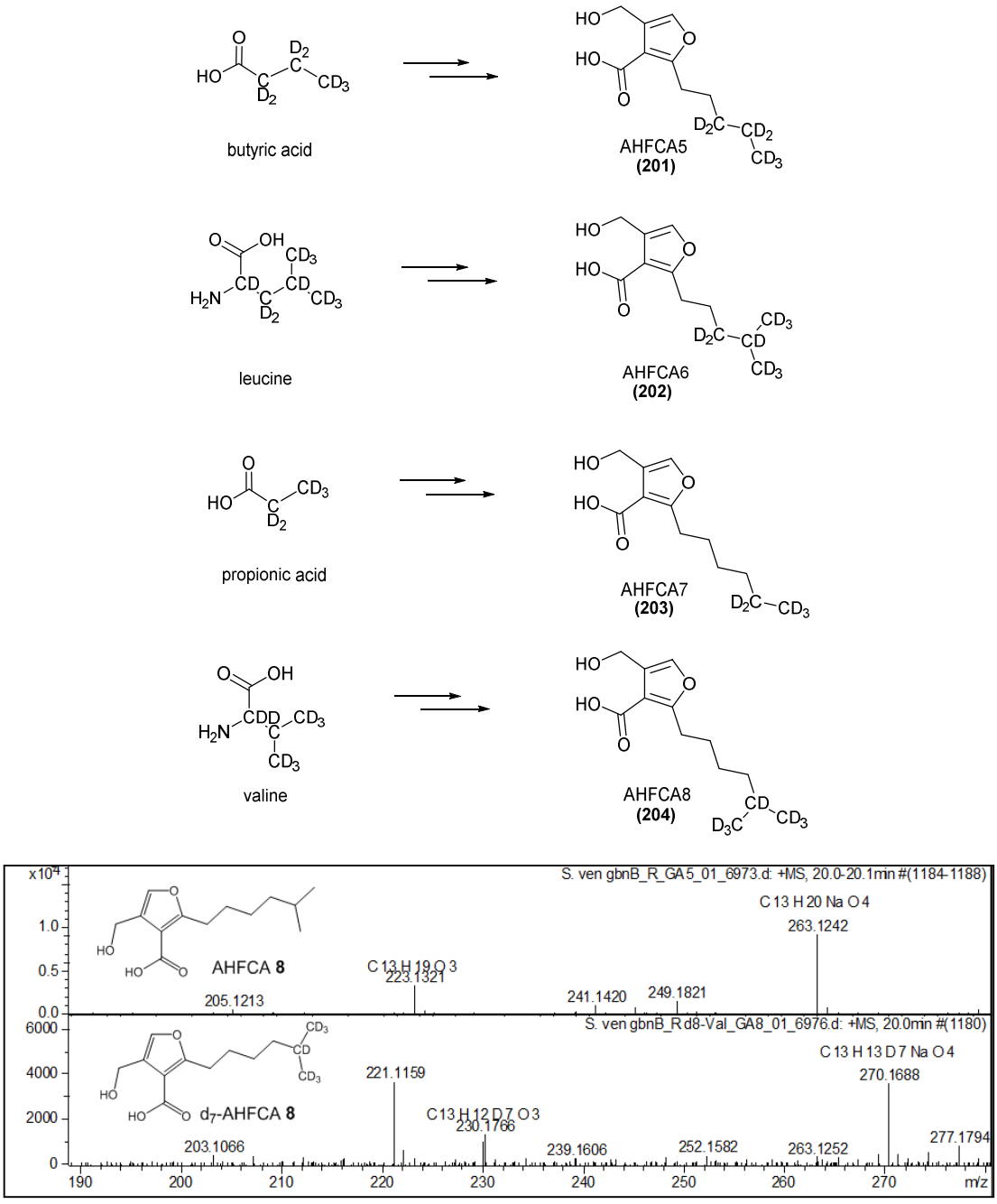


Figure 6.9 – AHFCAs produced by *S. venezuelae gbnR* deletion mutants as demonstrated by incorporation of the specific precursor molecules shown, and UHR-MS spectra of AHFCA8 (**204**) unlabelled (top panel) and AHFCA8 labelled with 7 deuteriums from incorporation of d_8-valine (bottom panel) and molecular formulae generated for the $[M\text{-}H_2O+H]^+$ and $[M+Na]^+$ and ions are shown

Upon feeding with 10 mM d_{10}-isoleucine and d_{10}-leucine, respectively, there were two new AHFCA-like metabolites **205** and **206**, with ions corresponding of *m/z* = 246.0 and 286.0 in their mass spectra, as would be expected for incorporation of nine deuterium

Table 6.2 – UHR-MS of $[M+Na]^+$ ions of AHFCAs 5-10 identified in culture extract of *S. venezuelae gbnB/gbnR* grown on SMMS supplemented with different stable isotope precursor molecules

retention time / min	*compound*	*precursor molecule fed*	*Molecular formula*	*Observed* m/z	*Calculated* m/z
19.4	AHFCA5 (**201**)	none	$C_{11}H_{16}NaO_4$	235.0936	235.0941
19.4	d_7-AHFCA5 (**201**)	d_7-butyric acid	$C_{11}H_9D_7NaO_4$	242.1377	242.1386
19.7	AHFCA6 (**202**)	none	$C_{12}H_{18}NaO_4$	249.1089	249.1087
19.7	d_9-AHFCA6 (**202**)	d_{10}-L-leucine	$C_{12}H_9D_9NaO_4$	258.1651	258.1668
19.8	AHFCA7 (**203**)	none	$C_{12}H_{18}NaO_4$	249.1089	249.1087
19.8	d_5-AHFCA7 (**203**)	d_5-propionic acid	$C_{12}H_{13}D_5NaO_4$	254.0*	254.1417
20.1	AHFCA8 (**204**)	none	$C_{13}H_{20}NaO_4$	263.1242	263.1254
20.1	d_7-AHFCA8 (**204**)	d_8-DL-valine	$C_{13}H_{13}D_7NaO_4$	270.1688	270.1699
20.5	d_9-AHFCA9 (**205**)	d_{10}-L-isoleucine	$C_{14}H_{13}D_9NaO_4$	286.1969	286.1981
20.5	d_9-AHFCA10 (**206**)	d_{10}-L-leucine	$C_{14}H_{13}D_9NaO_4$	286.1*	286.1981

* intensity of peak in UHR-MS too low to generate molecular formula

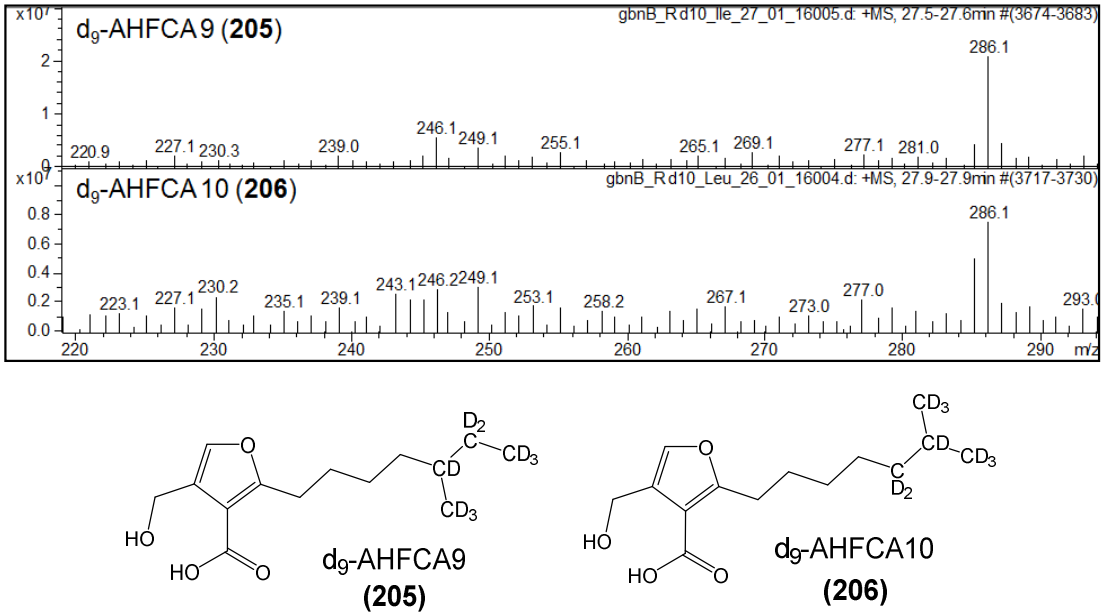


Figure 6.10 – Mass spectra and proposed structures of **205** produced upon feeding the *gbnB/R* strain with d_{10}-isoleucine (top panel) and **206** produced upon feeding the *gbnB/R* strain with d_{10}-leucine (bottom panel)

labels from d_{10}-isoleucine and d_{10}-leucine, respectively (Figure 6.10). These AHFCAs were named AHFCA9 (isoleucine-derived) and AHFCA10 (leucine-derived).

In order to further verify the structures of **201**-**203** as corresponding to AHFCAs 5-7 identified in the culture extract of the *S. venezuelae gbnB/R* strain, LC-MS data were obtained in which the metabolites extracted from this strain were co-injected with authentic standards of AHFCAs 5, 6 and 7 (synthesised by N. Malet and S. Zhou in Prof. G. L. Challis group, Warwick).[100] As shown by Figure 6.11, compounds **201**, **202** and **203** co-elute with AHFCAs 5, 6 and 7 respectively.

The discovery that AHFCAs are produced by *S. venezuelae* represents the first example of AHFCAs being produced by species different from *S. coelicolor*. Interestingly, one of these AHFCAs – AHFCA5 has an identical structure to one of the AHFCAs initially isolated from *S. coelicolor*, but the other AHFCAs produced by *S. venezuelae* have longer alkyl chains. The AHFCAs 6-10 are novel natural products, although AHFCA7

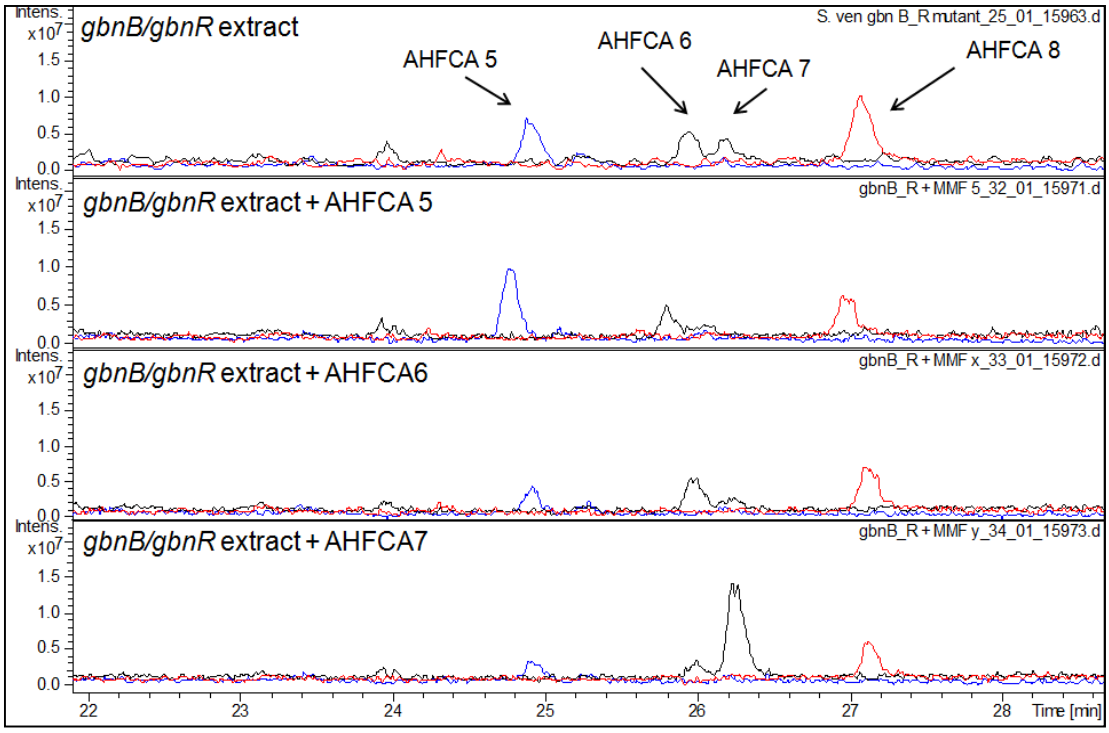


Figure 6.11 – Extracted ion chromatograms for $[M\text{-}H_2O+H]^+$ ions of AHFCA species present in the metabolites extracted from the *S. venezuelae gbnB/gbnR* with no AHFCA added and with AHFCAs 5, 6 and 7 co-injected. Chromatograms correspond to *m/z* = 195.0 (blue traces), 209.0 (black traces) and 223.0 (red traces)

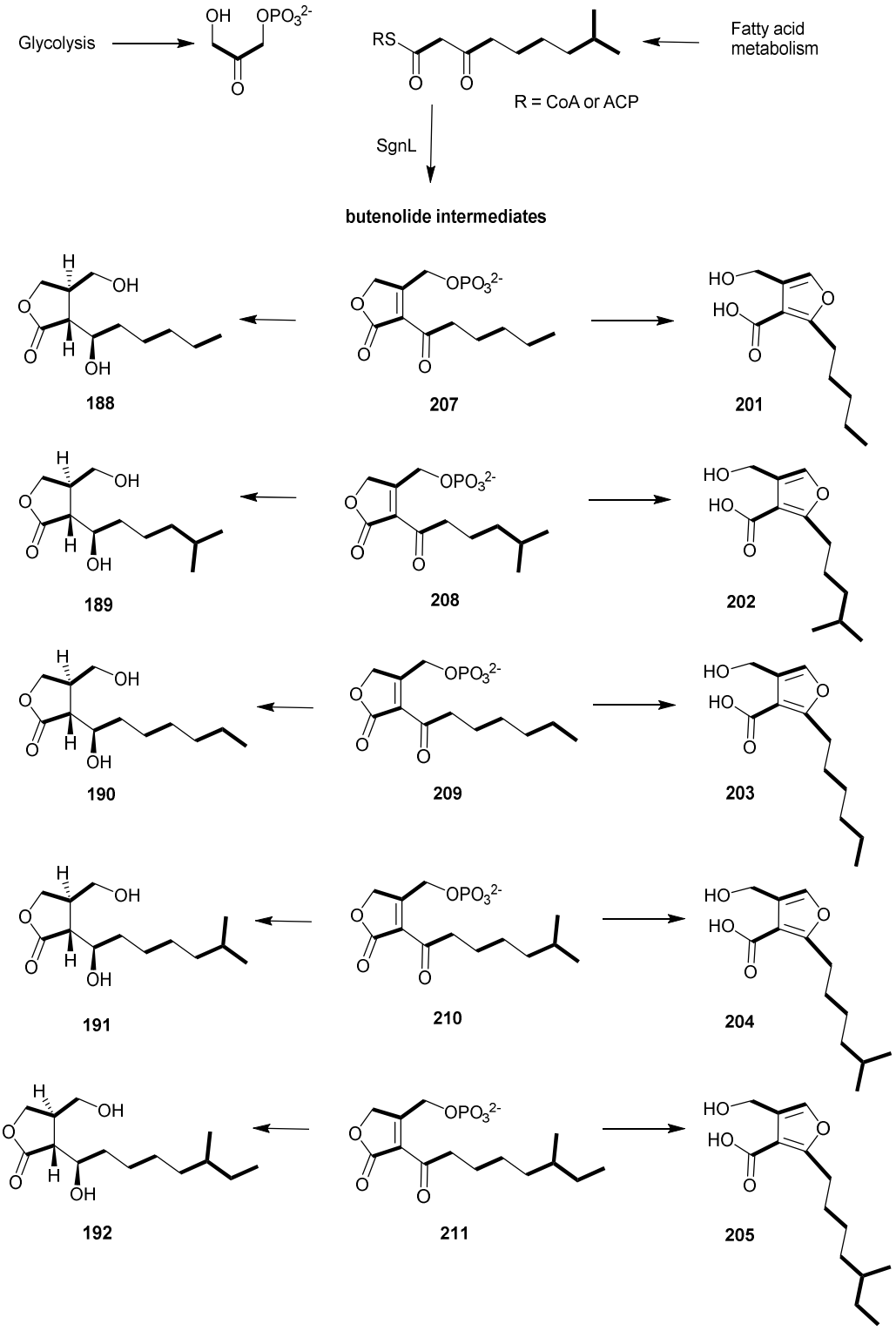
has been previously synthesised when the ligand tolerance of MmfR was investigated (see Section 7.2).[100]

6.1.5 Deductions about SCB/AHFCA biosynthesis in *S. venezuelae*

Interestingly, the GBLs and AHFCAs shown to be produced by the *S. venezuelae gbnR* deletion mutants are both derived from the same precursor molecules (Scheme 6.3). Indeed, upon feeding with the deuterium labelled precursors d_7-butyric acid, d_{10}-leucine, d_5-propionic acid, d_8-valine and d_{10}-leucine, there was overproduction of compounds **188**, **189**, **190**, **191** and **192** respectively in the *S. venezuelae gbnB/R* mutant, consistent with the structures proposed for these molecules in Section 6.1.3.

This result is consistent with the previously reported hypothesis that *scbA*, *mmfL* and their homologues produce butenolide phosphate intermediates, and the downstream enzymatic steps determine the final structure of GBL and AHFCA signalling molecules (Scheme 6.2). The observation that exactly the same precursors appear to be utilised for GBL and AHFCA biosynthesis in *S. venezuelae gbnR* deletion mutant might suggest that SgnL is the shared enzyme in the biosynthesis of these specific signalling molecules.

It is possible that the deletion of *gbnR* – and therefore de-repression of transcription of *sgnL* – leads to overproduction of the butenolide intermediates to such an extent that the intracellular concentration is too great for the downstream enzymes SgnP and SgnH to turn over, leading to production of gamma-butyrolactones as a series of shunt metabolites. Indeed, the potential for cross-talk between the *sgnL* and *jadW1* pathways cannot be excluded, as SCB3 was recently reported as being produced a *S. venezuelae* mutant in which *jadW1* had been deleted. Upon deletion of *jadW1* and *sgnL*, there was



Scheme 6.3 – Proposed structures of the GBLs and AHFCAs produced by *S. venezuelae gbnR* deletion mutants. The same series of butenolide phosphate intermediates **207**-**211** are proposed to be utilized in GBL and AHFCA biosynthesis

still some SCB3 produced. However, upon construction of a *jadW1*, *sven_5093* double deletion mutant, SCB3 production was abolished.[95] This observation implies that *sgnL*

is not responsible for SCB biosynthesis in *S. venezuelae*, therefore the presence of SCBs in *S. venezuelae gbnR* mutant extracts may be due to other *afsA* homologues.

Furthermore, the observation that the SCBs and AHFCAs observed in the *S. venezuelae gbnR* mutants are derived from the same precursor molecules, may simply reflect the availability of β-ketothioesters with different alkyl substituents which are available in the cell to be scavenged from fatty acid biosynthesis by SgnL, or by JadW1 or Sven_5093. Alternatively, if the AREs in the *sgn* and *jad* clusters are very similar, it is possible that the GbnR repressor binds to both. Therefore upon deletion of *gbnR* there may be some derepression of transcription of *jadW1*, leading also to SCB biosynthesis.

The observation that *S. venezuelae gbnR* mutants produce AHFCA as well as GBL signalling molecules raises the question of whether SCBs or AHFCAs control the biosynthesis of gaburedins, or whether the presence of both AHFCAs and SCBs in the *gbnR* deletion mutants is the result of crosstalk between different signalling pathways. In order to determine the nature of the signalling molecules (GBLs or AHFCAs) produced by SgnL, the butenolide synthase gene *sgnL* was heterologously expressed in *E. coli* to determine what profile of different AHFCAs and/or GBLs are produced by an *E. coli* strain overexpressing *sgnL*.

6.2 Production of *S. venezuelae* signalling molecules in *E. coli*

6.2.1 Overexpression of *sgnL* in *E. coli*

When the AHFCAs produced by *S. coelicolor* were first characterised, it was shown that heterologous expression of *mmfL* alone was sufficient to lead to production of AHFCAs in *E. coli*.[79] Therefore it was hypothesised that the heterologous expression of

the *mmfL* homologue from *S. venezuelae, sgnL*, would also lead to production of AHFCAs. However, if GBLs were produced as well, this would account for the above observation that both AHFCAs and GBLs are produced by *S. venezuelae* when *gbnR* is deleted.

The biosynthetic gene *sgnL* was amplified by PCR from the SV4 F01 cosmid containing the whole *gbn-sgn* gene cluster. Specific primers S/S' were designed to amplify the whole *sgnL* gene and add a 5'-CACC tag required for cloning into the pET151 expression vector (TOPO cloning is discussed in Section 5.1). Successful colonies were identified by selection with ampicillin, followed by preparation of overnight cultures, purification of their plasmids and subsequent digestion with *Pst*I.

Upon verifying the sequence of the *sgnL* gene insert, the construct pJDS08 was used to transform *E. coli* BL21star and protein expression induced using IPTG. In parallel, this *E. coli* BL21star/pJDS08 strain was grown without the addition of IPTG to the growth medium. Metabolites produced by this strain were compared by LC-MS, leading to the identification of several metabolites **212-214** overproduced in the *E. coli* BL21star/pJDS08 induced with IPTG (Figure 6.12).

Molecules **212** and **213** with the *m/z* values and fragmentation patterns expected for the AHFCAs 2 and 5 ($[M\text{-}H_2O+H]^+$ ions with *m/z* = 167.0 and 195.0, respectively) were identified. Comparison of these metabolites with an authentic standard of AHFCA2 confirmed the retention times and fragmentation of AHFCA2 authentic standard and **212** extracted from *E. coli* BL21star/pJDS8 to be the same, implying that **212** is AHFCA2 (compound **32**).

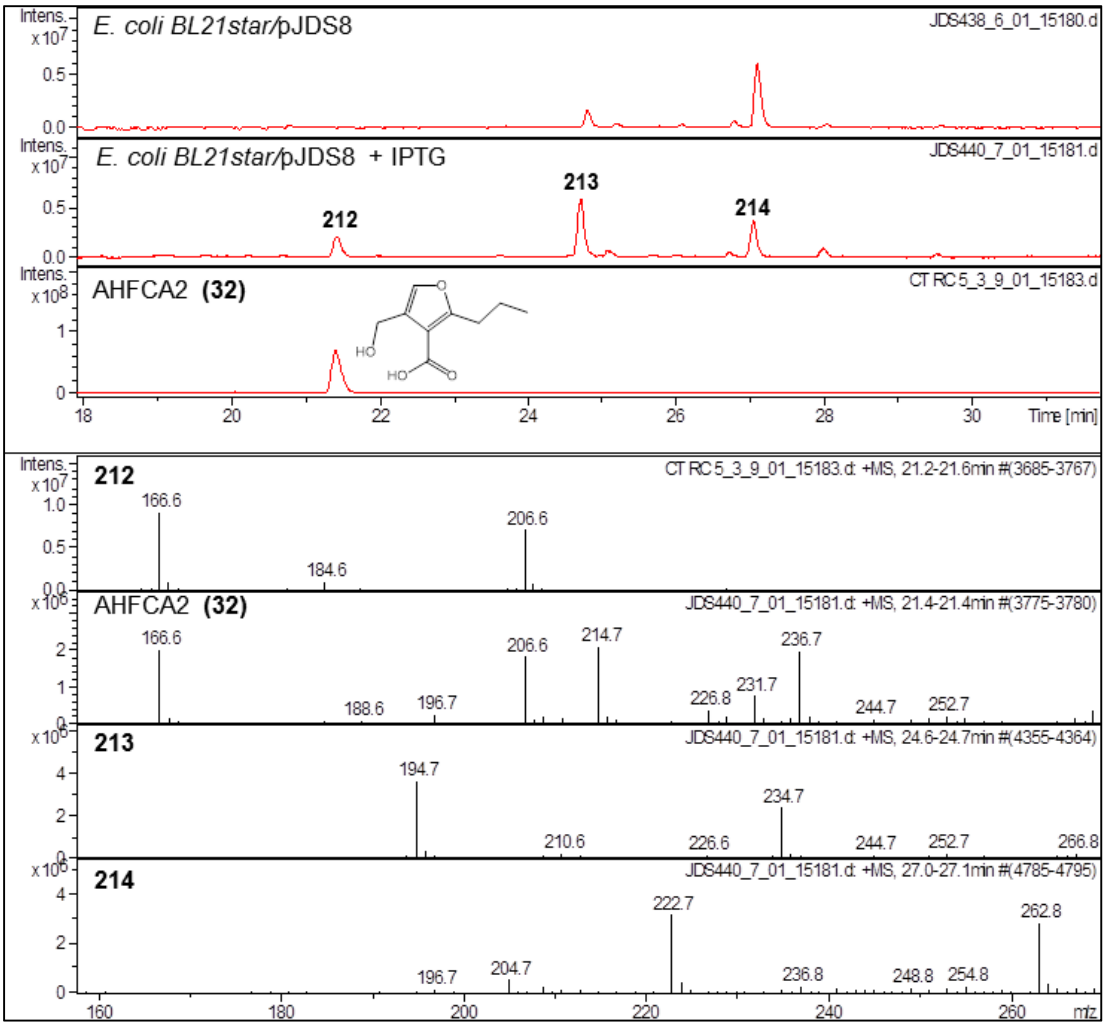


Figure 6.12 **–** Extracted ion chromatograms *m/z* = 167.0, 195.0 and 223.0 for $[M\text{-}H_2O+H]^+$ ions of **212-214** present in the metabolites extracted from the *E. coli* BL21star/pJDS8 strain and *E. coli* BL21star/pJDS8 strain induced with IPTG. A synthetic standard of AHFCA2 (**32**) was run in parallel. Mass spectra of AHFCA **212-214** are also shown

Furthermore, an additional peak with *m/z* = 262.8 and the same characteristic AHFCA fragmentation pattern was also observed (Figure 6.12). UHR-MS analysis confirmed this new AHFCA-like compound **214** to have the same molecular formulae that would be predicted for an AHFCA of this size: $C_{13}H_{20}NaO_4$ for the $[M+Na]^+$ ion (calculated *m/z* = 263.1254, observed 263.1245), $C_{13}H_{21}O_4$ for the $[M+H]^+$ ion (calculated *m/z* = 241.1434, observed 241.1419) and $C_{13}H_{19}O_3$ for the $[M+H\text{-}H_2O]^+$ ion (calculated *m/z* = 223.1329, observed 223.1321). The retention time at which this metabolite elutes is also consistent with the retention time that one would expect for a larger molecule

belonging to this same series (AHFCAs with longer alkyl chains would be expected to be retained on the column for longer).

It can be deduced that this new AHFCA **214** is distinct from the AHFCA8 produced by *S. venezuelae*, as AHFCA8 is derived from valine. However, *E. coli* is known to only produce unbranched fatty acids (the predominant starter unit in fatty acid biosynthesis in *E. coli* is acetyl-CoA), therefore the AHFCAs produced would be expected to have no branching in their alkyl chains, either.[79,136] Hence this new compound **214** was named AHFCA8E, the E to signify it is only produced by SgnL when overproduced in the heterologous host *E. coli*.

The observation that *E. coli* carrying *sgnL* is able to produce AHFCA2 is also of interest, as no AHFCA2 has been observed to be produced by the native host *S. venezuelae gbnR*. This difference in the AHFCA profiles of *E. coli* BL21star/pJDS08 compared with the *S. venezuelae gbnR* mutants would imply that the substrate specificity of SgnL is sufficiently broad that shorter β-ketothioester intermediates than can be tolerated by SgnL, to produce smaller butenolides than those shown in Scheme 6.3. Incorporation of the shorter β-ketothioester to allow AHFCA2 be produced by *E. coli* compared to *S. venezuelae* may simply reflect the intracellular availability of different β-ketothioesters in *E. coli* and *S. venezuelae*.

6.2.2 Implications for AHFCA/GBL biosynthesis

The observation that only *sgnL* is required for AHFCA biosynthesis in *E. coli* complements the previously published work in which expression of *mmfL* alone resulted in AHFCA production.[79] Interestingly, no GBLs (neither of the 6-keto or 6-hydroxy types) were observed in the culture extract of *E. coli* BL21star/pJDS08.

However, as the heterologous host *E. coli* contains an MmfH homologue (acyl-CoA dehydrogenase, NCBI accession no. WP_032718200.1 with 26% identity and 41% similarity to MmfH at the amino acid level), this could be the reason for the production of AHFCAs instead of GBLs. If the presence of an *mmfH* homologue in *E. coli* was the only reason for production of AHFCAs, then one would expect that expression of other *afsA* homologues in *E. coli* might also lead to AHFCA (rather than GBL) production.

However, upon heterologous expression of *afsA* from *S. griseus* in *E. coli*, Kato and co-workers noted that 6-keto GBL analogues of A-factor **71** and **72** were produced (Figure 1.17). If the MmfH homologue acyl-CoA dehydrogenase acts on the butenolide phosphate formed (as suggested in the case of heterologous expression of *mmfL* in *E. coli*) then one would expect AHFCAs to be produced by *E. coli* carrying other *afsA*. The observation that *afsA* alone in *E. coli* is sufficient to produce A-factor implies that the downstream enzyme required (BprA or a homologue) is present in *E. coli*. Indeed, a BLAST search of BprA homologues in *E. coli* revealed that BprA has significant homology to quinine reductase 2 (NCBI accession no. WP_016159986.1, which has 30% identity, 46% similarity with BprA at the amino acid level).

As there are homologues of both BprA and MmfH in *E. coli*, when expressing *afsA* homologues (therefore producing butenolide intermediates) then one might expect both AHFCAs and GBLs to be produced in *E. coli*. However, if the BprA or MmfH homologues are specific for substrates with different length alkyl chains, this may account for the different signalling molecules observed when expressing *sgnL*, compared with expressing *afsA* in *E. coli*. On the other hand, the result that only AHFCA signalling molecules are produced by *E. coli* overexpressing *sgnL* could imply that *sgnL* catalyses a different reaction to other *afsA* homologues. Hence, the signalling

molecule (GBLs or AHFCAs) resulting from the reaction of the *afsA* homologue may be determined at least in part by the AfsA homologue itself (rather than the downstream reactions proposed to be catalysed by MmfH/P and BprA to yield AHFCAs or GBLs, respectively).

The result that only *sgnL* is required for AHFCA biosynthesis lends support to the hypothesis that AHFCAs are produced by the SgnLPH system in *S. venezuelae*, yet the ligands which actually induce gaburedin biosynthesis had not yet been confirmed. Therefore, bioinformatic analyses to compare the AHFCA binding site in MmfR to the putative binding site in SgnR were performed, as a high degree of similarity of the MmfR and SgnR binding pockets would provide evidence for SgnR being an AHFCA-responsive protein.

6.3 Comparison of SgnR and MmfR ligand binding sites

The ligand binding tolerances of ArpA and homologues BarA and FarA been studied. Generally, these studies have focused on previously characterised signalling molecules (to determine whether a GBL receptor responds to 6-keto or 6-hydroxy GBLs with different stereochemistry at carbon 6).[112,142,126] Of the MmfR sub-family of ArpA-like transcriptional repressors identified, only MmfR has been studied extensively *in vitro* and *in vivo*. More recently, the crystal structure of MmfR has been solved, both in its *apo* form and its ligand-bound *holo* form with the ligand AHFCA2 (**32**) bound.[100] Up until this point, MmfR is the only ArpA-like repressor known to respond to AHFCA-like signalling molecules.

Figure 6.13 – Clustal Omega alignments of SgnR (SVEN_4182), MmfR and homologues in other *Streptomyces* species. The amino acid residues identified as interacting with the AHFCA ligand are highlighted in yellow. The key tyrosine residues identified are highlighted in cyan

In order to provide *in silico* evidence for SgnR also being able to bind to the same AHFCA-type ligands (rather than GBLs), sequence alignments of MmfR and SgnR were produced using Clustal Omega (Figure 6.13) and the active site residues were compared. The tyrosines 85 and 144 are conserved throughout the SgnR homologues identified, yet there is some variation in the amino acid residues interacting with the alkyl side chain of the AHFCA ligand (Figure 6.13). Amino acid residues of MmfR

identified to be interacting with AHFCA ligand were compared with the amino acid residue in the same positions in SgnR. Interestingly, the relative positions of the amino acids interacting with AHFCA2 in the ligand binding pocket are 100% conserved between the MmfR and SgnR sequences. This provides evidence for AHFCAs – rather than GBLs – as being the ligands interacting with SgnR in the *S. venezuelae gbn-sgn* system.

To provide more evidence for the potential for AHFCAs to bind in the same active site in SgnR as in MmfR, a simulated crystal structure for SgnR was generated using Phyre2 and the solved MmfR-AHFCA2 structure as a template.[184] As shown in Figure 6.14, the amino acids in the active site of the structure generated for SgnR align almost

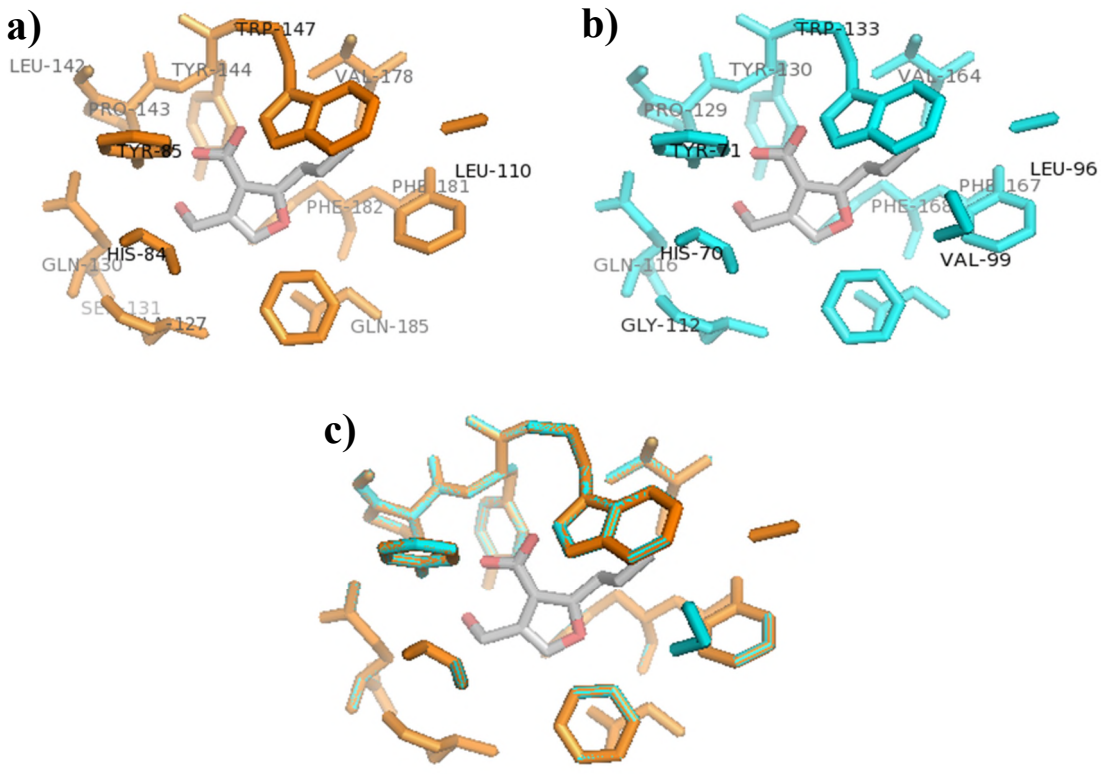


Figure 6.14 – a) Active site residues of MmfR interacting with AHFCA2 from solved crystal structure[100] b) Active site residues of SgnR proposed to interact with AHFCA2 from SgnR structure generated using Phyre2 c) overlaid active sites residues of MmfR (orange) and SgnR (cyan) showing that the active site residues are predicted to be conserved between MmfR and SgnR. Images generated using PyMOL with views restricted to 6Å around the AHFCA2 ligand

perfectly with those amino acids of MmfR shown to interact with AHFCA2, providing evidence that SgnR is likely to be furan-responsive. However, experimental evidence to verify these *in silico* analyses has not been presented thus far.

6.4 Biological role of *S. venezuelae* AHFCA signalling molecules

6.4.1 AHFCAs induce gaburedin production in an *S. venezuelae sgnL-P* mutant

To establish if particular AHFCA or GBL signalling molecules could induce gaburedin biosynthesis, an induction assay was devised in which the *sgnL-P::apra* mutant (unable to produce AHFCAs) was fed with a series of culture extracts containing AHFCAs/GBLs, in order to establish whether gaburedin biosynthesis could be induced in the *sgnL-P::apra* strain. This chemical complementation approach could be an effective way to determine structure-activity relationships, and has been used to assay induction of methylenomycin biosynthesis in *S. coelicolor* W81, a strain which lacks *mmfLHP*, but the rest of the methylenomycin cluster is intact.[79,100] In contrast to the activity-based methylenomycin induction assay previously reported, gaburedin production was analysed by LC-MS.

Initial assays were carried out in which culture supernatant of the *S. venezuelae gbnB/R* double mutant (which produces AHFCAs and GBLs) was added to the *S. venezuelae sgnL-P* mutant, however no inductive effect was observed. SMMS enriched with pure authentic standards AHFCAs 6 and 7 were also prepared in which AHFCAs were dissolved in DMSO to a concentration of 10 mM and then added to 5 mL re-melted SMMS to give final concentrations of 100 μM, 10 μM and 1 μM. It was noted however that DMSO appeared to restrict the growth of the *S. venezuelae sgnL-P::apra*, therefore

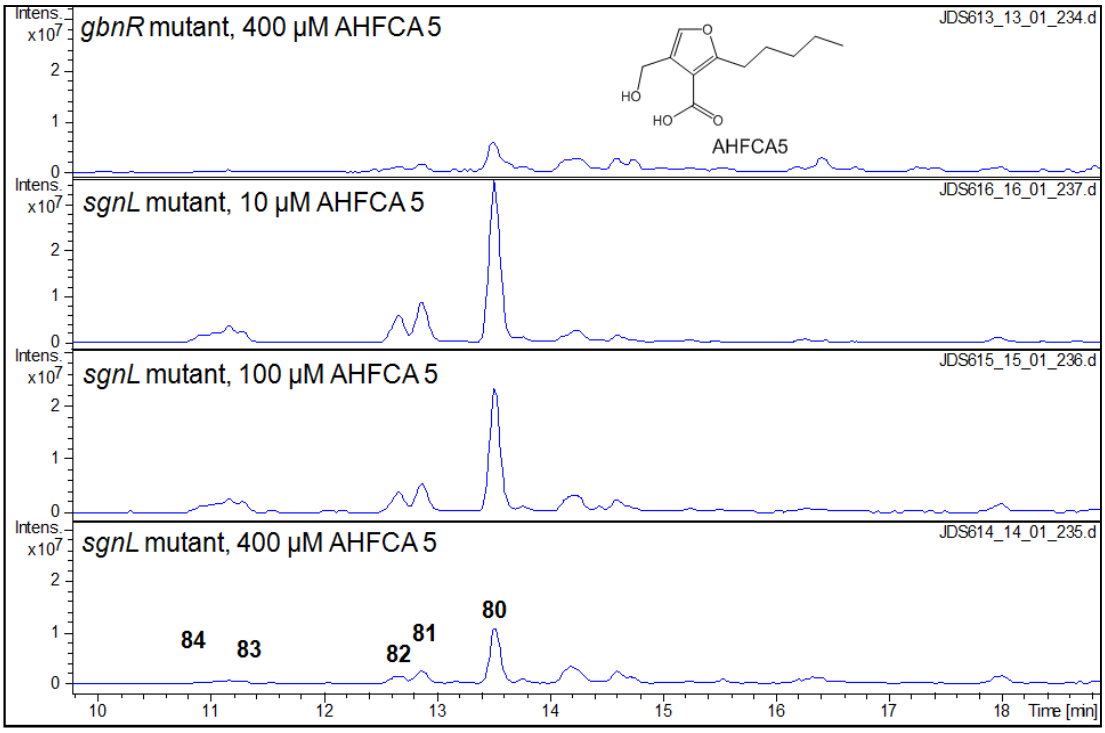


Figure 6.15 – Extracted ion chromatograms *m/z* = 292.9, 279.0, 247.0, 261.0, 249.9 for gaburedins A-F (**80-85**) in metabolites extracted from *S. venezuelae sgnL-P::apra* strain upon addition of 10 µM, 100 µM and 400 µM AHFCA5

a much more concentrated solution of each AHFCA was made to minimise the quantity of DMSO which was required to be added to each plate.

Therefore, stock solutions of AHFCAs 5, 6 and 7 were prepared, and then diluted with DMSO so that 2 µL of each diluted solution was added to each 5 mL SMMS to give final AHFCA concentrations of 400 µM, 100 µM, and 10 µM. LC-MS analysis of culture extracts of the *S. venezuelae sgnL-P::apra* mutant grown on these media for 3 days revealed that gaburedin biosynthesis was induced at all concentrations (Figure 6.15). These experiments demonstrate that the inducing molecules in gaburedin biosynthesis are indeed AHFCAs, and provide the first example after the methylenomycin system in which AHFCAs have been shown to control biosynthesis of a natural product.

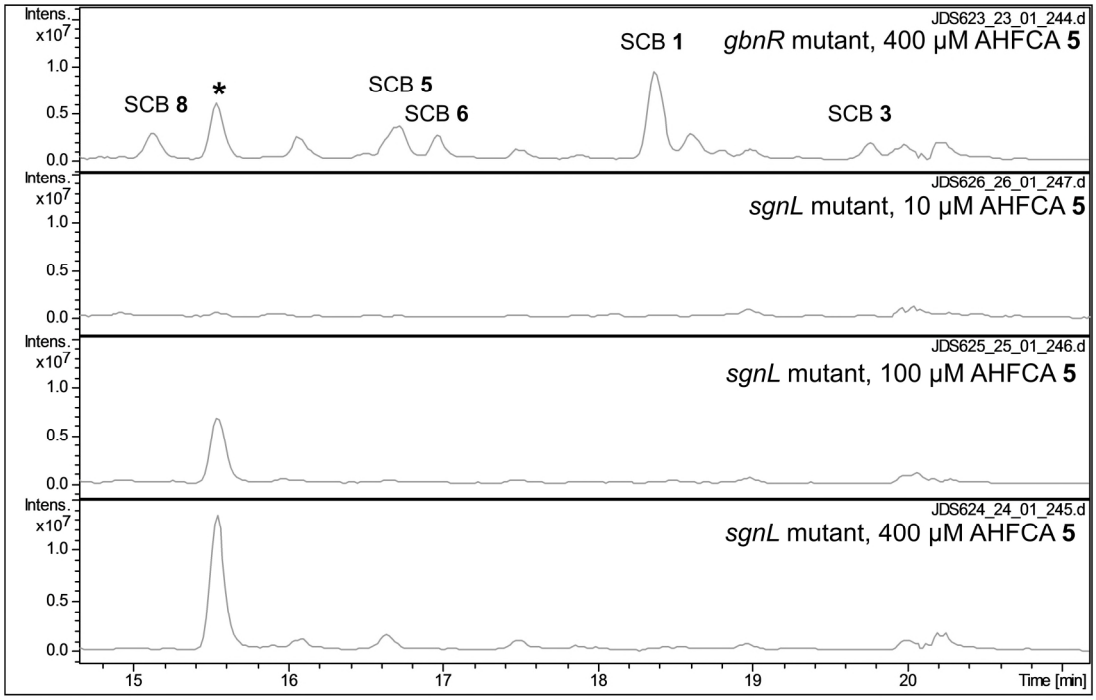


Figure 6.16 **–** Extracted ion chromatograms *m/z* = 239.0, 253.0, 267.0, 281.0 for SCBs in metabolites extracted from *S. venezuelae sgnL-P::apra* strain upon addition of 10 µM, 100 µM and 400 µM AHFCA5, with a positive control experiment adding 400 µM AHFCA5 to the *gbnR* mutant. The compound denoted by the * has *m/z* = 267 and is unrelated to SCBs

It was also noted during these experiments that SCBs 1, 3, 5, 6 and 8 were still produced by the *gbnR* strain, but no SCBs were produced in the *sgnL-P::apra* strain induced with AHFCAs (Figure 6.16). Therefore, it can be deduced that gaburedins can be produced in the absence of SCBs. This result appears to confirm that AHFCAs – not SCBs – are the true inducer molecules in the *gbnABC-sgnLHP* system.

6.4.2 Effect of AHFCAs on *S. venezuelae gbnR* mutant and wild-type strains

Interestingly, the yields of gaburedins produced by the *sgnL-P::apra* strain upon addition of AHFCA 6 and 7 were similar for all three concentrations of AHFCAs 6 and 7 used. However, at a higher concentration of AHFCA5 (400 µM) there was actually less of an inductive effect than at 100 µM concentration, suggesting an upper limit on concentration of AHFCA required for inducing gaburedin biosynthesis. In control experiments in which the *gbnR* mutant (which contains the AHFCA biosynthetic genes

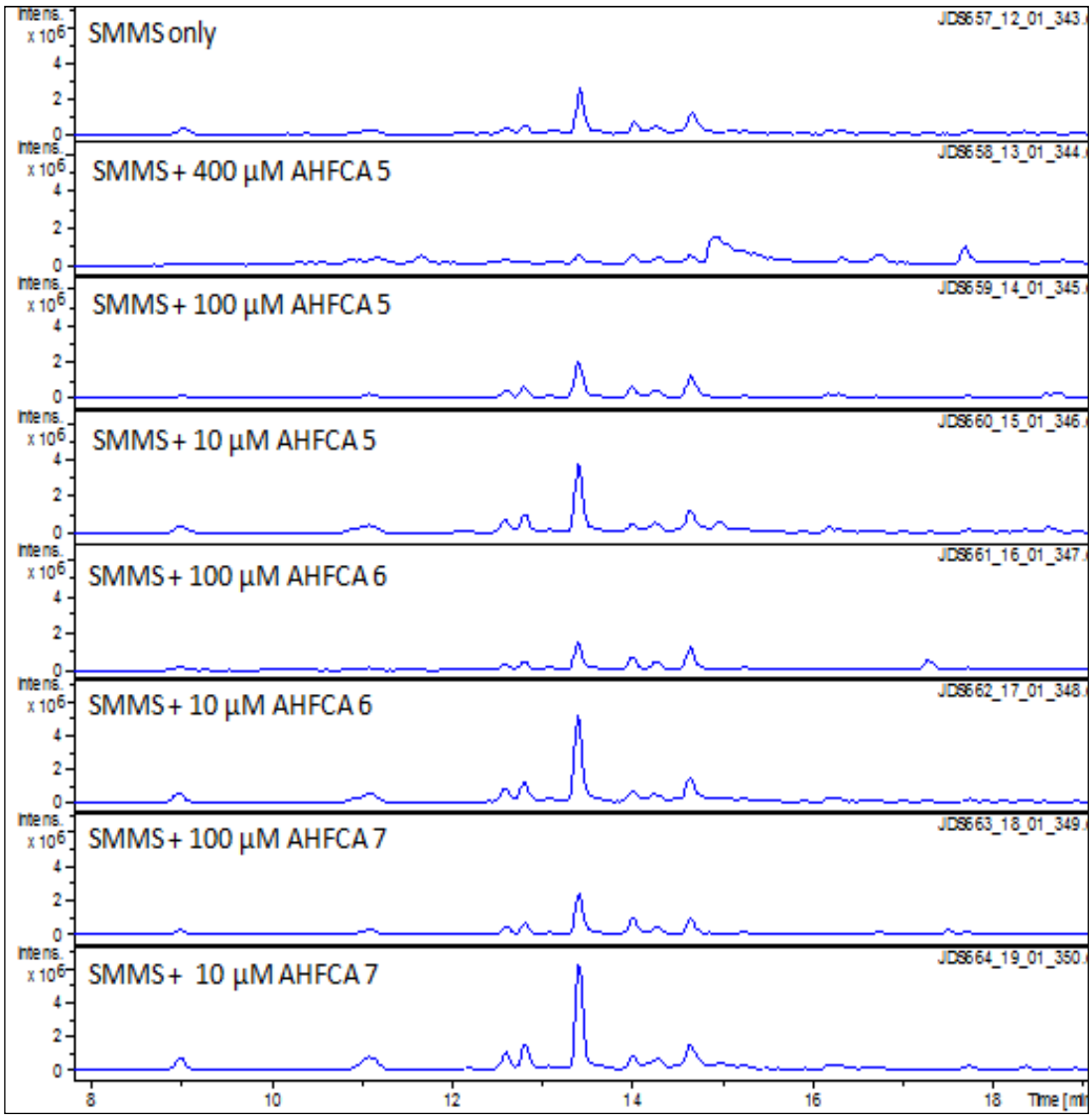


Figure 6.17 – Extracted ion chromatograms *m/z* = 292.9, 279.0, 247.0, 261.0, 249.9 for gaburedins A-F (**80-85**) in metabolites extracted from *S. venezuelae gbnR::apra* strain upon addition of different AHFCAs to the growth media

sgnL and *sgnP*, and overproduces gaburedins) was fed with 400 µM AHFCAs 5, 6 and 7, there were reduced yields of gaburedins compared with the yield of gaburedins induced in the *sgnL-P::apra* strain (Figure 6.15 illustrates this effect with AHFCA5), also implying that high concentrations of AHFCA may have an inhibitory effect on gaburedin biosynthesis.

Induction assays with AHFCAs were repeated with the *S. venezuelae sgnL-P*, *gbnR* and wild-type strains each grown on SMMS containing different concentrations of

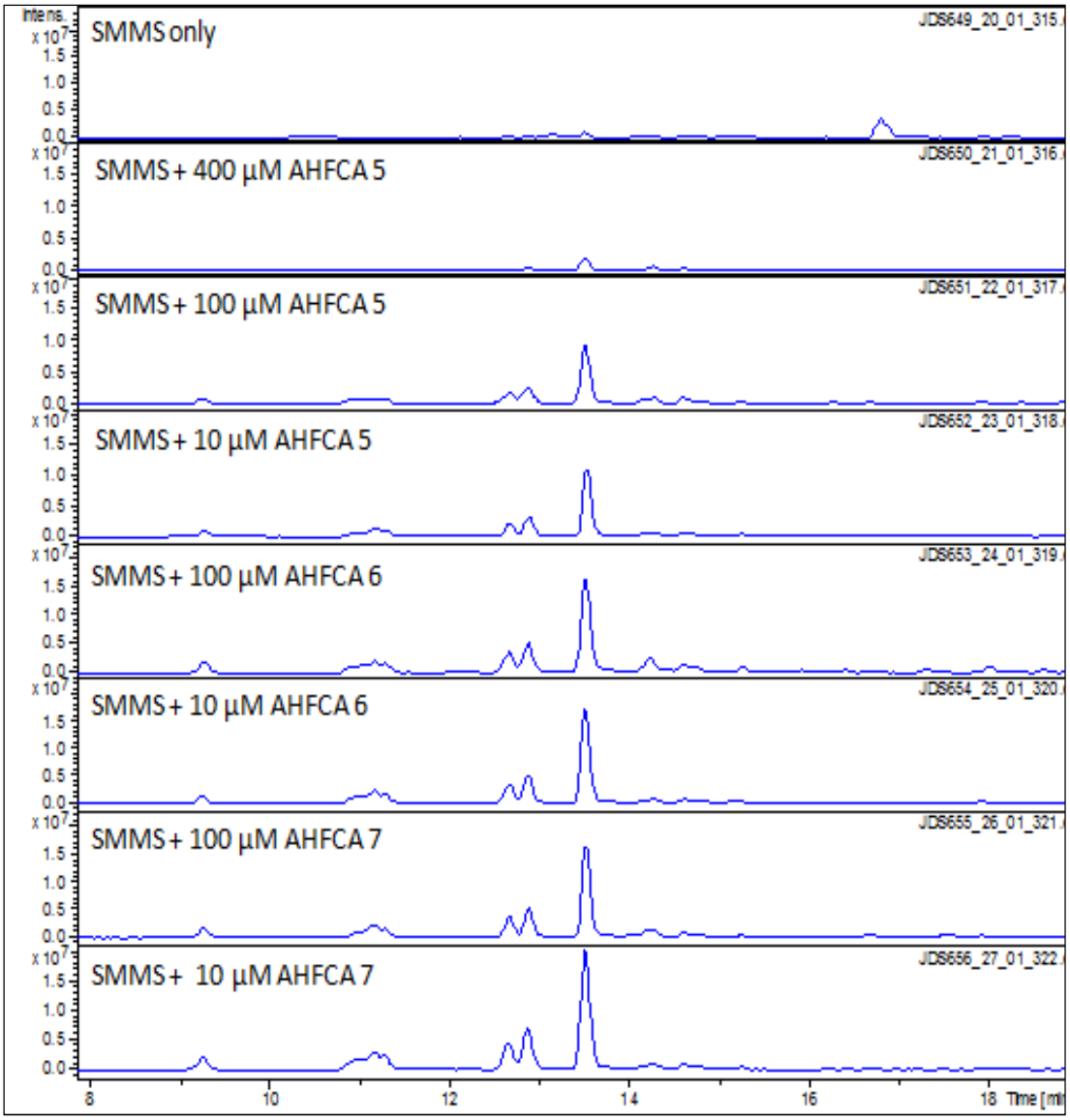


Figure 6.18 – Extracted ion chromatograms *m/z* = 292.9, 279.0, 247.0, 261.0, 249.9 for gaburedins A-F (**80**-**85**) in metabolites extracted from *S. venezuelae sgnL-P::apra* strain upon addition of different AHFCAs to the growth media

AHFCAs 5, 6 and 7. As shown in Figure 6.17, 10 µM AHFCAs 5-7 increased the yield of gaburedins produced by the *gbnR* mutant, compared to no AHFCA being added at all. AHFCA concentrations of 100 µM added gave yields similar to when no AHFCAs were added, and 400 µM concentrations of AHFCA5 lowered the yield of gaburedins produced by the *gbnR* strain, relative to the level of production without addition of AHFCAs.

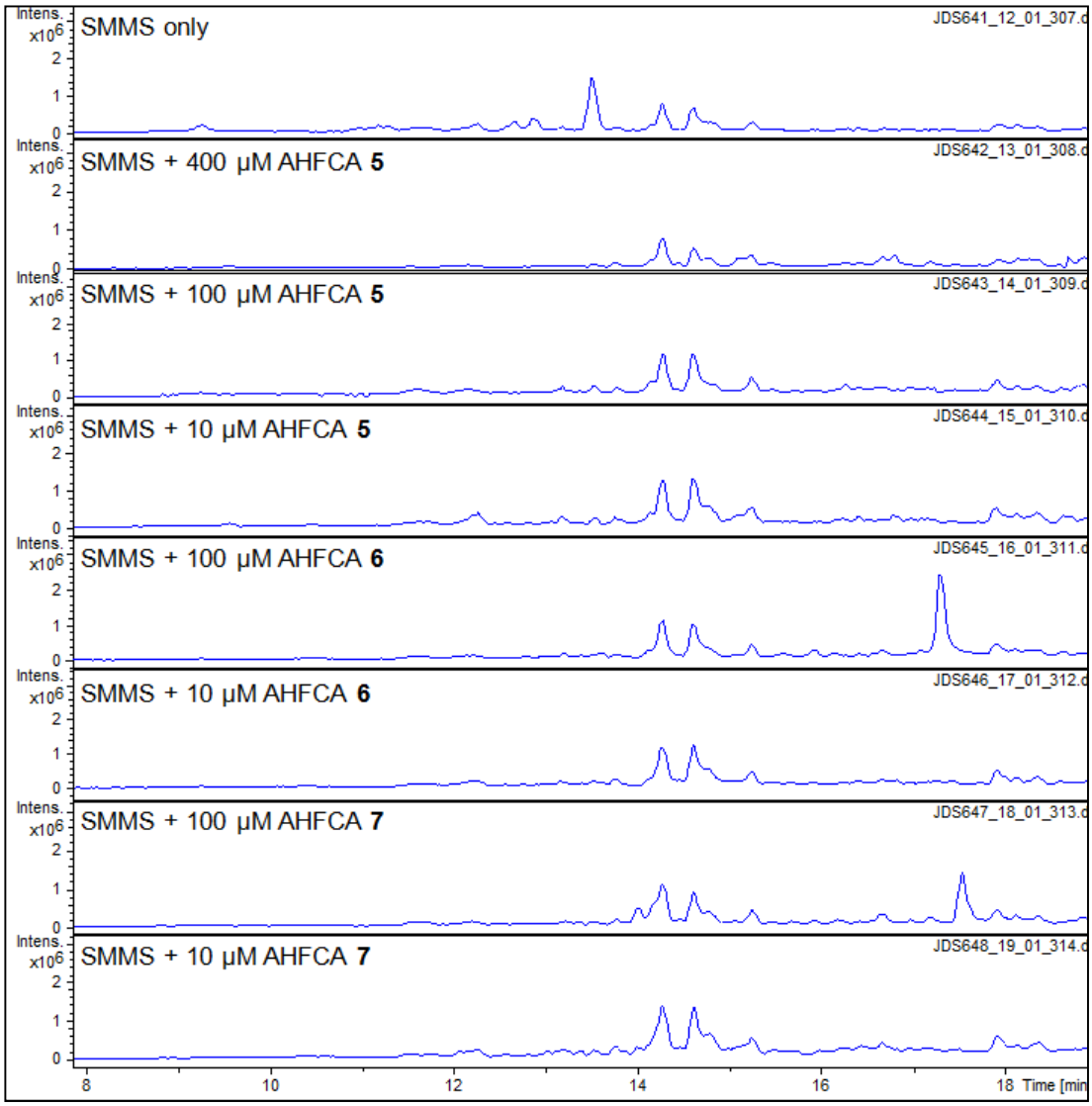


Figure 6.19 – Extracted ion chromatograms *m/z* = 292.9, 279.0, 247.0, 261.0, 249.9 for gaburedins A-F (**80**-**85**) in metabolites extracted from *S. venezuelae* wild-type strain upon addition of different AHFCAs to the growth media

In the *sgnL-P* strain, it was confirmed that addition of 10 µM and 100 µM of all three AHFCAs 5, 6 and 7 to growth media induced gaburedin production, with the yield of gaburedins being greatly reduced in the sample which had 400 µM AHFCA5 added (Figure 6.18). Additional control experiments in which the *S. venezuelae* wild-type was grown on SMMS containing the same AHFCA concentrations showed that in all conditions, gaburedin biosynthesis was completely abolished (Figure 6.19).

Taken together, these results suggest that there is a narrow range of concentrations at which AHFCAs can induce gaburedin biosynthesis (concentrations <10 µM are sufficient to abolish production in the wild-type, concentrations between 100 and 400 µM decrease production in the *gbnR* mutant, and concentrations of 400 µM decrease production in the *sgnL-P* mutant). Therefore, the wild-type strain would appear to be more sensitive to exogenous AHFCAs than the *sgnL-P* or *gbnR* mutants.

Based upon the hypothesis for the regulation of methylenomycin biosynthesis discussed in Section 1.6, an analogous hypothesis was made for the regulation of gaburedin biosynthesis (Section 2.1 and Figure 6.20). There is a basal level of *sgnL* transcription, leading to AHFCA biosynthesis. At a critical concentration, there is sufficient AHFCA present to be able to cause SgnR to dissociate from its' cognate AREs, leading to expression of *gbnA*, *sgnL* and *gbnR*. Therefore, gaburedin biosynthesis is activated. This is accompanied by a sharp increase in the intracellular concentration of AHFCAs due to expression of *sgnL*. Due to the increased expression

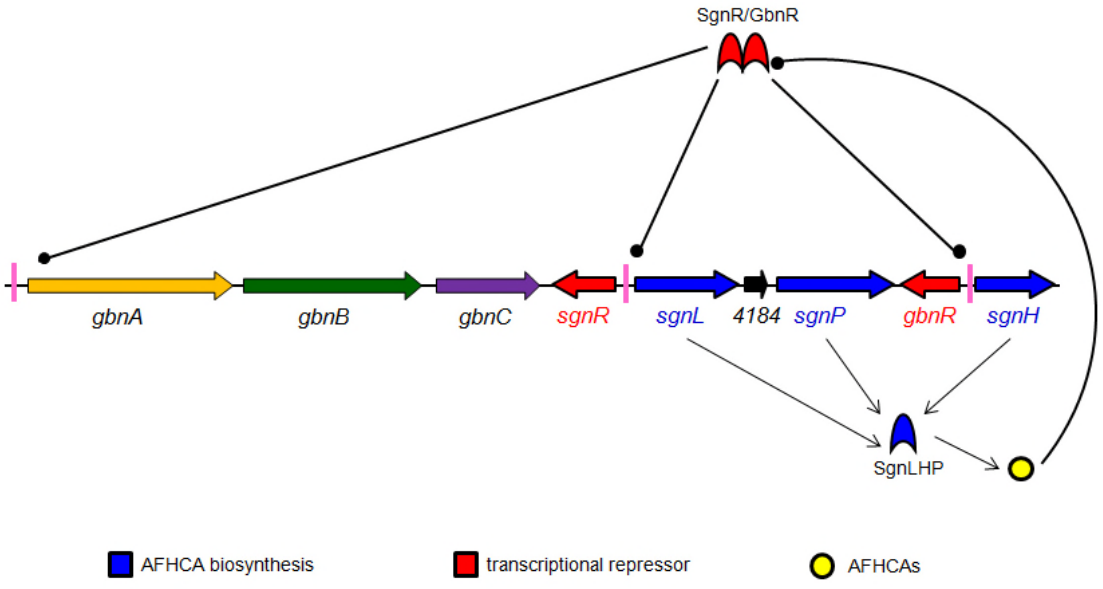


Figure 6.20 – Proposed signalling cascade involved in gaburedin biosynthesis. AREs are shown in pink. Inhibitory events are denoted by circular arrowheads

of *gbnR*, eventually the intracellular concentration of GbnR protein is sufficient to bind to the AREs in the *gbnA*, *sgnR-sgnL* and *gbnR-sgnH* promoters. This leads to *gbnABC* being silenced, even if there are high levels of AHFCAs present (as GbnR is proposed to be unresponsive to AHFCAs). At this time, even if there is a high intracellular concentration of SgnR (which may also silence *gbnABC*) is it possible that the intracellular concentration of AHFCAs is high enough to cause total dissociation of SgnR from its target ARE sequences.

In the *gbnR* deletion mutant, this final step (repression of *gbnABC* transcription) does not occur, therefore *gbnABC* continue to be transcribed, leading to constitutive – rather than transient – gaburedin biosynthesis (Figure 6.1). Because GbnR is not present, the only protein able to switch off gaburedin production in the *gbnR* mutant is SgnR, by binding to the ARE in the *gbnA* promoter. However, if the intracellular concentration of AHFCAs is high enough to cause complete dissociation of SgnR from the ARE sequences, then *gbnABC* transcription will continue.

However, when AHFCAs are added to the media (to between 100 μM and 400 μM concentration) it would appear that gaburedin production is decreased in the *gbnR* mutant. This implies that the extracellular AHFCA concentration plays a role in switching off gaburedin biosynthesis, as well as inducing gaburedin biosynthesis in the first place. If AHFCAs are produced in high enough quantities, they may induce expression of *sgnR* so much that the concentration of SgnR is great enough to then silence transcription of *gbnABC* again.

As SgnL is required for AHFCA biosynthesis, in the *sgnL-P* mutant no AHFCAs will be produced. Therefore, SgnR will not be dissociated from its ARE sequence in the

gbnA operater and gaburedin biosynthesis will not occur. However, upon addition of exogenous AHFCA to the growth media, SgnR will dissociate from the AREs, therefore *gbnABC* transcription (and subsequent gaburedin biosynthesis) will occur, as will overproduction of SgnR protein. Therefore, there will be a high concentration of intracellular SgnR protein available to bind to the ARE in the *gbnR* promoter. This may lead to repression of *gbnR* transcription, therefore there may be time lag in the increased intracellular concentration of GbnR protein. Therefore, addition of AHFCAs to the *sgnL-P* mutant can lead to gaburedin production (as shown in Figure 6.18). Eventually, a point is reached where the concentration of AHFCAs is high enough to cause so much de-repression of SgnR, that *gbnR* is transcribed (and enough GbnR protein is produced) to cause repression of transcription of *gbnABC*.

In the wild-type, it would appear that a much lower concentration of AHFCAs is required to repress gaburedin biosynthesis compared to the *sgnL-P* strain. This may be explained by the AHFCA concentration being high from the time of inoculation, therefore there could be early overexpression of *sgnR*, leading to overproduction of SgnR. Therefore, addition of AHFCAs to the wild-type could have abolished gaburedin biosynthesis, due to the high intracellular concentration of SgnR having silenced *gbnABC* transcription. This would be consistent with the hypothesis that *sgnL* mRNA cannot be translated until later in the developmental cycle (*sgnL* contains a TTA codon, as does *mmfL*). If SgnR is overproduced initially, this may have the effect of delaying onset of gaburedin biosynthesis in the wild type, therefore it would be interesting to repeat this feeding experiment to monitor production of gaburedins in wild-type upon addition of exogenous AHFCAs over a longer period than the three days reported here.

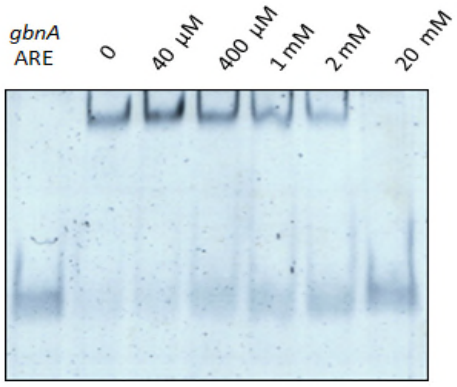


Figure 6.21 **–** Gel shift assay with SgnR protein (200 nM) and 28-bp sequence which includes the *gbnA* ARE sequence. Concentrations of AHFCA6 added to each sample shown above each lane. Similar experiments with the *sgnR-sgnL* and *gbnR-sgnH* promoters have also been performed (S. Zhou)

To summarise, without addition of AHFCAs to the strains, there is only transient AHFCA production in the cells (eventually the intracellular concentration of SgnR/GbnR is great enough compared to the intracellular concentration of AHFCAs to result in transcription of *sgnL* being totally repressed). The intracellular molar ratio of SgnR:GbnR:AHFCAs is vital for understanding this regulatory cascade. Transcription of *sgnL*, *sgnR*, *gbnR* and *gbnA* could be monitored by real-time PCR in the wild-type and mutant strains, allowing some insight into the effect of GbnR and SgnL on *sgnR* transcription. Indeed, time at which transcription of *gbnA*, *sgnL*, *sgnR* and *gbnR* occurs is distorted in the AHFCA feeding experiments presented here; therefore the results obtained cannot fully explain this complicated signalling cascade.

Moreover, the relative binding affinities of SgnR and GbnR for each of the AREs in the *gbnA*, *sgnR-sgnL* and *gbnR-sgnH* promoters are important to understand this regulatory mechanism, and work in our group is currently being undertaken to probe this (K. Styles). The relative binding affinities of SgnR and GbnR for each of the AREs could also be probed *in vitro* by EMSAs, and investigations are ongoing (S. Zhou, Prof. G. L. Challis group). Preliminary data has demonstrated the ability of AHFCA6 to dissociate SgnR from the *gbnA*, *sgnR-sgnL* and *gbnR-sgnH* promoters (Figure 6.21).

7. Summary and future work

7.1 Conclusions

7.1.1 Deletion of *S. venezuelae gbnR* leads to overproduction of novel natural products

As demonstrated in Chapter 2, deletion of the *arpA*-like repressor *gbnR* in *S. venezuelae* leads to overproduction of a series of novel natural products – the gaburedins. The approach used in the present study demonstrates the first genome-mining approach to discover a new series of metabolites in *Streptomyces* by this rational approach of de-repressing a cryptic biosynthetic pathway by identification and subsequent deletion of a transcriptional repressor gene.[152]

Deletion of *gbnR* leads to overexpression of *gbnABC*, therefore to gaburedin biosynthesis. Presumably, this is due to the mechanism by which *gbnABC* transcription is 'switched off' in *S. venezuelae* being disrupted by the removal of the *gbnR* gene. This is proposed to be analogous to the mechanism by which methylenomycin biosynthesis is controlled in *S. coelicolor*.[85]

7.1.2 Gaburedin biosynthesis proceeds via an intermediate derived from glutamic acid

Analysis of the metabolites produced by a *S. venezuelae gbnB*/*gbnR* deletion mutant demonstrated that gaburedins (which are overproduced in the *gbnR* deletion mutant) are not produced when *gbnB* is also deleted – therefore *gbnB* is essential for gaburedin biosynthesis. Bioinformatic analysis of the GbnA and GbnB proteins suggested that GbnA catalyses a decarboxylation, and that GbnB catalyses formation of an AMP-ester. This hypothesis was tested with a range of potential substrate analogues of the GbnA and GbnB enzymes, leading to the suggestion that GbnA is a glutamate

decarboxylase; this was inferred from the observation that four labels from [U-^{13}C]-L-glutamic acid are retained throughout the biosynthetic route to gaburedins. The substrate for GbnB is therefore proposed to be a derivative of GABA, although the *in vivo* experiments presented do not explicitly confirm this. The biosynthetic origin of the ureidyl carbon present in gaburedins also remains ambiguous.

A library of over 25 gaburedin analogues was produced by feeding experiments with a variety of nucleophiles (amino acids, amines and aminoalcohols). The observation such a variety of nucleophiles can be incorporated into gaburedins implies a conserved, reactive intermediate in gaburedin biosynthesis. It is proposed that this reactive intermediate is the carbamoyl ester **109** as proposed in the biosynthetic route proposed in Chapter 3.

7.1.3 GbnABC systems exist across a range of species

Genes encoding for *gbnABC* orthologues in fifteen other bacterial strains have been identified; these organisms belong to a diverse range of genera, and include Gram positive and Gram negative organisms. It would appear that GbnB and GbnC exist as a fusion protein in some of these bacteria. Of the strains identified which contain *gbnABC* orthologues, the regulatory cassette of *sgnR-sgnLPH-gbnR* is only conserved among some *Streptomyces*, suggesting that the mechanism by which gaburedin biosynthesis is regulated in *Streptomyces* is distinct to the regulatory mechanisms controlling expression of *gbnABC* systems in the other organisms.

7.1.4 Heterologous expression of *gbnB* in *E. coli* is not sufficient to produce gaburedins

Overexpression of GbnB in *E. coli* did not lead to production of any gaburedin-like molecules, even when fed with the putative precursor GABA. It was noted however

that overexpression of synthetic versions of *gbnB*, and the homologues *seag_B2133* and *smugs5_01685* from *Salmonella enterica* and *Streptococcus mutans* led to production of insoluble proteins in *E. coli*, which may account for their inactivity. Alternatively, this observation may also suggest that either GbnA or GbnC is required for gaburedin biosynthesis. In order to address this, a vector has been constructed in which synthetic *S. venezuelae gbnA* and *gbnB* should both be overexpressed in *E. coli*. A compatible vector carrying synthetic *gbnC* has also been constructed allowing for the whole GbnABC system to be reconstituted in *E. coli*, and for any metabolites produced to be compared. This will also help to deduce the role of GbnC in gaburedin biosynthesis and export.

7.1.5 Gaburedin biosynthesis is controlled by AHFCA signalling molecules

Comparative metabolic profiling of *S. venezuelae* strains revealed there to be nine signalling molecules overproduced by a *gbnR* deletion strain, compared with the wild-type and *sgnL-P* deletion strains. Five of these belong to the γ-butyrolactone family, and are likely to have the same stereochemistry as SCBs produced by *S. coelicolor* M1152. Of the four AHFCAs produced by *S. venezuelae gbnR* mutants, AHFCAs 6-8 represent new metabolites produced by a *Streptomyces* strain and are the first example of AHFCA signalling molecules to be produced by a strain other than *S. coelicolor* from which AHFCAs were first purified.

Herein, it has been shown that addition of AHFCAs to the *S. venezuelae sgnL-P* strain (unable to produce AHFCAs) can induce gaburedin biosynthesis in this strain. Therefore, AHFCA signalling in *Streptomyces* is not unique to the methylenomycin system. Addition of AHFCAs to the *S. venezuelae* wild-type strain did not have the

same effect of inducing gaburedin biosynthesis in this strain, suggesting a complex mechanism for AHFCA-mediated signalling in *S. venezuelae*.

7.2 Outlook

7.2.1 Deletion of *arpA*-like repressors is a valuable tool for natural product discovery

The present work demonstrates how identification of gene clusters encoding for signalling molecule biosynthesis and transcriptional repressors can be exploited for natural product discovery. As the signalling molecule/transcriptional repressor gene clusters are widely distributed across *Streptomyces* species, this offers a new approach to natural product discovery. This contrasts more traditional approaches, whereby the effect of deletion of transcriptional repressors has generally been used to overproduce already characterised natural products, rather than deletion of the repressor being used as a method of overproducing cryptic metabolites. This work sets a precedent for natural product discovery by this novel genome mining approach.

7.2.2 GbnB-type enzymes may have potential for producing drug-like molecules

The example presented in which 4-(3-cyclohexylureido)butanoic acid **103** was produced demonstrates the capability of GbnA and GbnB to introduce the ureidyl-GABA moiety into nucleophiles in order to create drug-like molecules *in vivo*. The structural similarity of gaburedins to other drug-like molecules such as the carboxypeptidase inhibitor **102** and the first-generation diabetes drug tolbutamide **105** demonstrate that gaburedins, or derivatives of gaburedins, could find use in a pharmaceutical setting. More work is needed to demonstrate the ability of GbnA/GbnB to introduce ureidyl-GABA moieties into other drug-like molecules *in vitro*.

7.2.3 Some signalling molecules produced by *S. venezuelae gbnR* mutants are also produced by other *Streptomyces*

The AHFCA signalling molecules discovered during this study demonstrate that *S. coelicolor* is not the only organism capable of biosynthesising AHFCAs, and this work provides a second example of natural product biosynthesis being induced by AHFCAs. One of the AHFCA molecules – AHFCA5 – is produced by both *S. coelicolor* and *S. venezuelae* and also has been shown to induce methylenomycin and gaburedin biosynthesis in both these organisms.[79,100] This could indicate a possible role for AHFCAs in interspecies signalling, and also indicates the possibility that AHFCAs could be used to induce natural product biosynthesis of cryptic metabolites in other species that contain *mmfR-mmfLHP-mmyR*-like gene clusters. Indeed, AHFCA5 and AHFCA7 (**201** and **203**) have been shown to induce methylenomycin production *in vivo* in *S. coelicolor* W81 and gaburedin production in *S. venezuelae sgnL-P:apra*.

Several of the SCB signalling molecules produced by *S. coelicolor* M1152 and *S. venezuelae gbnR* mutants have the same structures as SCBs which have previously been shown to interact with ScbR. For example, synthetic SCB6, SCB7 and SCB8 were used in the kanamycin resistance reporter assay to dissociate ScbR from its DNA target (though it must be noted that 480 times more SCB8, and 10 times more SCB6 or SCB7 were required to induce kanamycin resistance compared to SCB1).[132]

7.2.4 MmfR-MmfLHP-MmyR-like systems exist in a range of *Streptomyces*

To identify potential AHFCA-responsive systems in other species, a series of protein BLAST searches were performed, using both SgnR (SVEN_4182) and SgnR-SgnL (SVEN_4182-SVEN_4183) as the protein query. For the organisms containing SgnR protein analogues identified in Section 6.3, the genes surrounding the *sgnR-sgnL*

homologues were annotated. In addition to the *mmfR-mmfLHP-mmyR* regulatory cluster adjacent to the gaburedin clusters *S. globisporus* and *Streptomyces* sp. Amel 2Xe9 identified in Section 3.1, these results revealed the presence of eight new *mmfR-mmfLHP-mmyR* regulatory systems that had not been previously identified. These gene clusters from *S. exfoliatus* NRRL B-2924, *S. olindensis*, *Streptomyces* sp. WC-3641, *S. iakyrus*, *S. flavovariabilis*, *Streptomyces* sp. F-5135, *Streptomyces* sp. 351MFTsu5.1 and *S. roseoverticillatus* are shown in Figure 7.1. The other gene clusters shown in Figure 7.1 were identified by C. Corre, V. Poon and G. L. Challis during the present study.

For these systems, the autoregulatory response elements (AREs) were also identified using the MEME suite to first identify the consensus in the promoters of the *mmfL*, *mmfR* and *mmyR* homologues. This was followed by FIMO analysis of the nucleotide sequences that contain the *mmfR-mmfLHP-mmyR*-like gene clusters to reveal whether any other AREs located in the adjacent gene clusters. The results of these analyses are shown in Figure 7.1.

Interestingly, the *mmfR-mmfLHP-mmyR*-like gene clusters identified mostly appear adjacent to natural product biosynthetic gene clusters which probably encode for previously unknown natural products. AntiSMASH searches of these gene clusters revealed no analogous gene clusters. The exception is the *Streptomyces* S-31 gene cluster, which contains homologues of the entire methylenomycin gene cluster from the *S. coelicolor* SCP1 plasmid. Furthermore, the *S. iakyrus* and *S. flavovariabilis* gene clusters are highly homologous to one another (each protein sequence has a homologue in the other gene cluster with >80% identity at the amino acid level), as are the *S. olindensis* and *Streptomyces* sp. WC-3641 gene clusters identified.

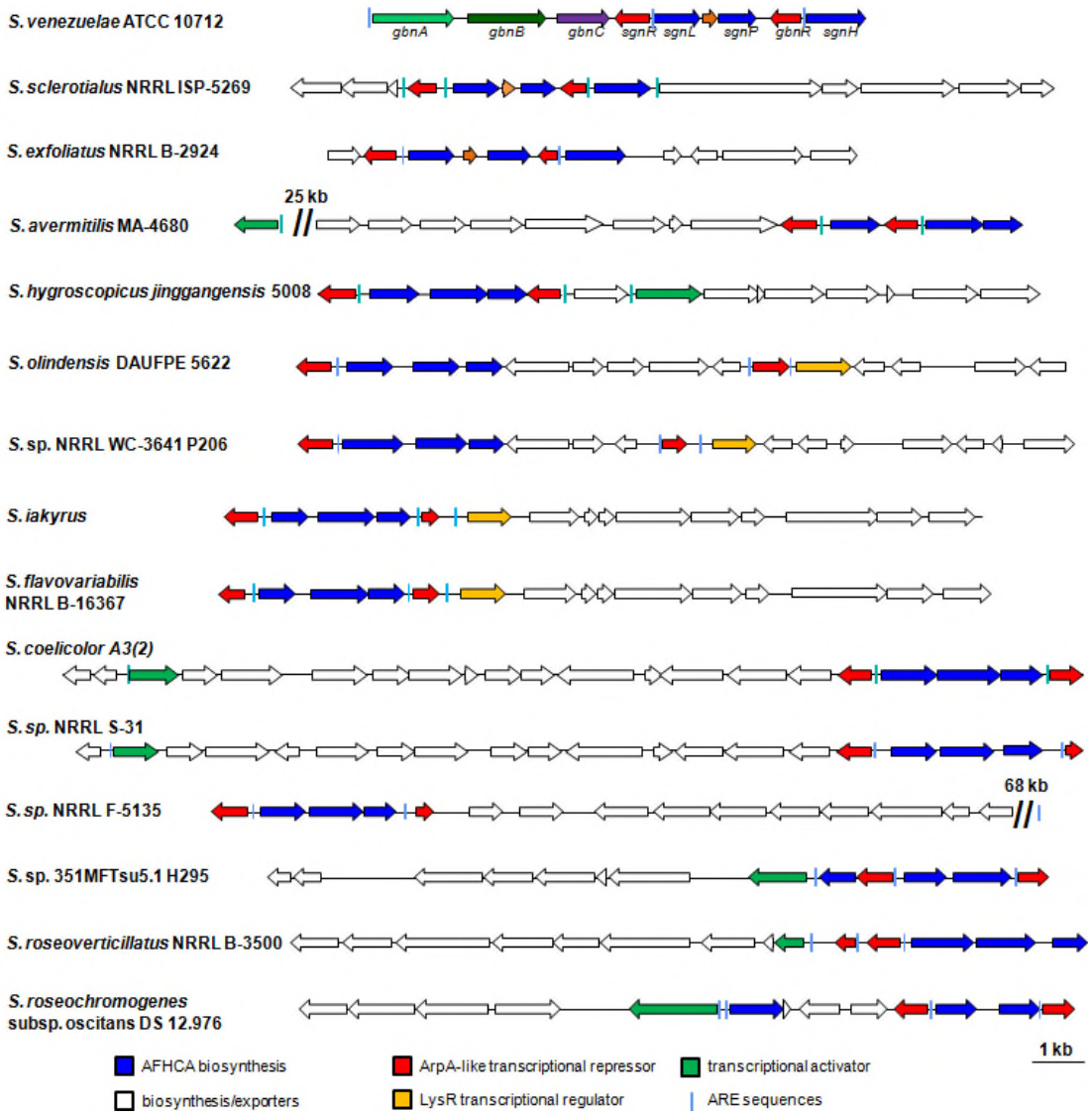


Figure 7.1 *– mmfR-mmfLHP-mmyR* gene clusters in other *Streptomyces* species shown with 10 kb upstream/downstream of the regulatory mini-cluster. AREs identified are shown, as are transcriptional activators identified in the adjacent clusters

7.3 Future work

7.3.1 Biological role of gaburedins

It will be important to establish the biological activity of gaburedins and to establish whether or not the *S. venezuelae gbnABC* system can be reconstituted in *E. coli* using the Duet vectors pJDS10 and pJDS11 constructed herein. If the co-expression of all three proteins can be optimised, the culture extracts can be screened for production of gaburedins or gaburedin-like natural products.

If gaburedins are produced by *E. coli* carrying pJDS10 and pJDS11, this could be followed by heterologous expression of the *gbnABC* genes from other bacteria in *E. coli*. Characterising the metabolites produced by *E. coli* carrying the *gbnA*, *gbnB* and *gbnC* genes of these other organisms will hopefully lead to discovery of further new natural products. This is interesting as there are currently very few examples of natural products produced by many of these other organisms in which *gbnABC* gene clusters have been identified.

7.3.2 Characterisation of SCB signalling molecules in *S. venezuelae*

Whilst the observation that the same SCBs can be produced by different species may suggest the potential for interspecies signalling,[95] further testing of the capability of the SCBs identified here to induce natural product biosynthesis in other strains *in vivo* will be required. Before this can be determined, however, it will be important to characterise the SCBs produced by *S. coelicolor* M1152 and *S. venezuelae gbnR* mutants, to confirm that their stereochemistry is actually (2*R*, 3*R*, 6R), as proposed.

7.3.3 AHFCA signalling in *S. venezuelae*

It would be of particular interest to perform further EMSAs with *S. venezuelae* SgnR and AHFCAs in order to determine whether or not addition of other AHFCAs would attenuate the ability of SgnR to bind to the AREs in the gaburedin cluster *in vitro*. Indeed, addition of other AHFCAs to the *S. venezuelae sgnL-P* mutant would also allow *in vivo* structure-activity relationships regarding ligands for SgnR to be established. These structure-activity relationships would complement the previous work already performed on the MmfR-AHFCA interactions *in vitro* and *in vivo* in order to determine whether the AFHCAs discovered in the extracts of the *gbnR* deletion

mutants are the AHFCAs which are most effective at disassembling the SgnR-ARE complex, as is the case for MmfR and the AHFCAs produced by *S. coelicolor*.[79,100]

This would contribute to understanding of AHFCA regulatory mechanisms and would provide further evidence as to whether or not AHFCA-mediated interspecies signalling could be possible in nature. Moreover, if the concentration of AHFCA required to induce gaburedin biosynthesis in the *S. venezuelae* wild-type strain can be determined, this will set a precedent for activation of cryptic gene clusters identified in other organisms containing *mmfR*-*mmfLHP*-*mmyR*-like clusters, without the need for genetic manipulation.

7.3.4 Deletion of *mmyR*-like repressors in other *Streptomyces* species

All of the gene clusters discussed in Section 7.2.4 contain homologues of both *mmfR* and *mmyR*. If the hypothesis of MmyR homologues being responsible for the switching off of biosynthetic pathways proves to be true – as suggested by this work with the gaburedin system – then deletion of the *mmyR*-like genes in these other *Streptomyces* species will offer a powerful method of natural product discovery by de-repression of other cryptic biosynthetic pathways.

8. Experimental

8.1 Analytical Chemistry

8.1.1 Mass spectrometry and NMR spectroscopy

Mass spectra were obtained using a Bruker Esquire 2000, Agilent 6130 Quadrupole mass spectrometer or Bruker High Capacity Trap (HCT) + ion trap mass spectrometer with an electrospray source and ^{1}H, ^{13}C DEPT, COSY, HSQC and HMBC NMR spectra were obtained using a Bruker Avance DPX-400, DRX-500, or AV-700 spectrometers. Samples prepared for NMR in d_6-DMSO were dissolved in 750 µL DMSO from glass ampoules and samples run in d_4-MeOD or $CDCl_3$ were dissolved in d_4-MeOD or $CDCl_3$ that had been stored under argon.

8.1.2 Extraction protocol

Unless otherwise stated, bacterial cultures were stored at 4 °C until the time of extraction, then ethyl acetate was added (in equal volume to the volume of culture media used) and acidified to pH 3 with 37% HCl (HCl was not added to extracts of *S. venezuelae gbnB::apra*, *gbnR::scar* and *S. coelicolor* M1152 strains used for comparative analysis of signalling molecules). The ethyl acetate layer was removed, evaporated under reduced pressure and the remaining residue was redissolved in 500 µL 50:50 HPLC grade methanol/water. For extracts from liquid cultures, the cells were pelleted (5000 rpm, 20 minutes) and the above extraction performed on the supernatant.

8.1.3 Liquid-chromatography-mass spectrometry (LC-MS)

20 µL of prepared extracts were injected through a reverse phase column (Zorbax C_{18}, size 46 x 150 mm, particle size 5 µm) connected to an Agilent 1100 HPLC. The

outflow was routed to a Bruker High Capacity Trap (HCT) + ion trap mass spectrometer with an electrospray source, operating in positive ion mode. A 5 min isocratic elution (95:5 solventA/solvent B) was followed by gradient elution to 0:100 solvent A/solvent B over 25 min. *m/z* = 50 to 1500 scan range was used.

The same method was used for samples analysed using the Bruker AmaZon mass spectrometer. 15 μL injections of microbial culture extracts diluted by a factor of five were used and the same liquid chromatography method with acetonitrile in place of methanol.

The high-resolution data were obtained by performing UPLC-MS through a reverse phase column (Zorbax Eclipse Plus C_{18}, size 2.1 x 100 mm, particle size 1.8 μm) connected to a Dionex 3000RS UHPLC coupled to Bruker Ultra High Resolution (UHR) Q-TOF MS MaXis mass spectrometer with an electrospray source. Sodium formate (10mM) was used for internal calibration and *m/z* = 50 to 1500 scan range was used. A 0.4 min isocratic elution (95:5 solvent A/solvent B) was followed by a gradient elution from 95:5 solvent A/solvent B to 0:100 solvent A/solvent B over 12.8 minutes. Solvents A and B were water (0.1 % HCOOH) and acetonitrile (0.1 % HCOOH). Some UHR-LC-MS analyses were performed with following method: a 5 min isocratic elution (95:5 solventA/solvent B) was followed by gradient elution to 0:100 solvent A/solvent B over 15.3 min. Solvents A and B were water (0.1 % HCOOH) and acetonitrile (0.1 % HCOOH), respectively.

Other high-resolution analyses were performed using the following gradient: a 5 min isocratic elution (95:5 solvent A/solvent B) was followed by a gradient elution from

95:5 solvent A/solvent B to 0:100 solvent A/solvent B over 25 minutes. Solvents A and B were water (0.1 % HCOOH) and methanol (0.1 % HCOOH).

8.1.4 Reverse phase HPLC

Reverse-phase HPLC was performed using a Zorbax XBD-C_{18} column (212 x 150 mm, particle size 5 μm) connected to an Agilent 1200 HPLC equipped with a binary pump and DAD detector. Gradient elution was used (solvent A: water with 0.1 % HCOOH, solvent B: methanol) with a flow rate of 20 mL min^{-1}. Fractions were collected by time or absorbance at 210 nm. Before each experiment was run, an isocratic equilibration using the initial solvent profile was performed for 15-20 min.

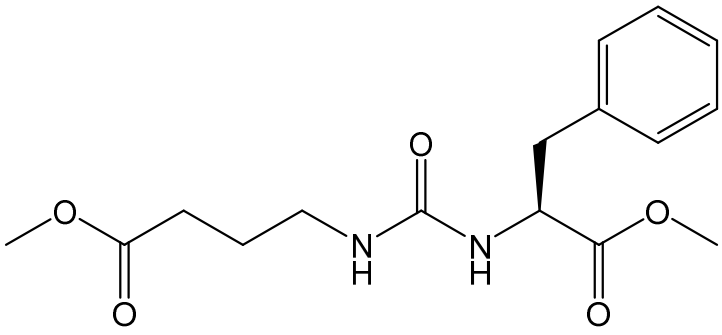
For purifying gaburedin A **80**, The following elution profile was used: 80:20 solvent A/solvent B for 5 min, then 20:80 solvent A/solvent B over 25 min. Solvent A was water + 0.1 % formic acid; solvent B was methanol. Fractions were collected (at the retention time of interest), pooled, methanol removed under vacuum and the metabolite of interest was re-extracted from the remaining water (2 x 50 mL ethyl acetate). The ethyl acetate was removed under reduced pressure and the sample re-dissolved in 50:50 methanol/water. This concentrated sample was re-run using the following gradient: 65:35 solvent A/solvent B for 5 min, then 35:65 solvent A/solvent B over 25 min. Fractions were collected every 0.2 minutes and samples corresponding to the centre of peak at t = 7.5 min which had the desired *m/z* = 295 were then combined. This sample was reconstituted by removal of the organic solvent, re-extracted from the remaining water and the remaining residue was dried and redissolved in d_6-DMSO.

8.1.5 Chiral HPLC analysis

Samples dissolved in 50:50 methanol/water were injected through an Agilent 1100 equipped with a quaternary pump through a Chiral-PAK IA column coupled to a Bruker High Capacity Trap (HCT) + ion trap mass spectrometer with an electrospray source, operating in positive ion mode. A 5 min isocratic elution (75:25 solvent A/solvent B) was followed by gradient elution to 0:100 solvent A/solvent B over 25 min. Solvents A and B were water (0.1 % TFA) and methanol (0.1 % TFA), respectively.

8.2 Synthetic Chemistry

Unless stated otherwise, all reactions were performed in oven-dried glassware under an atmosphere of argon using anhydrous solvents. Pyridine was distilled over CaH_2. All reagents that were commercially available were obtained from Sigma-Aldrich or Fischer Scientific, with the exceptions of some of the ingredients for the growth media, and were used as provided without further purification unless otherwise stated. Reaction temperatures were controlled using an IKA RCT basic hot plate stirrer with an IKA ETS-D5 temperature controller.



8.2.1 Preparation of methyl (*S*)-4-(3-(1-methoxy-1-oxo-3-phenylpropan-2-yl)ureido)butanoate (**90**)

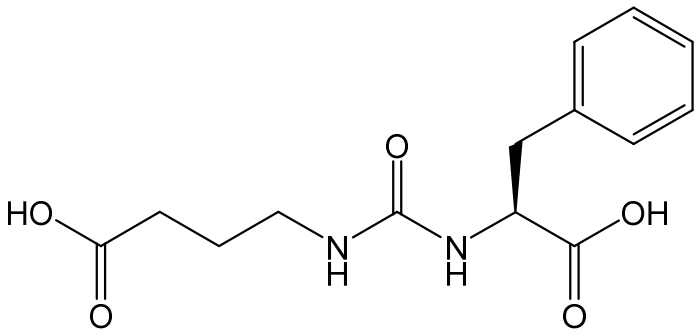
To a stirred solution of mono-methyl glutarate (1.2 mL, 9.6 mmol, 1.4 eq) in dry pyridine (6 mL), diphenylphosphorylazide (1.5 mL, 7 mmol, 1 eq) was added and stirred at 90 °C for 1 h under argon. After 1 h, L-phenylalanine *O*-methyl ester hydrochloride (0.736 g, 3.41 mmol, 0.49 eq) was added and the reaction mixture stirred at 90 °C for a further 4 h. After being allowed to cool, 30 mL ethyl acetate was added to the reaction mixture, the mixture washed with 2M HCl (5 x 30 mL), and the ethyl acetate layer dried ($MgSO_4$) and removed under reduced pressure. The product was then purified by flash column chromatography (4:1 ethyl acetate: hexane eluent), yielding 331 mg (1.03 mmol, 30% yield) of protected L-gaburedin-Phe, which was a white solid. This procedure is modified from the work of Takaya, *et al*.[160]

The yield reported here is the overall yield from the two-step, one-pot reaction to yield the dimethyl ester from mono-methyl glutarate, treating L-phenylalanine *O*-methyl ester as the limiting reagent.

δ_H (400 MHz, $CDCl_3$); 1.77 (2H, s, $CH_2CH_2CH_2$), 2.33 (2H, t, CH_2CH_2CO, J = 7.3 Hz), 3.08 (2H, m, CH_2Ph, J = 6.1 Hz), 3.14 (2H, t, CH_2CH_2NH, J = 7.1 Hz), 3.66 (3H, s, CH_2COOCH_3), 3.71 (3H, s, CH_3COCH), 4.75 (1H, t, CH, J = 6.2 Hz), 7.12-7.27 (5H, multiplet, ArH)

δ_C (100 MHz, $CDCl_3$); 25.1 ($CH_2CH_2CH_2$), 31.0 (CH_2CH_2CO), 38.4 (CH_2ArH), 39.8 (CH_2CH_2NH), 51.8 (CH_2COOCH_3), 52.3 (CH_3COCH), 54.2 (CH), 127.0 (*para* ArH), 128.5 (*meta* ArH), 129.3 (*ortho* ArH), 136.2 (ipso Ar), 157.7 (NHCONH), 172 (CH_3OCOCH), 174.0 (CH_2COOCH_3)

HRMS (ESI) calc'd for $C_{16}H_{22}N_2O_5Na$ $[M+Na]^+$ 345.1418, found 345.1421



8.2.2 Preparation of 4-[(1-benzyl-2-hydroxy-2-oxo-ethyl)carbamoylamino]butanoic acid (**91**)

The dimethyl ester protected L-gaburedin-Phe **90** (28.8 mg, 8.94 x 10^{-5} mol) was redissolved in 18.5 mL of a stirred 1:1 mixture of methanol/aqueous NaOH to a final concentration of 2M NaOH and heated to 50 °C. After 24 h, the methanol was removed and the remaining aqueous solution washed with ethyl acetate (3 x 20 mL), the pH of the mixture reduced to 1 using 2M HCl, and ethyl acetate extractions (3 x 20 mL) performed. The ethyl acetate extracts were combined, dried ($MgSO_4$) and the ethyl acetate removed under reduced pressure. The remaining residue was redissolved in 3 mL of a 1:1 mixture of methanol/water and purified by reverse-phase HPLC using the same conditions as detailed below, with the gradient as follows: 65:35 solvent A/solvent B for 5 min, to 35:65 solvent A/solvent B over 25 min. Methanol was removed under reduced pressure and the water removed by lyophilization using a VirTis benchtop K freeze-dryer, furnishing 12.6 mg (4.27 x 10^{-5} mol, 48% yield) of 4-[(1-benzyl-2-hydroxy-2-oxo-ethyl)carbamoylamino]butanoic acid (gaburedin **91**), which was a white solid. The product was dissolved in d_6-DMSO for NMR analysis.

δ_H (500 MHz, d_6-DMSO); 1.54 (2H, m, $CH_2CH_2CH_2$, J = 8.2 Hz), 2.17 (2H, t, CH_2CO_2H, J = 8.5 Hz), 2.84-3.01 (4H, m, CH_2Ph and $NHCH_2CH_2$, J = 8.3 Hz, 8.2 Hz, 8.3 Hz, 13.6 Hz), 4.29 (1H, m, CH, J = 8.3 Hz, 8.3 Hz), 5.96 (1H, d, CHNHCO, J = 8.8 Hz), 6.17 (1H, t, CH_2NHCO, J = 8.3 Hz), 8.16-8.25 (5H, m, ArH)

δ_C (125 MHz, d_6-DMSO); 28.5 ($CH_2CH_2CH_2$), 31.1 (CH_2CO_2H), 38.7 (CH_2Ph), 38.5 ($NHCH_2CH_2$), 54.1 (*C*H), 126.2 (*Ar*H, *para*), 128.0 (*Ar*H, *meta*), 129.3 (*Ar*H, *ortho*), 138.8 (*Ar*, *ipso*), 158.5 (NH*C*ONH), 174.0, (CO_2H), 174.3 (CO_2H)

HRMS (ESI) calc'd for $C_{14}H_{19}N_2O_5$ $[M+H]^+$ 295.1288, found 295.1283

8.3 Microbiological methods

8.3.1 Bacterial strains

Strains of *S. venezuelae* wild type and *sgnL-P::apra* mutant strains and *S. coelicolor* M1152 were obtained from the John Innes centre, Norwich, UK (M. Al-Bassam and M. J. Buttner). The *gbnR::apra* mutant was constructed in our group, University of Warwick (C. Corre). The *gbnR::scar* and *gbnB::apra, gbnR::scar* mutant strains were constructed in our group in collaboration with the G. L. Challis group (O. Lazos and A.-C.Olivan).

8.3.2 Culture media

Mannitol Soya Flour Medium Solid (SFM) was prepared by dissolving Bacto agar (20 g), mannitol (20 g) and soya flour (20 g) in tap water (1 L) and autoclaving before use.

LB medium was prepared using LB Broth, Miller (Fisher) to a final concentration of 25 gL^{-1}. For solid LB, Bacto agar was added (15 gL^{-1}).

Mannitol yeast medium (MYM) was made in batches with the following ingredients added to the concentrations as follows: maltose (4 gL^{-1}), yeast extract (4 gL^{-1}), malt extract (10 gL^{-1}) and Bacto agar (18 gL^{-1}).

Unless otherwise stated, Modified Supplemented Minimal Medium Solid (SMMS) was made up in batches with the following ingredients added to the concentrations as follows: Difco casaminoacids (2 gL^{-1}), TES buffer (8.68 gL^{-1}) and Bacto agar (15 gL^{-1}) were dissolved in distilled water, the pH adjusted to 7.2 using 10 N NaOH and autoclaved. At the time of use, the media was re-melted and the following ingredients were added to the stated concentrations: NaH_2PO_4 + K_2HPO_4 (50 mM each, 10 mL per litre of culture), $MgSO_4.7H_2O$ (1 M, 5 mL per litre of culture), glucose (50% w/v, 18 mL per litre of culture), trace elements solution (2 mL per litre of culture). Trace elements solution contained 0.1 gL^{-1} each of $ZnSO_4.7H_2O$, $FeSO_4.7H_2O$, $MnCl_2.4H_2O$, $CaCl_2.6H_2O$ and NaCl. The solution was stored at 4°C in a refrigerator.

8.3.3 Modified culture media

Unless otherwise stated, modified Supplemented Minimal Medium Solid (SMMS) was prepared as above, with the addition of each amino acid to final concentration of 20 mM before adjusting the pH to 7.2. In the cases of the amino acid analogues *n*-butylamine, isobutylamine, isopentylamine, cyclohexylamine, L-ala-ala and L-valinol, a separate stock of the amine was dissolved in sterile water and added to the culture medium after re-melting, and the amount of water used to dilute the initial ingredients was altered so that the final concentrations of each ingredient was consistent with those of the other experiments. The prepared media was poured (15 mL per plate).

8.3.4 Culture media enriched with stable isotope precursors

For the feeding experiments with labelled precursors (L-glutamate-1-^{13}C and L-glutamate-$^{13}C_5$), 5 mL of media was made to a final concentration of 10 mM, keeping the final concentration and pH of all other ingredients the same as above. 3 µL bacterial

spores were used to inoculate each 5 mL plate. Metabolites were extracted at pH 3 and the culture extracts were redissolved in 200 µL 50:50 HPLC grade methanol/water.

For experiments with deuterium labelled precursors (d_{10}-L-leucine, d_{10}-L-isoleucine, d_8-DL-valine), 15 mL SMMS was made to a final concentration of 10 mM precursor molecule, keeping the final concentration and pH of all other ingredients the same as above. d_5-propionic acid and d_7-butyric acids were added after autoclaving to a final concentration of 1 mM or 10 mM. Metabolites were extracted at pH 6 and the culture extracts were redissolved in 500 µL 50:50 HPLC grade methanol/water.

8.3.5 Feeding *Streptomyces coelicolor* M1152 with stable isotope precursors

AlaMM liquid medium (50 mL) was inoculated with 20 µL spores of *Streptomyces coelicolor* M1152 and incubated (30 °C, 180 rpm). After 24 h, a solution of sterile filtered precursor (d_5-propionic acid, d_7-butyric acid, d_8-DL-valine, d_{10}-L-leucine or d_{10}-L-isoleucine) was added to give a final concentration of 1 mM. This feeding procedure was repeated at 12 h intervals until a total incubation time of 5 days, after which the metabolites were extracted from these cultures and analysed as described in Sections 8.1.2 and 8.1.3.

8.3.6 Preparation of bacterial spores

Mannitol SFM was used for the production of bacterial spores. Media was remelted and apramycin added to a final concentration of 50 µg mL^{-1} for growth of the *sgnL-P::apra* and *gbnR::apra* and *gbnB::apra, gbnR::scar* mutant strains. No apramycin was added to the SFM used for growth of the wild type *S. venezuelae*. Each plate of SFM was inoculated with 20 µL of bacterial spore stocks. After incubating at 30 °C for 5 days, sterile water was added to each plate, the spores resuspended, and filtered. The

suspensions were centrifuged (3000 rpm, 5 minutes) and the supernatants removed, leaving approximately 1 mL of suspension. To each suspension, approximately 1 mL of glycerol (50 %) was added and the spore stocks stored at -20 °C. This a variation of the method detailed by Hopwood, *et al*.[155]

8.3.7 Preparation of *S. venezuelae* Genomic DNA

Genomic DNA was prepared using the MP Bio FastDNA Spin Kit for Soil extraction kit and enclosed protocol.

8.3.8 Preparation of SV4 F01 Cosmid DNA

Cosmid DNA was initially prepared using the Fermentas GeneJET Plasmid Miniprep kit, heating the elution buffer to 70 °C and incubating the samples for 5 min before the final elution step. For further preparations, the following procedure was performed: 60 µL *E. coli* MC1061 stock was transformed using 2 µL SV4 F01 cosmid DNA by electroporation at 1700 V, 1 mL ice cold liquid LB immediately added and these cultures incubated at 37 °C for 16 h in liquid LB medium with kanamycin added to a final concentration of 50 µg mL^{-1}. The cells were centrifuged (5 min at 2000 rpm) and the pellet resuspended in 50 mM Tris/HCl, pH 8; 10 mM EDTA, (200 µL) and 200 mM NaOH, 1% SDS (400 µL).

The mixture was immediately inverted 10 times and 3 M KOAc, pH 8.5 (300 µL) added and the mixture inverted another 5 times. The mixture was centrifuged (5 min at 10,000 rpm) and the supernatant extracted with phenol/chloroform (400 µL), vortexed for 2 min and centrifuged (5 min at 10,000 rpm). The phenol/chloroform layer was removed, isopropanol (600 µL) added and centrifuged (20 min at 10,000 rpm, 4 °C). The isopropanol was removed and the pellet washed with 70% ethanol (200 µL),

centrifuged (5 min at 10,000 rpm) and left open overnight for the ethanol to evaporate. The following day, the prepared cosmid DNA was resuspended in 35 μL sterile water. This is a variation on the method presented by Gust, *et al*.[82]

8.3.9 Chemical transformations

For subsequent transformations, 50 μL chemically competent *E. coli* DH5α (New England Biolabs), *E. coli* TOP10 (Invitrogen) or *E. coli* BL21star(DE3) (Invitrogen) were incubated on ice with 1 μL cosmid for 30 minutes before heat-shock in a 42 ˚C water bath (30 s). 250 μL LB (or SOC) media was added and cells incubated (37 ˚C, 180 rpm, 1 h) after which the cells were added to 25 mL LB plates inoculated with appropriate resistance marker (50 μg mL^{-1} kanamycin, 100 μg mL^{-1} ampicillin, 50 μg mL^{-1} apramycin or 34 μg mL^{-1} chloramphenicol) and incubated (37 ˚C, 16 h). Single colonies were picked and incubated further in 10 mL LB containing appropriate resistance marker(s) (37 ˚C, 180 rpm, 16 h). 0.7 mL of each culture was taken and added to 0.7 mL 50 % glycerol for storage at -80 °C.

For transformations using the *E. coli* iba TOP10 StarGate cells, the same procedure as above was used with the following modifications: cells were heat-shocked at 37 ˚C for 5 minutes, and 900 μL LB was used for the 1 h pre-incubation before spreading the cells onto selection plates.

8.5.10 Restriction digests

Unless manufacturer's guidelines recommended otherwise, 0.5 μL of restriction endonuclease enzyme was incubated with 2 μL of the recommended buffer and 15 μL of the prepared plasmid DNA for 2 h, after which 7 μL DNA loading dye was added to

the reaction mixture and a 1 % agarose gel run of the resulting fragments. Reactions using FastDigest enzymes were incubated for 30 min.

8.3.11 Polymerase Chain Reaction (PCR) conditions

An Eppendorf Mastercycler gradient and an Eppendorf Mastercycler personal PCR machine were used for variable annealing temperature and constant annealing temperature PCR reactions, respectively. Unless otherwise stated, the following quantities of each reagent were used per reaction: sterile water (36 μL), DMSO (2.5 μL), dNTPs, 10 mM each (2 μL), Fermentas Taq DNA polymerase, 5 $U\mu L^{-1}$ (1 μL), $(NH_4)_2SO_4$ buffer (5 μL), $MgCl_2$ buffer (3 μL) DNA template (0.7 μL). 1 μL of each forward and reverse primers were used for each reaction.

Expand High Fidelity polymerase (Roche) was used in PCR reactions with the following quantities of each reagent per reaction: sterile water (37 μL), DMSO (2.5 μL), dNTPs, 10 mM each (2 μL), Expand High Fidelity polymerase (Roche), 5 $U\mu L^{-1}$ (1 μL), $MgCl_2$ buffer (5 μL) and DNA template (0.5-0.7 μL). 1 μL of each forward and reverse primers were used for each PCR reaction.

Primers were obtained from Sigma-Genosys and diluted to 100 μM as per the manufacturer's guidelines before use. The typical program below was used for the single annealing temperature PCR reactions: Reaction mixture heated to 95 °C (5 min), 30 cycles consisting of denaturation at 95 °C (45 s), annealing at 55 °C (45 s) and elongation at 72 °C (90 s). This was followed by a further elongation step at 72 °C for 15 min. The reaction(s) were then held at 4 °C until PCR products were used for gel electrophoresis.

8.3.12 Agarose Gel Electrophoresis

To 5 μL PCR product, 2 μL loading dye was added, and the samples loaded onto a 1% agarose gel, prepared by addition of agarose (1 g) to 100 mL 1x TBE buffer and Gel Red (5 μL) after allowing the gel mixture time to cool. TBE buffer contained Tris base (10.8 gL^{-1}), boric acid (5.5 gL^{-1}), and EDTA (1.17 gL^{-1}).

8.3.13 Protein production in *E. coli* BL21star

10 μL *E. coli* BL21star stock carrying the plasmid of interest was in 10 mL LB containing appropriate resistance marker (concentrations as in Section 8.3.9) and incubated at 37 °C for 16 h. This culture was used to prepare a 1% v/v culture of LB and grown until cell density had reached OD_{600nm} = 0.6, when isopropyl β-D-1-thiogalactopyranoside (ITPG) was added to a final concentration of 0.5 mM. After addition of IPTG, the cultures were grown for a further 16 h at 15 °C. For initial pilot expressions, control samples were also prepared, to which no IPTG was added.

8.3.14 Preparation of cell lysates

Cells were pelleted by centrifugation at 5000 rpm for 10 minutes (or 3000 rpm, 30 minutes for smaller volumes), and resuspended in wash buffer containing 100 mM NaCl, 20 mM Tris-HCl (pH 8.8), and 10% v/v glycerol (6 mL buffer used per 50 mL, 15 mL buffer for 500 mL original cell culture). Phenylmethanesuphonylfluoride (PMSF) was also added (100 μL per 10 mL resuspended cells). Resuspended cell pellets stored at -20 °C until time of use. The resuspended cells were lysed using a Constant Systems E1061 cell disrupter, and the cell debris pelleted by centrifugation at 18,000 rpm for 30 minutes. The supernatant was removed and stored at 4 °C for later use as a cell-free extract containing soluble proteins. A pipette tip was used to transfer a

small quantity of the cell debris to 1 ml of the above wash buffer, and the resuspended cell debris was stored at 4 °C for use later as a sample containing the insoluble proteins.

8.3.15 Sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE)

Samples were prepared by addition of 3 μL of SDS-dye (Thermo) to 20 μL sample and incubated at room temperature for at least 1 h before loading onto the prepared gel. 10% SDS-PAGE gels were prepared as follows – for 5 mL resolving gel: distilled water (2.025 mL), 30% acrylamide solution (1.67 mL), 1.5 M Tris-HCl, pH 8.8 (1.25 mL), 20% w/v sodium dodecyl sulphate (25 μL), 1% w/v ammonium persulphate (100 μL) and *N*,*N*,*N*,*N*-tetramethylethylenediamine (TEMED) (5 μL). Stacking gel was prepared as follows: distilled water (705 μL), 30% acrylamide solution (167 μL), 1.5 M Tris-HCl, pH 8.8 (127 μL), 20% w/v sodium dodecyl sulphate (5 μL), 1% w/v ammonium persulphate (100 μL) and (TEMED) (5 μL).

After allowing the gel to set for 30 min, gels were pre-run at 200 V for 1 h, and then run at 120V for 10 min after the samples were loaded, which was increased to 200 V until the second dye front had reached the base of gel. All samples were run concurrently with PageRuler Plus prestained protein dye (Thermo).

8.4 Biological assays

8.4.1 Induction assay to screen for gaburedin production by non-producing strain

S. venezuelae gbnB/R was grown on 5 x 25 mL plates of SMMS for 3 days, after which metabolites were extracted at pH 6 as detailed in Section 8.1.2, and the culture extract redissovled in 2.4 mL MeOH/H_2O, sterile filtered and stored at 4 °C until the time of use. Culture extract (10 μL) was loaded onto sterile discs and the discs placed in the

centre of a 25 mL SMMS plate inoculated with 10 µL *S. venezuelae sgnL-P::apra* spores after 0, 1 or 2 days of growth.

1 cm x 1 cm agar plugs cut out of a 3-day old plate of *S. venezuelae gbnB/R* culture were also added to the centre of a 25 mL SMMS plate inoculated with 10 µL *S. venezuelae sgnL-P::apra* spores after 0, 1 or 2 days of growth. Control plates were also prepared in which MeOH/H_2O was loaded onto a sterile disc or sterile agar plug was added to plates inoculated with *sgnL-P* spores. Plates inoculated with *sgnL-P* were all extracted at pH 3 and screened for gaburedin production by LC-MS as detailed in Section 8.1.3.

For future experiments, metabolites were extracted from 16 x 3 day old 50 mL SMMS plates inoculated with *S. venezuelae gbnB/R* and redissovled in 8 mL MeOH/H_2O. 1 mL of this extract was diluted by a factor of 10 in MeOH/H_2O. Both culture extracts were sterile filtered and 1 mL added directly to the *sgnL-P* strain after 0, 1 or 2 days of growth. Control plates were also prepared to which MeOH/H_2O was added. Plates inoculated with *sgnL-P* were all extracted and screened for gaburedin production by LC-MS as in Section 8.1.3.

8.4.2 Induction assay with purified AHFCA compounds

AHFCAs 6 and 7 were dissolved in DMSO to 10 mM concentrations. These solutions were then diluted up to 500 µL with sterile dH_2O and added to re-melted SMMS (5 mL) to give final concentrations of 100 µM, 10 µM and 1 µM AHFCA. 3 µL *S. venezuelae sgnL-P::apra* spores were inoculated into each 5.5 mL plate. These cultures were incubated at 30 °C for 3 days, the metabolites extracted at pH 3 and analysed by LC-MS.

For future experiments using purified AHFCA compounds, stock solutions of AHFCAs 5, 6 and 7 (1 M, 0.25 M and 0.025 M) dissolved in DMSO were prepared. 2 µL of each of these solutions were added to re-melted SMMS to give final AHFCA concentrations of 400 µM, 100 µM and 10 µM and each 5 mL plate was inoculated with 3 µL of *S. venezuelae sgnL-P::apra* spores. These cultures were incubated at 30 °C for 3 days, the metabolites extracted at pH 3 and analysed by LC-MS.

8.4.3 Antibiotic assay

Cultures of the target strains of *E. coli* rosetta *kanR*, *S. coelicolor* M1152 and *Bacillus subtilis* were grown in LB to a final OD_{600} of 0.4. 300 µL of these cultures were each added to 3 mL warmed 'soft' LB (0.7% agar) and this mixture added to 25 mL of pre-prepared LB. *S. venezuelae gbnR* and *gbnB/R* was grown on 5 x 50 mL plates of SMMS for 3 days, after which neutral extractions were performed, and the culture extracts redissovled in 2.4 mL MeOH, sterile filtered and stored at 4 °C.

7 µL of prepared extract was added to a sterile disc and placed on top of the soft agar inoculated with the target strains. A control disc was also prepared on each plate, inoculated with 7 µL ampicillin (5 mg mL^{-1} final concentration). The plates were incubated at 30 °C (*S. coelicolor*) or 37 °C (*E. coli* and *B. subtilis*) for 24 hr and inspected for zones of inhibition.

8.4.4 *In vivo* assay for gaburedin or gaburedin analogue production by *E. coli* BL21star/pJDS02, *E. coli* BL21star/pJDS04 and *E. coli* BL21star/pJDS06

As in Sections 8.3.13 and 8.3.14, *E. coli* BL21star, *E. coli* BL21star/pJDS02, *E. coli* BL21star/pJDS04 and *E. coli* BL21star/pJDS06 were grown and induced with IPTG, and GABA was added to a final concentration of 10 mM. The cells were grown and

pelleted (4000 rpm, 40 min), after which metabolites were extracted from the supernant (50 mL ethyl acetate, pH 2). The cells were lysed (5 mL buffer) and metabolites extracted from the cell lysates (10 mL ethyl acetate, pH 2).

8.4.5 *In vivo* assay for gaburedin or gaburedin analogue production by *E. coli* BL21star/pJDS10

As in Section 8.4.4, except cultures were inoculated with 1 mM L-glutamate final concentration at the same time as addition of IPTG.

8.5 Molecular biology

8.5.1 Construction of plasmids pJDS02, pJDS04, pJDS05, pJDS06 and pJDS08

Each gene of interest was amplified by PCR with High Fidelity Polymerase (Roche) as described in Section 8.3.11 and appropriately designed primer pairs (Table 8.1). PCR products were run on a 1 % agarose gel, the bands of interest excised, and purified using the GeneJet Gel Extraction Kit (Fermentas).

Concentrations of each pure PCR product were determined by use of a Thermo NanodropLite, and diluted accordingly to fall within the desired range of 1-5 ng μL^{-1} for cloning into pET151 as directed by the Invitrogen TOPO cloning manual. 1 µL of each of the diluted PCR products were mixed with sterile water and the other components of the kit and incubated for 5 min as directed in the TOPO cloning manual. *E. coli* TOP10 cells were transformed (as in Section 8.3.9) using 3 µL of each TOPO cloning reaction mixture and poured onto selection plates with 100 µg mL^{-1} final concentration of ampicillin.

Table 8.1– Primers used for cloning *gbnA* and *gbnB* from *Streptomyces venezuelae* into pET151 for protein overproduction. Restriction sites are highlighted in italics, and 5'-CACC overhangs are underlined

TOPO cloning primers for cloning native *S. venezuelae* genes:

Primer	*Gene to amplify*	*Sequence*
A	*gbnA* (*sven_4179*)	5′-CACCATGACCGCATCCACCAGC-3′
A'	*gbnA* (*sven_4179*)	5′-GGGAGGTGAGGTGTGGTC-3′
B	*gbnA* (*sven_4179*)	5′-CACCATGACCGCATCCACCAGCTCCAC-3′
B'	*gbnA* (*sven_4179*)	5′-GAGCCGGGAGGTGAGGTGTGGTC-3′
C	*gbnA* (*sven_4179*)	5′-CACCATGACCGCATCCACCA-3′
C'	*gbnA* (*sven_4179*)	5′-AGCCGGGAGGTGAG-3′
D	*gbnB* (*sven_4180*)	5′-CACCATGATCAAGATCGCGG-3′
D'	*gbnB* (*sven_4180*)	5′-GGCGGGGGCGGCCGGGTC-3′
E	*gbnB* (*sven_4180*)	5′-CACCATGATCAAGATCGCGGACATCCG-3′
E'	*gbnB* (*sven_4180*)	5′-GGGTCGGCGGGGGCGGCCGGGTC-3′
S	*sgnL* (*sven_4183*)	5′-CACCTCCATCGGTGTGCTGCAG-3′
S'	*sgnL* (*sven_4183*)	5′-CCGGCCCTCACGCG-3′
T	*sgnR* (*sven_4182*)	5′-CACCGCGACACCCCGAAGCCAA-3′
T'	*sgnR* (*sven_4182*)	5′-TCAGCGGCGCAGGCC-3′

Screening primers:

Primer	*Gene to amplify*	*Sequence*
F	*gbnR* (*sven_4187*)	5′-CCTTGAGTGGTGGTACTG-3′
F'	*gbnR* (*sven_4187*)	5′-GACGTCACACTCGTCATC-3′
G	*gbnA* (*sven_4179*)	5′-CTTGCGGCCTTGCGGCCTTG-3′
G'	*gbnA* (*sven_4179*)	5′-TCGGGATCGGAGGCGGAGTG-3′
H	*gbnB* (*sven_4180*)	5′-GCCACCAGCTGACCACAC-3′
H'	*gbnB* (*sven_4180*)	5′-TCCGGCTGGGAGTTGAGG-3′
I	*gbnA* (*sven_4179*)	5′-CTTGCGGCCTTGCGG-3′
I'	*gbnA* (*sven_4179*)	5′-TCGGGATCGGAGGCG-3′

TOPO cloning of synthetic gbnB from *sven*, *seag* and *smug*:

Primer	*Gene to amplify*	*Sequence*
J	synthetic *sven_gbnB*	5′-CACCATCAAAATTGCCGATA-3′
J'	synthetic *sven_gbnB*	5′-AAGGCCCATGAGGC-3′
K	synthetic *seag_gbnB*	5′-CACCATGATTACCCTGCGTCG-3′
K'	synthetic *seag_gbnB*	5′-CCCATGAGGCCCAG-3′
L	synthetic *smugs5_gbnB*	5′-CACCATGCTGGCACTGGAAAATC-3′
L'	synthetic *smugs5_gbnB*	5′-GCCCATGAGGCCCAG-3′

Subcloning synthetic *gbnB* into pJDS05 and synthetic *gbnA* into pJDS5:

Primer	*Gene to amplify*	*Sequence*
M	synthetic *sven_gbnA*	5′-CACCACCGCAAGCACCAG-3′
M'	synthetic *sven_gbnA*	5′-CCCATGAGGCCCAG-3′
N	synthetic *sven_gbnA*	5′-*GAGCTC*ATGACCGCAAGCACCA-3′
N'	synthetic *sven_gbnA*	5′-*GAGCTC*TTAGCTGGTCGCAACTG-3′
O	synthetic *sven_gbnB*	5′-*GAGCTC*ATGATCAAAATTGCCGATATC-3′
O'	synthetic *sven_gbnB*	5′-*GAGCTC*TTAAACACCAACGGTTTCTG-3′

Sequencing primers:

Primer	*Gene to amplify*	*Sequence*
T7_for	T7 promoter forwards	5′-TAATACGACTCACTATAGGG-3′
T7_rev	T7 terminator reverse	5′-CTAGTTATTGCTCAGCGG-3′

Colonies (typically 5 to 10) were picked from these plates, added to 10 mL LB containing 100 μg mL^{-1} final concentration of ampicillin and incubated at 37 °C for 16 h. Glycerol stocks were prepared and plasmids purified from these overnight cultures. The purified plasmids were confirmed as containing the correct gene inserts by PCR with primers specific for each gene (Tables 8.1 and 8.2) and by restriction digests with appropriate enzyme(s). Plasmids that gave both the expected sized PCR products and expected sized DNA fragments resulting from the restriction digests were sent for sequencing by GATC Biotech, during which the gene of interest was sequenced using the T7 primer pairs (Table 8.1).

8.5.2 Cloning of *SgbnB* into pJDS05 and *SgbnA* into pJDS06

Synthetic *S. venezuelae gbnA* was amplified from pMK-RQ+*sven_gbnA* using primers N/N' and Synthetic *S. venezuelae gbnB* was amplified from pMK-RQ+*sven_gbnB* using primers O/O' (Table 8.1). PCR products were gel extracted from 1% agarose gel and concentrations determined by use of the Thermo NanodropLite. Plasmids pJDS05 and pJDS06 were digested with *Sac*I, followed by treatment with shrimp alkaline phosphatase (1 μL, 37 °C, 30 min). Digested plasmid was run on a 0.8% agarose gel and extracted using the gel extraction kit (Fermentas).

Following the product information protocol, the digested pJDS05 was mixed together with the synthetic *gbnB* PCR and the digested pJDS06 was mixed with the synthetic

Table 8.2 – pET-151-derived plasmids constructed during this study

pJDS	*template*	*gene amplified*	*primers used*	*enzyme used*	*length / bp*
02	pMK-RQ	synthetic *seag_gbnB (seag_B2133)*	K/K'	*Nde*I	7271
04	pMK-RQ	synthetic *smug_gbnB (smugs5_01685)*	L/L'	*Pst*I	7239
05	pMK-RQ	synthetic *sven_gbnA* (*sven_4179*)	M/M'	*Sph*I	7448
06	pMK-RQ	synthetic *sven_gbnB* (*sven_4180*)	J/J'	*Nde*I	7374
08	SV4 F01	*sven_sgnL* (*sven_4183*)	S/S'	*Pst*I	6802

Table 8.3 – Conditions for ligations of *Sac*I digested products and plasmids

for pJDS05 and synthetic *gbnB*:

digested PCR product / fmol	*digested plasmid / fmol*	*alkaline phosphatase used?*	*Ligation mixtures diluted prior to transformations?*
24.7	8.48	No	x1, x10, x100
24.7	8.44	Yes	x1, x10
74.1	8.48	Yes	x1, x10

for pJDS06 and synthetic *gbnA*:

digested PCR product / fmol	*digested plasmid / fmol*	*alkaline phosphatase used?*	*Ligation mixtures diluted prior to transformations?*
22.4	6.80	No	x1, x10, x100
24.4	8.00	Yes	x1, x10
68.2	6.80	Yes	x1, x10
73.2	8.10	Yes	x1, x10

gbnA PCR product. These solutions were incubated (25 °C, 2 h) with 0.1 μL (0.1 U) T4 DNA ligase (see Table 8.3 for conditions tested). The ligation product (2 μL) was used to transform 25 μL *E. coli* DH5α, and colonies were picked, grown overnight and plasmids purified from them. Plasmids were checked for the correct insert by PCR and restriction digest with *Nde*I.

8.5.3 Construction of plasmids pJDS09-pJDS11 from Duet vectors

pET-Duet and pACYC-Duet (Novagen) were used to transform *E. coli iba* StarGate cells using ampicillin and chloramphenicol to select for positive colonies for pET-Duet and pACYC-Duet, respectively. Plasmids were purified and verified by restriction digest with *Acc*I.

Primers P/P' (Table 8.4) were used to amplify synthetic *gbnA* from the pMK-RQ template. Purified PCR product was digested: DNA (30 μL), NEB CutSmart buffer (4 μL) and *Bam*HI-HF/*Hind*III-HF (2 μL each) were combined and incubated at 37 °C for 1 h. pET-Duet (15 μL) was also digested in NEB CutSmart buffer (2 μL) using *Bam*HI-HF/*Hind*III-HF (1 μL each) and incubated at 37 °C for 1 h. Digested PCR product and

Table 8.4 – Primers used for construction and sequencing of plasmids pJDS09-pJDS11. Restriction site sare shown in italics

Cloning primers:

Primer	*Gene to amplify*	*Sequence*
P	synthetic *gbnA*	5′-AATT*GGATCC*GACCGCAAGCACCAG-3′
P’	synthetic *gbnA*	5′-ATAT*AAGCTT*TTAGCTGGTCGCAACT-3′
Q	synthetic *gbnB*	5′-TATA*GGCCGGCC*ACATCAAAATTGCCGATATC-3′
Q’	synthetic *gbnB*	5′-ATAT*GGTACC*TTAAACACCAACGGTTTCT-3′
R	synthetic *gbnC*	5′-ATTA*GGATCC*GAATAGCCAGAATACCTATACC-3′
R’	synthetic *gbnC*	5′-TATA*AAGCTT*TTAGCTACGGGTGCC-3′

Sequencing primers:

Primer	*Gene to amplify*	*Sequence*
pETup	See Appendix 3	5′-ATGCGTCCGGCGTAG-3′
pACYCup	See Appendix 3	5′-GGATCTCGACGCTCTCCCT-3′
UP2	See Appendix 3	5′-TGTACACGGCCGCATAATC-3′
DOWN1	See Appendix 3	5′-GATTATGCGGCCGTGTACA-3′
T7_term	See Appendix 3	5′-GCTAGTTATTGCTCAGCG-3′

pET-Duet were re-purified (gel extraction kit), followed by ligation with T4 ligase (21.8 fmol digested synthetic *gbnA* and 8.95 fmol digested pET-Duet were incubated with 4 μL T4 ligase buffer and 1 μL T4 ligase in total reaction volume of 20 μL at 4 °C for 24 h).

The ligation mixture (16 μL) was used to transform *E. coli iba* StarGate cells, positive colonies picked from selection plates, grown overnight and plasmid pJDS09 purified from these overnight cultures. Plasmids were verified as containing the correct insert by PCR using the specific primers ‘pETup’ and ‘T7_term’ (Table 8.4) used to amplify synthetic *gbnA*, and also by restriction digest (15 μL plasmid was incubated with 2 μL buffer and 0.5 μL *Pst*I (FastDigest) at 37 °C for 30 min). Plasmids that gave both the expected sized PCR products and expected sized DNA fragments resulting from the restriction digests were sent for sequencing (GATC biotech) using primers ‘pETup’ (forwards direction) and ‘pET-RP’ (reverse direction).

For the subsequent cloning of synthetic *gbnB* into MCS 2 of pET-Duet, synthetic *gbnB* was amplified from pMK-RQ using specific primers Q/Q' to introduce *Fse*I and *Kpn*I restriction sites at the 5' and 3' ends of synthetic *gbnB*, respectively. PCR product and pJDS09 (25 µL each) were digested with *Fse*I and *Kpn*I (2 µL each, 3 h with 3 µL NEB Buffer 1.1) and digestion products re-purified. 113 fmol digested synthetic *gbnB* and 35.5 fmol digested pJDS09 were incubated with 4 µL T4 ligase buffer and 1 µL T4 ligase in total reaction volume of 20 µL at 4 °C for 24 h. The ligation mixture (16 µL) was used to transform *E. coli iba* StarGate cells, positive colonies picked from selection plates, grown overnight and the plasmid pJDS10 purified from these overnight cultures. Plasmids were verified as containing the correct insert by digestion with *Hind*III, and plasmids that gave the expected band pattern were sent for sequencing (GATC biotech) using primers 'UP2' (forwards direction) and 'T7_term' (reverse direction), see Table 8.4.

Primers R/R' (Table 8.4) were each used to amplify synthetic *gbnC* by PCR. PCR product and pACYC-Duet (25 µL each) were digested with *Bam*HI and *Hind*III (2 µL each, 4 h with 3 µL NEB CutSmart buffer) and digestion products re-purified. 78.9 fmol digested synthetic *gbnC* and 21.2 fmol digested pACYC-Duet were incubated with 4 µL T4 ligase buffer and 1 µL T4 ligase in total reaction volume of 20 µL at 4 °C for 24 h.

16 µL of the ligation mixture was used to transform *E. coli iba* StarGate cells, positive colonies picked from selection plates, grown overnight and the plasmid pJDS11 purified from these overnight cultures. Plasmids were verified as containing the correct insert both by digestion with *Acc*I and by double digestion with *Bam*HI/*Hind*III. Plasmids that gave the expected band pattern were sent for sequencing (GATC biotech)

Table 8.5 – Duet-derived co-expression vectors constructed during this study

pJDS	*parent vector*	*gene inserted*	*primers used*	*enzymes used*	*length / bp*
09	pET-Duet	synthetic *gbnA (sven_4179)*	P/P'	*Bam*HI, *Hind*III	7064
10	pJDS09	synthetic *gbnB (sven_4180)*	Q/Q'	*Fse*I, *Kpn*I	8645
11	pACYC-Duet	synthetic *gbnC (sven_4181)*	R/R'	*Bam*HI, *Hind*III	5040

using primers 'pACYCup' and 'T7_term' (Table 8.4). A summary of the plasmids pJDS09-pJDS11 constructed from Duet cloning vectors are shown in Table 8.5.

9. References

1. M. Ventura, C. Canchaya, A. Tauch, G. Chandra, G. F. Fitzgerald, K. F. Chater and D. van Sinderen, *Microbiol. Mol. Biol. R.* 2007, **71**, 495–548

2. P Dyson, '*Streptomyces*' *Encyclopedia of Microbiology*, *3rd Ed, Academic Press, Waltham* **2009**

3. D. A. Hopwood, *Microbiology* 1999, **145**, 2183-2202

4. K. Flärdh and M. J. Buttner, *Nat. Rev. Microbiol.* 2009, **7**, 36–49

5. D. Glazebrook, J. Doull, C. Stuttard and L. Vining, *Microbiology* **1990**, 136, 581-588

6. M. J. Bibb, A. Domonkos, G. Chandra and M. J. Buttner, *Mol. Microbiol.* 2012, **84**, 1033–1049

7. J. P. Gomez-Escribano and M. J. Bibb, *J. Ind. Microbiol. Biotechnol.* 2014, **41**, 425–431

8. D. A. Hopwood, '*Streptomyces* in Nature and Medicine: The Antibiotic Makers' *Oxford University Press, New York* **2007**

9. J. Houbraken, J. C. Frisvad and R. A. Samson, *IMA Fungus* 2011, **2**, 87–95

10. A. Schatz, E. Bugie & S. Waksman, *Proc. Soc. Exp. Biol. Med.* 1944, **55**, 66–69

11. B. Singh, and D. A. Mitchison, *British Medical Journal*, 1954, **1**(4854), 130-132

12. R. W. Fairbrother and J. E. Southall, *J. Clin. Path.* 1951, **4**, 183–189

13. D. Gottlieb, P. W. Robbins and H. E. Carter, *J. Bacteriol.* 1956, **72**, 153-156

14. K. Izaki, I. K. A. N. Kiuchi and K. E. I. Arima, *J. Bacteriol.* 1966, **91**, 628–633

15. D. J. Newman and G. M. Cragg, *J. Nat. Prod.* 2007, **70**, 461–477

16. H. Zhu, S. K. Sandiford and G. P. van Wezel, *J. Ind. Microbiol. Biotechnol.* 2014, **41**, 371–386

17. R. W. Burg, B. M. Miller, E. E. Baker, J. Birnbaum, A. Currie, R. Hartman, Y. Kong, R. L. Monaghan, G. Olson, I. Putter, J. B. Tunac, H. Wallick, E. Stapley and S. Omura, *Antimicrob. Agents Ch.* 1979, **15**, 361–367

18. T. Murakami, H. Anzai, S. Imai, A. Satoh, K. Nagaoka and C. J. Thompson, *Mol. Gen. Genet.* 1986, **205**(1), 42–50

19. D. R. D. Bignell, R. F. Seipke, J. C. Huguet-tapia, A. H. Chambers, R. J. Parry and R. Loria, *Mol. Plant Microbe Interact.* 2010, **23**(2), 161–175

20. D. J. Payne, M. N. Gwynn, D. J. Holmes and D. L. Pompliano, *Nat. Rev. Drug Discov.* 2007, **6**, 29-40

21. A. J. Alanis, *Arch. Med. Res.* 2005, **36**, 697-705

22. J. Garau, D. P. Nicolau, B. Wullt and M. Bassetti, *J. Global Antimicrobial Res.* 2014, **2**, 245-253

23. M. A. Cooper and D. Shales, *Nature* 2011, **472**, 32

24. S. J. Projan, *Curr. Opin. Microbiol.* 2003, **6**, 427-430

25. J. Trias, *Curr. Opin. Microbiol.* 2001, **4**, 520-525

26. K. Lewis, *Nature* 2012, **485**, 439-440

27. U. Theuretzbacher, *Int. J. Antimicrob. Ag.* 2009, **34**, 15-20

28. J. Clardy, M. A. Fischbach and C. T. Walsh, *Nat. Biotechnol.* 2006, **24**, 1541–1550

29. R. H. Baltz, *Microbe* 2007, **2**, 125–131

30. G. Bonadonna, S. Monfardini, M. de Lena and F. Fossati-Bellani, *Brit. Med. J.* 1969, **3**, 503-506

31. H. Umezawa, Y. Suhara, T. Taketa and K. Maeda, *J. Antibiotic* 1966, **19**, 210-215

32. S. N. Sehgal, H. Baker and C. Vezina, *J. Antibiotic* 1975, **28**, 727-732

33. P. Caffrey, S. Lynch, E. Flood and S. Finnan, *Chem. Biol.* 2001, **8**, 713–723

34. V. Yoon and J. R. Nodwell, *J. Ind. Microbiol. Biotechnol.* 2014, **41**, 415–24

35. S. Rochfort, *J. Nat. Prod.* 2005, **68**, 1813-1820

36. R. S. Varghese, A. Cheema, P. Cheema, M. Bourbeau, L. Tuli, B. Zhou, M. Jung, A. Dritschilo and H. W. Ressom, *J. Proteome Res.* 2010, **9**, 2786-2793

37. C. T. Walsh and M. A. Fischbach, *J. Am. Chem. Soc.* 2010, **132**, 2469-2493

38. T. Ito, T. Odake, H. Katoh, Y. Yamaguchi and M. Aoki, *J. Nat. Prod.* 2011, **74**, 983–988

39. S. W. Ayer, A. G. McInnes, P. Thibault, J. A. Walter, J. L. Doull, T. Parnell and L. C. Vining, *Tetrahedron Lett.* 1991, **32**(44), 6301-6304

40. J. L. Doull and L. C. Vining, *Appl. Microbiol. Biotechnol.* 1990, **32**, 449–454

41. J. F. Martín, *J. Bacteriol.* 2004, **186**, 5197–5201

42. H. B. Bode, B. Bethe, R. Höfs and A. Zeeck, *ChemBioChem* 2002, **3**, 619–627

43. O. E. Christian, J. Compton, K. R. Christian, S. L. Mooberry, F. A. Valeriote and P. Crews, *J. Nat. Prod.* 2005, **68**, 1592–1597

44. Y. Liao, Z.-H. Wei, L. Bai, Z. Deng and J.-J. Zhong, *J. Biotechnol.* 2009, **142**, 271-274

45. H. Schiewe and A. Zeeck, *J. Antibiotic* 1999, **52**, 635-642

46. M. E. Rateb, W. E. Houssen, W. T. A. Harrison, H. Deng, C. K. Okoro, J. A. Asenjo, B. A. Andrews, A. T. Bull, M. Goodfellow, R. Ebel and M. Jaspars, *J. Nat. Prod.* 2011, **74**, 1965–1971

47. L. F. Wright and D. A. Hopwood, *J. Gen. Microbiol.* 1976, **95**, 96-106

48. G. Hobbs, A. I. C. Obanye, J. Petiy, J. C. Mason, E. Barratt, D. C. J. Gardner, F. Flett, C. P. Smith, P. Broda and S. G. Oliver, *J. Bacteriol.* 1992, **174**, 1487–1494

49. A. Craney, C. Ozimok, S. M. Pimentel-Elardo, A. Capretta and J. R. Nodwell, *Chem. Biol.* 2012, **19**, 1020–1027

50. S. Ahmed, A. Craney, S. M. Pimental-Elardo and J. R. Nodwell, *ChemBioChem* 2013, **14**(1), 83-91

51. Y. Imai, S. Sato, Y. Tanaka, K. Ochi and T. Hosaka, *Appl. Envrion. Microbiol.* 2015, DOI: 10.1128/AEM.04214-14

52. K. Yamanaka, H. Oikawa, H. Ogawa, K. Hosono, F. Shinmachi, H. Takano, S. Sakuda, T. Beppu and K. Ueda, *Microbiology* 2005, **151**, 2899–905

53. K. Ueda, S. Kawai, H. Ogawa, A. Kiyama, T. Kubota, H. Kawanobe and T. Beppu, *J. Antibiotic* 2000, **53**(9), 979-982

54. A. Marmann, A. H. Aly, W. Lin, B. Wang and P. Proksch, *Mar. Drugs* 2014, **12**, 1043–1065

55. A. S. Khokhlov, I. I. Tovarova, L. N. Borisova, S. A. Pliner, L. N. Shevchenko, E. Kornitskaia, N.S. Ivkina and I. A. Rapoport, *Dokl. Akad. Nauk. SSSR* 1967, **177**, 232–235

56. G. P. van Wezel and K. J. McDowall, *Nat. Prod. Rep.* 2011, **28**, 1311–1333

57. M. J. Bibb, *Curr. Opin. Microbiol.* 2005, **8**, 208-215

58. L. L. Ling, T. Schneider, A. J. Peoples, A. L. Spoering, I. Engels, B. P. Conlon, A. Mueller, T. F. Schäberle, D. E. Hughes, S. Epstein, M. Jones, L. Lazarides, V. a Steadman, D. R. Cohen, C. R. Felix, K. A. Fetterman, W. P. Millett, A. G. Nitti, A. M. Zullo, C. Chen and K. Lewis, *Nature* 2015, **517**, 455-459

59. R. J. Scheffler, S. Colmer, H. Tynan, a L. Demain and V. P. Gullo, *Appl. Microbiol. Biot.* 2013, **97**, 969–978

60. F. Malpartida and D. A. Hopwood, *Nature* 1984, **309**, 462-464

61. S. D. Bentley, K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch,M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Weitzorrek, J. Woodward, B. G. Barrell and J. Parkhill and D.A. Hopwood, *Nature* 2002, **417**, 141-147

62. H. Ikeda, J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori and S. Omura, *Nat. Biotechnol.* 2003, **21**, 526–531

63. S. T. Pullan, G. Chandra, M. J. Bibb and M. Merrick, *BMC Genomics* 2011, **12**, 175-188

64. Y. Ohnishi, J. Ishikawa, H. Hara, H. Suzuki, M. Ikenoya, H. Ikeda, A. Yamashita, M. Hattori and S. Horinouchi, *J . Bacteriol.* 2008, **190**, 4050–4060

65. B. O. Bachmann, S. G. Van Lanen and R. H. Baltz, *J. Ind. Microbiol. Biotechnol.* 2014, **41**, 175–184

66. C. Corre and G. L. Challis, *Nat. Prod. Rep.* 2009, **26**, 977-986

67. C. Corre and G. L. Challis in '*Comprehensive Natural Products II Chemistry and Biology Volume 1*', *1st ed, Elsevier* **2010**, pp429-453

68. M. Zerikly and G. L. Challis, *ChemBioChem.* 2009, **10**, 625-633

69. B. Aigle, S. Lautru, D. Spiteller, J. S. Dickschat, G. L. Challis, P. Leblond and J.-L. Pernodet, *J. Ind. Microbiol. Biotechnol.* 2014, **41**, 251–263

70. M. H. Medema, E. Takano and R. Breitling, *Mol. Biol. Evol.* 2013, **30**, 1218-1223

71. K. Blin, M. H. Medma, D. Kazempour, M. A. Fischbach, R. Brietling, E. Takano and T. Weber, *Nucl. Acids Res.* 2013, **41**, W204-W212

72. M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, *Nucl. Acids Res.* 2011, **39**, W362-W367

73. G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol.* 2000, **7**, 211-224

74. S. Lautru, R. J. Deeth, L. M. Bailey, and G. L. Challis, *Nat. Chem. Biol*. 2005, **1**, 265-269

75. H.-M. Park, B.-G. Kim, D. Chang, S. Malla, H.-S. Joo, E.-J. Kim, S.-J. Park, J. K. Sohng and P. I. Kim, *Appl. Microbiol. Biotechnol.* 2013, **97**, 1213–1222

76. J. Claesen and M. J. Bibb, *J. Bacteriol.* 2011, **193**, 2510–2516

77. A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. a Yamashita, D. Zhang, C. Zheng and S. H. Bryant, *Nucl. Acids Res.* 2014, **43**, D222–226

78. H. Gross, V. O. Stockwell, M. D. Henkels, B. Nowak-Thompson, J. E. Loper and W. H. Gerwick, *Chem. Biol.* 2007, **14**, 53–63

79. C. Corre, L. Song, S. O. Rourke, K. F. Chater and G. L. Challis, *Proc. Natl. Acad. Sci. USA* 2008, **105**(45), 17510-17515

80. X. Lin, R. Hopson and D. E. Cane, *J. Am. Chem. Soc.* 2006, **128**, 6022–6023

81. J. P. Gomez-Escribano and M. J. Bibb, *Microbial Biotechnol.* 2011, **4**, 207–215

82. B. Gust, G. L. Challis, K. Fowler, T. Kieser and K. F. Chater, *Proc. Natl. Acad. Sci. USA* 2003, **100**, 1541-1546

83. L. Song, F. Barona-Gomez, C. Corre, L. Xiang, D. W. Udwary, M. B. Austin, J. P. Noel, B. S. Moore and G. L. Challis, *J. Am. Chem. Soc.* 2006, **128**, 14754–14755

84. B. Aigle and C. Corre, *Methods in Enzymol.* 2012, **517**, 343-366

85. S. O'Rourke, A. Wietzorrek, K. Fowler, C. Corre, G. L. Challis and K. F. Chater, *Mol. Microbiol.* 2009, **71**, 763-778

86. L. Laureti, L. Song, S. Huang, C. Corre, P. Leblond, G.L. Challis and B. Aigle, *Proc. Natl. Acad. Sci. USA* 2011, **108**, 6258-6263

87. A. Aroonsri, S. Kitani, J. Hashimoto, I. Kosone, M. Izumikawa, M. Komatsu, N. Fujita, Y. Takahashi, K. Shin-ya, H. Ikeda and T. Nihira, *Appl. Environ. Microbiol.* 2012, **78**(22), 8015–8024

88. M. Gottelt, S. Kol, J. P. Gomez-Escribano, M. J. Bibb and E. Takano, *Microbiology* 2010, **156**, 2243-2253

89. J.-P. Gomez-Escribano, L. Song, D. J. Fox, V. Yeo, M. J. Bibb and G. L. Challis, *Chem. Sci.* 2012, **3**, 2716-2720

90. Y. Ohnishi, H. Yamazaki, J.-Y. Kato, A. Tomono and S. Horinouchi, *Biosci. Biotechnol. Biochem.* 2005, **69**, 431–439

91. Y. Onishi, S. Kameyama, H. Onaka and S. Horinouchi, *Mol. Microbiol.* 1999, **34**(1), 102-111

92. L. Cuthbertson and J. R. Nodwell, *Microbiol. Mol. Bol. Rev.* 2013, **77**, 440–475

93. S. E. Maddocks and P. C. F. Oyston, *Microbiology* 2008, **154**, 3609–3623

94. Y. Mast, T. Weber, M. Gölz, R. Ort-Winklbauer, A. Gondran, W. Wohlleben and E. Schinko, *Microb. Biotechnol.* 2011, **4**, 192–206

95. Z. Zou, D. Du, Y. Zhang, J. Zhang, G. Niu and H. Tan, *Mol. Microbiol.* 2014, **94**, 490–505

96. J. L. Ramos, M. Martínez-Bueno, J. Antonio, M. Martı, A. J. Molina-Henares and W. Tera, *Microbiol. Mol. Biol. Rev.* 2005, **69**, 326-356

97. W. Hillen, G. Klock, I. Kaffenberger, L. V Wray and W. S. Reznikoff, *J. Biol. Chem.* 1982, **257**, 6605–6613

98. I. Meier, L. V Wray and W. Hillen, *EMBO J.* 1988, **7**, 567–572

99. L. Cuthbertson, S. K. Ahn and J. R. Nodwell, *Chem. Biol.* 2013, **20**, 232–240

100. a) P. J. Harrison, N. Malet, D. Rea, S. Zhou, K. Styles, V. Fülop, G. L. Challis and C. Corre, *manuscript in preparation* b.) N. Malet, PhD thesis, *University of Warwick* **2012**

101. C. Corre, *Chem. Biol.* 2013*,* **20***,* 140-142

102. H. Onaka, N. Ando, T. Nihira, Y. Yamada, T. Beppu and S. Horinouchi, *J. Bacteriol.* 1995, **177**(210), 6083-6092

103. N.-H. Hsiao, M. Gottelt and E. Takano, *Methods Enzymol.* 2009, **458**, 143–157

104. E. Takano, *Curr. Opin. Microbiol.* **2006**, 9, 287-294

105. J. M. Willey and A. A. Gaskell, *Chem. Rev*. 2011, **111**, 174-187

106. H. Onaka and S. Horinouchi, *Mol. Microbiol.* 1997, **24**(5), 991–1000

107. Y. Ohnishi, Y. Furusho, T. Higasi, H.-K. Chun, K. Furihata, S. Sakuda and S. Horinouchi, *J. Antibiotic* 2004, **57**, 218-223

108. N. Ando, N. Matsumori, S. Sakuda, T. Beppu and S. Horinouchi, *J. Antibiotic* 1997, **50**, 847-852

109. J. Kato, N. Funa, H. Watanabe, Y. Ohnishi and S. Horinouchi, *Proc. Nat. Acad. Sci.* 2007, **104**(7), 2378-2383

110. D. D'Alia, D. Eggle, K. Nieselt, W.-S. Hu, R. Breitling and E. Takano, *Microbial Biotechnol.* 2011, **4**, 239–251

111. E. Takano, T. Nihira, Y. Hara, J. J. Jones, C. J. Gershater, Y. Yamada and M. Bibb, *J. Biol. Chem.* 2000, **275**, 11010–11016

112. N.H. Hsiao, S. Nakayama, M. E. Merlo, M. de Vries, R. Bunet, S. Kitani, T. Nihira, E. Takano, *Chem. Biol.* 2009, **16**, 951-960

113. K. Sato, T. Nihira, S. Sakuda, M. Yanagimoto and Y. Yamada, *J. Ferment. Bioeng.* 1989, **68**, 170–173

114. S. Kitani, A. Iida, T. Izumi, A. Maeda, Y. Yamada and T. Nihira, *Gene* 2008, **425**, 9–16

115. Y. Yamada, K. Sugamura, K. Kondo, M. Yanagimoto, H. Okada, *J. Antibiotics* 1987, **40**, 496-504

116. R. Kawachi, T. Akashi, Y. Kamitani, A. Sy, U. Wangchaisoonthorn, T. Nihira and Y. Yamada, *Mol. Microbiol.* 2000, **36**, 302–313

117. S. Kitani, K. T. Miyamoto, S. Takamatsu, E. Herawati, H. Iguchi, K. Nishitomi, M. Uchida, T. Nagamitsu, S. Omura, H. Ikeda and T. Nihira, *Proc. Natl. Acad. Sci.* 2011, **108**, 16410–16415

118. K. Arakawa, N. Tsuda, A. Taniguchi and H. Kinashi, *ChemBioChem* 2012, **13**, 1447–1457

119. R. Bunet, M. V Mendes, N. Rouhier, X. Pang, L. Hotel, P. Leblond and B. Aigle, *J. Bacteriol.* 2008, **190**, 3293–3305

120. E. Mingyar, L. Feckova, R. Novakova, C. Bekeova and J. Kormanec, *Appl. Microbiol. Biotechnol.* 2015, **99**, 309–325

121. F. G. Healy, K. P. Eaton, P. Limsirichai, J. F. Aldrich, A. K. Plowman and R. R. King, *J. Bacteriol.* 2009, **191**, 4786–4797

122. D. R. D. Bignell, N. Bate and E. Cundliffe, *Mol. Microbiol.* 2007, **63**, 838–847

123. U. Gräfe, W. Schade, I. Eritt and W. F. Fleck, *J. Antibiotics.* 1982, **35**(12), 1722-1723

124. U. Gräfe, G. Reinhardt, W. Schade, I. Eritt, W. F. Fleck and L. Radics, *Biotechnol. Lett.* 1983, **5**(9), 591-596

125. S. Kitani, H. Kinoshita, T. Nihira and Y. Yamada, *J. Bacteriol.* 1999, **181**(16), 5081-5084

126. H. Kinoshita, H. Ipposhi, S. Okamoto, H. Nakano and T. Nihira, *J. Bacteriol.* 1997, **179**(22), 6986-6993

127. G. Wei, N. Zhu, Y. Zeng, Y. Shen and P. Zhao, *Ann. Microbiol.* 2010, **60**, 249–253

128. H. Nishida, Y. Ohnishi, T. Beppu and S. Horinouchi, *Environ. Microbiol* 2007, **9**, 1986–1994

129. S. Kitani, M. Doi, T. Shimizu, A. Maeda and T. Nihira, *Arch. Microbiol.* 2010, **192**, 211–220

130. O. Hara and T. Beppu, *J. Antibiotics* 1982, **35**(9), 1208-1215

131. S. Sakuda, A. Higashi, S. Tanaka, T. Nihira and Y. Yamada, *J. Am. Chem. Soc.* 1992, **114**, 663–668

132. N.-H. Hsiao, J. Soding, D. Linke, C. Lange, C. Hertweck, W. Wohlleben and E. Takano, *Microbiology* 2007, **153**, 1394-1404

133. S. Horinouchi, Y. Kumada and T. Beppu, *J. Bacteriol.* 1984, **158**, 481–487

134. S. Sakuda, S. Tanaka, K. Mizuno, O. Sukcharoen, T. Nihira and Y. Yamada, *J. Chem. Soc. Perkin. Trans. 1* 1993, **19**, 2309-2315

135. G. L. Challis, *Microbiology* 2008, **154**, 1555–1569

136. Y. Li, G. Florova and K. A. Reynolds, *J. Bacteriol.* 2005, **187**, 3795–3799

137. E. Takano, R. Chakraburtty, T. Nihira, Y. Yamada and M. J. Bibb, *Mol. Microbiol.* 2001, **41**, 1015–1028

138. N. Shikura, J. Yamamura and T. Nihira, *J. Bacteriol.* 2002, **184**(18), 5151-5157

139. Y. J. Lee, S. Kitani, H. Kinoshita and T. Nihira, *Arch. Microbiol.* 2008, **189**, 367–374

140. C. Corre, S. W. Haynes, N. Malet, L. Song and G. L. Challis, *Chem. Commun.* 2010, **46**, 4079-4081

141. J. B. Davis, J. D. Bailey and J. K. Sello, *Org. Lett.* 2009, **11**, 2984–2987

142. M. Waki, T. Nihira and Y. Yamada, *J. Bacteriol.* 1997, **179**(16), 5131-5137

143. J. R. Nodwell, *Mol. Microbiol.* 2014, **94**, 483–485

144. M. Folcher, H. Gaillard, L. T. Nguyen, K. T. Nguyen, P. Lacroix, N. Bamas-Jacques, M. Rinkel and C. J. Thompson, *J. Biol. Chem.* 2001, **276**, 44297–44306

145. G.-Y. Tan, L. Bai and J.-J. Zhong, *Biotechnol. Bioeng.* 2013, **110**, 2984–2993

146. G. Xu, J. Wang, L. Wang, X. Tian, H. Yang, K. Fan, K. Yang and H. Tan, *J. Biol.Chem.* 2010, **285**, 27440–27448

147. T. L. Bailey, M. Boden, F. a Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li and W. S. Noble, *Nucl. Acids Res.* 2009, **37**, W202–208

148. K. Matsuno, Y. Yamada, C.-K. Lee and T. Nihira, *Arch. Microbiol.* 2004, **181**, 52–59

149. J. Wang, W. Wang, L. Wang, G. Zhang, K. Fan, H. Tan and K. Yang, *Mol. Microbiol.* 2011, **82**, 236–250

150. G.-Y. Tan, Y. Peng, C. Lu, L. Bai and J.-J. Zhong, *Metabol. Eng.* 2015, **28**, 74–81

151. Q. Xu, G. P. van Wezel, H.-J. Chiu, L. Jaroszewski, H. E. Klock, M. W. Knuth, M. D. Miller, S. A. Lesley, A. Godzik, M.-A. Elsliger, A. M. Deacon and I. A. Wilson, *PloS One* 2012, **7**, e41359

152. J. D. Sidda, L. Song, V. Poon, M. Al-Bassam, O. Lazos, M. J. Buttner, G. L. Challis and C. Corre, *Chem. Sci.* 2014, **5**, 86-89

153. Y. Xue, L. Zhao, H.-W. Liu and D. H. Sherman, *Proc. Natl. Acad. Sci. USA* 1998, **95**, 12111-12116

154. J. Ren, D. Liu, L. Tian, Y. Wei, P. Proksch, J. Zeng and W. Lin, *Bioorg. Med. Chem. Lett.* 2013, **23**(1), 301-304

155. D. A. Hopwood, T. Keiser, M. J. Bibb, M. J. Buttner and K. Chater 'Practical *Streptomyces* Genetics' *2nd Ed. John Innes Foundation, Norwich* **2000**

156. M. Ma, J. Zhao, S. Wang, S. Li, Y. Yang, J. Shi, X. Fan and L. He, *J. Nat. Prod.* 2006, **69**, 206–210

157. T. Wrkamiya, Y. Kobayashi, T. Shiba, K. Setogawa and H. Matsutani, *Tetrahedron* 1984, **40**(1), 235-240

158. A. P. Kozikowski, F. Nan, P. Conti, J. Zhang, E. Ramadan, T. Bzdega, B. Wroblewska, J. H. Neale, S. Pshenichkin and J. T. Wroblewski, *J. Med. Chem.* 2001, **44**, 298–301

159. A. N. Kravchenko and I. E. Chikunov, *Russian Chem. Rev.* 2006, **75**, 191–206

160. T. Tayaka and Z.Tozuka, United States Patent **4,349,552** Sept 14th 1982

161. D. H. Williams and I. Fleming, 'Spectroscopic methods in organic chemistry' *5th Ed. McGraw-Hill, New York* **2007**

162. J. H. Mueller and P. A. Miller, *J. Immunol.* 1941, **40**, 21-32

163. W. Zhang, B. Ostash and C. T. Walsh, *Proc. Nat. Acad. Sci. USA* 2010, **107**(39), 16828-16833

164. W. Zhang, I. Ntai, M. L. Bolla, S. J. Malcolmson, D. Kahne, N. L. Kelleher and C. T. Walsh, *J. Am. Chem. Soc.* 2011, **133**, 5240–5243

165. H. J. Imker, C. T. Walsh and W. M. Wuest, *J. Am. Chem. Soc.* 2009, **131**, 18263–18265

166. Y. Konda, Y. Takahashi, S. Arima, N. Sato, K. Takeda, K. Dobashi, M. Baba and Y. Harigaya, *Tetrahedron*, 2001, **57**, 4311–4321

167. G. A. Gomez, C. Morisseau, B. D. Hammock and D. W. Christianson, *Protein Sci.* 2006, **15**, 58–64

168. L. Su, H. Fang, K. Yang, Y. Xu and W. Xu, *Bioorg. Med. Chem.* 2011, **19**, 900–906

169. G. A. Cook, *J. Biol. Chem.* 1987, **262**, 4968–4972

170. R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L.L. Sonnhammer, J. Tate and M. Punta, *Nucl. Acids Res.* 2014, **42**, D222-D230

171. M. L. Hackert, D. W. Carroll, L. Davidson, S. Kim, C. Momany, G. L. Vaaler and L. Zhang, *J. Bacteriol.* 1994, **176**, 7391–7394

172. J. Andréll, M. G. Hicks, T. Palmer, E. P. Carpenter, S. Iwata and M. J. Maher, *Biochemistry* 2009, **48**, 3915–3927

173. A. M. Gulick, V. J. Starai, A. R. Horswill, K. M. Homick and J. C. Escalante-Semerena, *Biochemistry* 2003, **42**, 2866–2873

174. E. Conti, T. Stachelhaus, M. A. Marahiel and P. Brick, *EMBO J.* 1997, **16**, 4174–4183

175. I. Franke, A. Resch, T. Daßler, T. Maier and A. Böck, *J. Bacteriol.* 2003, **185**, 1161–1166

176. L. Inbar and A. Lapidot, *J. Bacteriol.* 1991, **173**, 7790-7801

177. S. Baumberg, H. Krugel and S. Noack, 'Genetics and product formation in *Streptomyces*' *Springer Science, New York*, **1991**

179. T. E. Tipple and L. K. Rogers, *Methods Mol. Biol*. 2012, **889**, 315-324

180. N. H. Tolia and L. Joshua-Tor, *Nat. Methods* 2006, **3**, 55-64

181. A. Radeghieri, M. Bonoli, F. Parmeggiani and A. Hochkoeppler, *Biochim. Biophys. Acta*, 2007, **1774**, 243–248

182. H. Joo, Y. Yang, C. Lee, J. Kim and B. Kim, *Rapid Commun. Mass Spectrom.* 2007, **21**, 764–770

183. K. Mizuno, S. Sakuda, T. Nihira and Y. Yamada, *Tetrahedron* 1994, **50**(37), 10849–10858

184. L. A. Kelley and M. J. E. Sternberg, *Nat. Protocols* 2009, **4**, 363-371

Appendices

Appendix 1 – UV spectrum of gaburedin A (**80**) obtained during the preliminary study

Appendix 2 – Genetic context of *gbnABC* orthologues in other microorganisms

Appendix 3 – Sequences of synthetic genes used in this study

Appendix 4 – Map of pET-Duet and pACYC-Duet multiple cloning sites

Appendix 5 – UHR-MS of GBLs identified in *S. coelicolor* M1152 extract

Appendix 6 – UHR-MS of AHFCAs identified in this study

Appendix 1 - UV spectrum of gaburedin A (80) obtained during the preliminary study

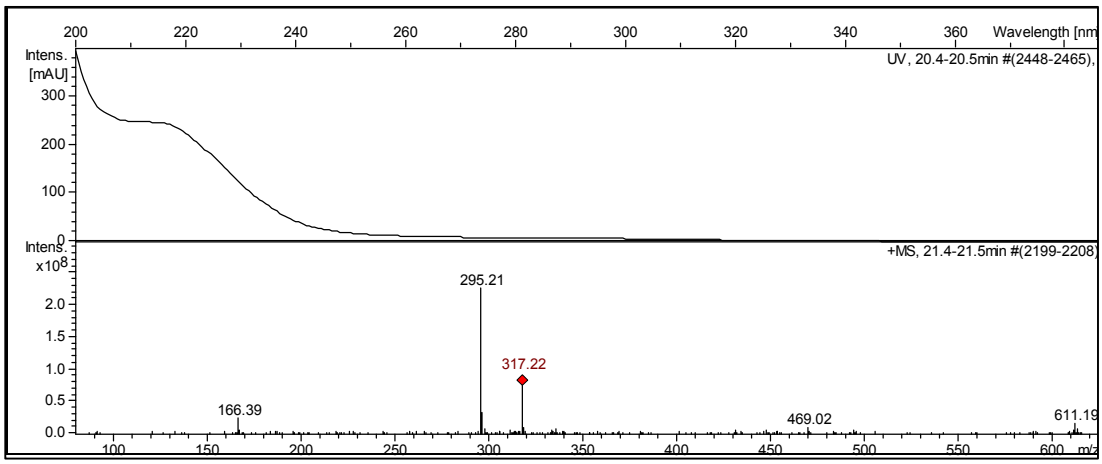
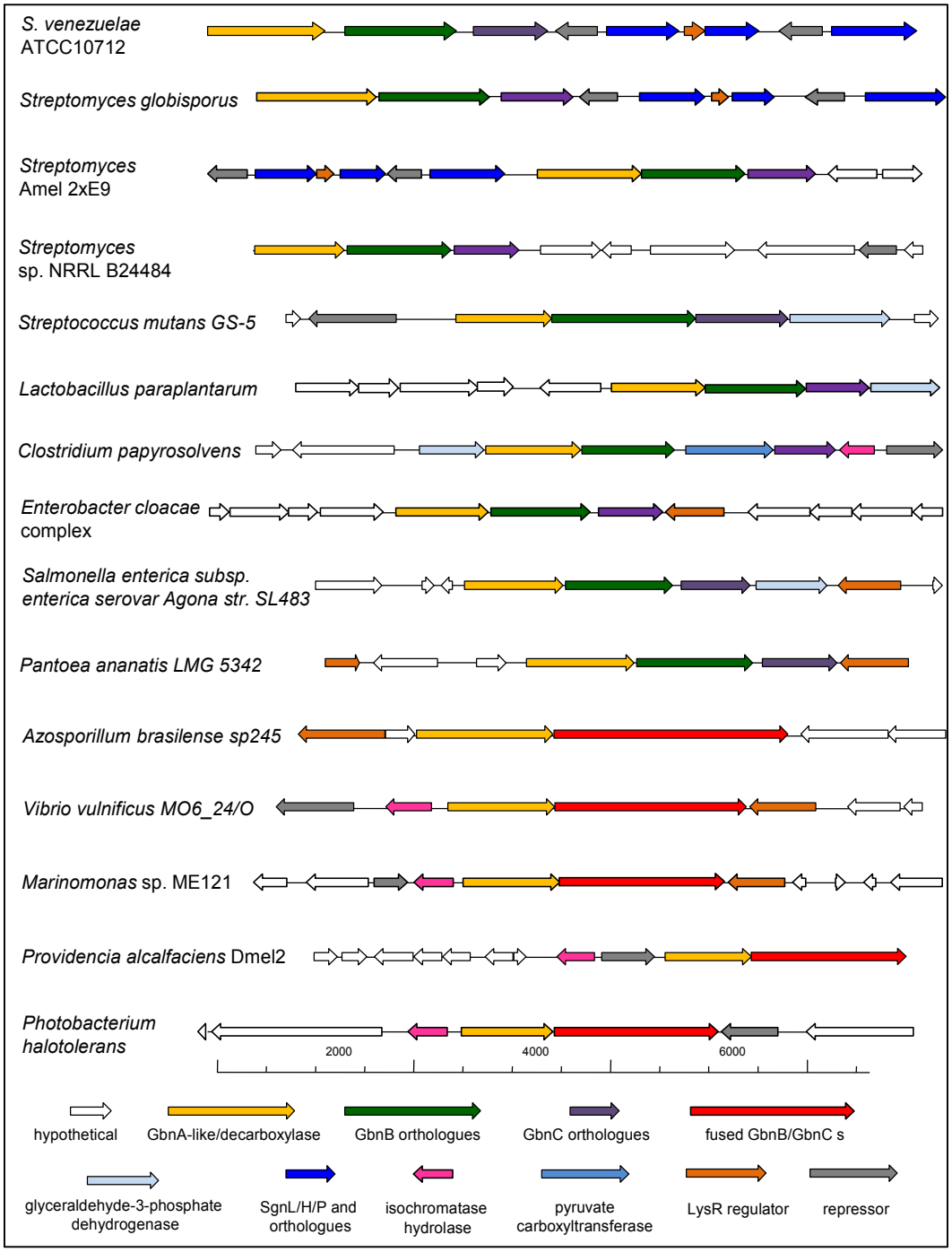


Figure T – UV and mass spectrum of compound **80** from preliminary data

Appendix 2 – Genetic context of *gbnABC* orthologues in other microorganisms



Appendix 3 - Sequences of synthetic genes used in this study

Synthetic *sven_4179* (*S. venezuelae gbnA*) native sequence is NCBI Gene ID: 13819337

```
  1  ATG ACC GCA AGC ACC AGC AGC ACA CCG GAA CCG GCA CAT CGT CCG GCA CGT CCG GGT GCA
     >>.............................synthetic_gbnA.............................>
       m   t   a   s   t   s   s   t   p   e   p   a   h   r   p   a   r   p   g   a

 61  TAT CCG GAT GGT CGT ACA CCG CTG GCA GAT GCA GTT CTG GCA GTT GCA CGT CGT GAT ATT
     >.............................synthetic_gbnA.............................>
       y   p   d   g   r   t   p   l   a   d   a   v   l   a   v   a   r   r   d   i

121  GCA ACC TTT CAT GCA CTG CCG CTG AGC CAT GGT CGT AGC ATT CGT GAT AGC CAT ATG CGT
     >.............................synthetic_gbnA.............................>
       a   t   f   h   a   l   p   l   s   h   g   r   s   i   r   d   s   h   m   r

181  GAA ACC TAT GAA GCA CTG TTT GGA GCA GCA CAG CTG GCA GCA GAT GTT AGC TAT AGC GGC
     >.............................synthetic_gbnA.............................>
       e   t   y   e   a   l   f   g   a   a   q   l   a   a   d   v   s   y   s   g

241  ACC ATG CTG GAT AGC TTT TTT CGT CCG CGT GGT CCG CTG CGT GAA GCA CAG CGT CTG GCA
     >.............................synthetic_gbnA.............................>
       t   m   l   d   s   f   f   r   p   r   g   p   l   r   e   a   q   r   l   a

301  GCA CGT AGC TTT GGT GCA GAT GCA ACC TTT TTT CTG AGC GCA GGC ACC AGC ACC GCA AAT
     >.............................synthetic_gbnA.............................>
       a   r   s   f   g   a   d   a   t   f   f   l   s   a   g   t   s   t   a   n

361  CGT GTT GCA CTG ACC GCA ATG ACC CGT CCG GGT AGC CGT GTT CTG GCA GAT CGT AGC TGT
     >.............................synthetic_gbnA.............................>
       r   v   a   l   t   a   m   t   r   p   g   s   r   v   l   a   d   r   s   c

421  CAT CAG AGC GTT CAT TTT GCA CTG AAT AGC CTG GGT GTT AGC GTT GCA TAT GCA CCG ATG
     >.............................synthetic_gbnA.............................>
       h   q   s   v   h   f   a   l   n   s   l   g   v   s   v   a   y   a   p   m

481  CAG CGT TGT TGT GAA GAT TGT CCG CGT ACC TAT GCA GAT CTG CCT CGT CTG CTG GAA ATG
     >.............................synthetic_gbnA.............................>
       q   r   c   c   e   d   c   p   r   t   y   a   d   l   p   r   l   l   e   m

541  TTT CGT GCA GCA GCA GTT GAA GGT CGT CCG TAT GAT ACC GTT GTG CTG AGC GCA GTT AGC
     >.............................synthetic_gbnA.............................>
       f   r   a   a   a   v   e   g   r   p   y   d   t   v   v   l   s   a   v   s

601  TAT GAT GGT GTT CGT TAT GAT CTG CCG ACC GTT CTG GCA GAA CTG GCA GCA GTT CAT CCG
     >.............................synthetic_gbnA.............................>
       y   d   g   v   r   y   d   l   p   t   v   l   a   e   l   a   a   v   h   p

661  GGT GTT GCA GTT CTG GTT GAT GAA GCA TGG GGT GCA GCA CAT CGT TTT CAT CCG CGT CTG
     >.............................synthetic_gbnA.............................>
       g   v   a   v   l   v   d   e   a   w   g   a   a   h   r   f   h   p   r   l

721  CGT CCG CTG ACC GCA CTG CAT GCC GTT GAA ACC CTG CGT GCA GCA GAT CCG CGT TTT ACC
     >.............................synthetic_gbnA.............................>
       r   p   l   t   a   l   h   a   v   e   t   l   r   a   a   d   p   r   f   t

781  ATG AAT GTT GCA GTT ACC CAT AGC GCA CAT AAA AGC ATG AGC GCA CTG CGT CAG GGT AGC
     >.............................synthetic_gbnA.............................>
       m   n   v   a   v   t   h   s   a   h   k   s   m   s   a   l   r   q   g   s

841  TAT CTG CAT CTG ATT GGT GAT GGT GAA GCA CGT GAA CGT ACC GCA CAG GCA CTG TTT CAG
     >.............................synthetic_gbnA.............................>
       y   l   h   l   i   g   d   g   e   a   r   e   r   t   a   q   a   l   f   q

901  CAT CAT ACC ACC AGT CCG AGC TGG CCT GTT CTG GCA AGC CTG GAT CTG GCA CGT CTG CAG
     >.............................synthetic_gbnA.............................>
       h   h   t   t   s   p   s   w   p   v   l   a   s   l   d   l   a   r   l   q

961  GCA GAA ACC GAA GGT GAA GCA CTG CTG GGT CGT AGT CTG GGG CTG GCA CGT ACC CTG CGT
     >.............................synthetic_gbnA.............................>
       a   e   t   e   g   e   a   l   l   g   r   s   l   g   l   a   r   t   l   r
```

```
1021  GTT GAA CTG GCA ACC GAT CCG CGT CTG AGC GCA TAT CGT CCG CTG GGT CCG GAT GGA CAT
      >..............................synthetic_gbnA..............................>
        v   e   l   a   t   d   p   r   l   s   a   y   r   p   l   g   p   d   g   h

1081  CTG ACC GAT CCT GCA CAG CTG GTT AGT GAT GAT CCG ACC CGT GTT CTG GTG GAT ATT AGC
      >..............................synthetic_gbnA..............................>
        l   t   d   p   a   q   l   v   s   d   d   p   t   r   v   l   v   d   i   s

1141  GCA CTG GGT ATT ACC GCA GCA GAT TTT CGT CGT ATG CTG TTT GAT GAA TAT GCA CTG TAT
      >..............................synthetic_gbnA..............................>
        a   l   g   i   t   a   a   d   f   r   r   m   l   f   d   e   y   a   l   y

1201  GTT GCA CGT GAA AGC GGT GAT GCC GTT CTG TTT CAT ATT CAT ATT GGT GTT GAT GAA GCC
      >..............................synthetic_gbnA..............................>
        v   a   r   e   s   g   d   a   v   l   f   h   i   h   i   g   v   d   e   a

1261  ACC CTG CTG CGT CTG CTG GAT GCA CTG CGT ACC ATT CAG CGT ACC TAT CGT AGC GCA AGC
      >..............................synthetic_gbnA..............................>
        t   l   l   r   l   l   d   a   l   r   t   i   q   r   t   y   r   s   a   s

1321  GAA GCA CTG GCA GCA GGT ACA AGC GAA CAT TTT ATC ATT GCA TAT CCG CCT GGT ATT CCG
      >..............................synthetic_gbnA..............................>
        e   a   l   a   a   g   t   s   e   h   f   i   i   a   y   p   p   g   i   p

1381  ATT ACC GTT CCG GGT GAA AAA CTG TGT GAA CGT ACC CTG AGC CAG ATT GAT ACC CTG CGT
      >..............................synthetic_gbnA..............................>
        i   t   v   p   g   e   k   l   c   e   r   t   l   s   q   i   d   t   l   r

1441  AGC AGT GGT TGT GAA ATC TAT ACC CTG CGC AAT CCG ACC GTT AGT GCC GGT GCA CTG GCT
      >..............................synthetic_gbnA..............................>
        s   s   g   c   e   i   y   t   l   r   n   p   t   v   s   a   g   a   l   a

1501  GCC CAG GCA GCA GCA ACA GGT GCA CTG CCT GTT CCG GCA GGT CAT CTG GCA GCA CCG CCT
      >..............................synthetic_gbnA..............................>
        a   q   a   a   a   t   g   a   l   p   v   p   a   g   h   l   a   a   p   p

1561  CCG ACC AGC GAA CGT CTG GCT GGT GCA GCA GCC GCA GTT AGC GCA GCA GGC ACC ACC GCA
      >..............................synthetic_gbnA..............................>
        p   t   s   e   r   l   a   g   a   a   a   a   v   s   a   a   g   t   t   a

1621  GCA ACC GTT AGC GCA GCT GCA GCA ACC GCA GTG GCA ACC GCA GTT GCG ACC AGC TAA
      >..............................synthetic_gbnA..............................>>
        a   t   v   s   a   a   a   a   t   a   v   a   t   a   v   a   t   s   -
```

Synthetic *sven_4180* (*S. venezuelae gbnB*) native sequence is NCBI Gene ID: 13819338

```
   1  ATG ATC AAA ATT GCC GAT ATC CGT AAA CAT GCA ACC CGT TGG CCT CAG CGT GTT GCA GTT
      >>..............................synthetic_gbnB..............................>
        m   i   k   i   a   d   i   r   k   h   a   t   r   w   p   q   r   v   a   v

  61  GTT GAT GGT GAA ACC CGT ATG ACC TGG GCA AGC TTT AGC GAT AAA GTT AGC CGT GTT AGC
      >..............................synthetic_gbnB..............................>
        v   d   g   e   t   r   m   t   w   a   s   f   s   d   k   v   s   r   v   s

 121  TGT GCA ATT CAG GAT CTG CTG CCG GAA ACC CGT CCG GCA CGT GCA GTT TTT CTG GCA GAA
      >..............................synthetic_gbnB..............................>
        c   a   i   q   d   l   l   p   e   t   r   p   a   r   a   v   f   l   a   e

 181  AAT TCA TGG GAA CTG GTT GTT AGC ATG GCA GCA GTT AGC AGC CTG GGT ATT CCG TGT GTT
      >..............................synthetic_gbnB..............................>
        n   s   w   e   l   v   v   s   m   a   a   v   s   s   l   g   i   p   c   v

 241  GGT CTG GAT TAT ACC GCA GGT CCG GTT GCA ACC CGT GCA GCA CTG GAT CAG CTG GAA CCG
      >..............................synthetic_gbnB..............................>
        g   l   d   y   t   a   g   p   v   a   t   r   a   a   l   d   q   l   e   p

 301  ACC GTT GTT ATT ACC AGC CGT GCA CAT CGT GCC CTG CTG GAA GAA TGT GGT TGG CCG AAT
      >..............................synthetic_gbnB..............................>
        t   v   v   i   t   s   r   a   h   r   a   l   l   e   e   c   g   w   p   n

 361  GGT CGT GAT ATT CTG GAT ATT TTT CTG GAT AGC GAA CCG GTT GGT CTG CCG GAT GGT CAG
      >..............................synthetic_gbnB..............................>
        g   r   d   i   l   d   i   f   l   d   s   e   p   v   g   l   p   d   g   q
```

```
 421  ACC GTT AGC GTT AGC TTT ACC GAT CTG CTG GTT AGC GCA CCG GGT GAT CCG CAG CCG GTT
      >.............................synthetic_gbnB.............................>
        t   v   s   v   s   f   t   d   l   l   v   s   a   p   g   d   p   q   p   v

 481  GAA CAG CCG TTT GAA GCA TTT AGC TTT ACC AGC GGC ACC AGC GGT GTT CCG AAA ATG GTT
      >.............................synthetic_gbnB.............................>
        e   q   p   f   e   a   f   s   f   t   s   g   t   s   g   v   p   k   m   v

 541  GTT CGT CGT ACC AGC TTT GAA GCA CGT CGT TTT GCA GAT CTG GTT GAT CAG TAT GCA TTT
      >.............................synthetic_gbnB.............................>
        v   r   r   t   s   f   e   a   r   r   f   a   d   l   v   d   q   y   a   f

 601  GAT GAA GAA GAT GTG CAT CTG GTT ACC GTT CCG CTG TAT CAT GCA AGC GGT CCG GGT TGG
      >.............................synthetic_gbnB.............................>
        d   e   e   d   v   h   l   v   t   v   p   l   y   h   a   s   g   p   g   w

 661  GCA CGT ATT TTT CTG ACC CTG GGT GGC ACC GTT GTT CTG GGT CCG CAT GAT GAT GCC GGT
      >.............................synthetic_gbnB.............................>
        a   r   i   f   l   t   l   g   g   t   v   v   l   g   p   h   d   d   a   g

 721  GAA CTG ATT CGT ATT ATT CAG GAT GAA GGT GTT AGC ACC ACC CTG ATG GTT CCG CCT GTT
      >.............................synthetic_gbnB.............................>
        e   l   i   r   i   i   q   d   e   g   v   s   t   t   l   m   v   p   p   v

 781  CTG GCA CGT GTT GTT GCA CAT CCG GAT AGC GAA CAT CTG CAT AAA ACC AGC CGT CTG CAT
      >.............................synthetic_gbnB.............................>
        l   a   r   v   v   a   h   p   d   s   e   h   l   h   k   t   s   r   l   h

 841  TTT GTT CTG AGC GGT GGT CGT CAT CTG AAT CGT TGG GTT ATT AAC AAT GCA TGG GAT CGT
      >.............................synthetic_gbnB.............................>
        f   v   l   s   g   g   r   h   l   n   r   w   v   i   n   n   a   w   d   r

 901  CTG GGT CCG GTT CTG CAT CTG TAT TAT GGC ACC ACC GAA ACC GGT GTT AAT GTT ATG ATT
      >.............................synthetic_gbnB.............................>
        l   g   p   v   l   h   l   y   y   g   t   t   e   t   g   v   n   v   m   i

 961  GGT CCG GAA GAA CTG CAT GTT GCA CCG TGT CGT AGC GGT CGT CCG ATG CCT GGT AAT ACC
      >.............................synthetic_gbnB.............................>
        g   p   e   e   l   h   v   a   p   c   r   s   g   r   p   m   p   g   n   t

1021  GTT GTT GTT CTG GAT GCA GAA AAT CGT CCG CTG CCG AAA GGT AGC CGT GGT CGT GTT GCA
      >.............................synthetic_gbnB.............................>
        v   v   v   l   d   a   e   n   r   p   l   p   k   g   s   r   g   r   v   a

1081  ATT GCC AGC TAT CAG CTG ATG GAT AGC TAT GCA ACC AGC GAA CCG GAT TTT ATG ACC CTG
      >.............................synthetic_gbnB.............................>
        i   a   s   y   q   l   m   d   s   y   a   t   s   e   p   d   f   m   t   l

1141  GAT TAT GGT GGT CGT ACC CAG CGT TTT CTG CTG ACC GGT GAT AGC GGT CTG GTT GAT GAA
      >.............................synthetic_gbnB.............................>
        d   y   g   g   r   t   q   r   f   l   l   t   g   d   s   g   l   v   d   e

1201  GCA GGT CGT CTG GAA CTG ACC GGT CGT AAT GAT GGT GTT GCA AAA GCA GAA GCA GGT AAA
      >.............................synthetic_gbnB.............................>
        a   g   r   l   e   l   t   g   r   n   d   g   v   a   k   a   e   a   g   k

1261  CCG CTG GAT GTT AAC ATT TTT GGT CTG GAA GCA GAT CTG ATG GAT CTG CCG TGT GTT CGT
      >.............................synthetic_gbnB.............................>
        p   l   d   v   n   i   f   g   l   e   a   d   l   m   d   l   p   c   v   r

1321  GAA ACC GCA GTT CTG CGT ACC AGC CTG CCT GGT GCC GGT GAA GTT CTG GTT GTT CCG TTT
      >.............................synthetic_gbnB.............................>
        e   t   a   v   l   r   t   s   l   p   g   a   g   e   v   l   v   v   p   f

1381  ACA CCG GTT GCA CCG GAA CGT CAA GAA AGC GGT TAT CGT GCA GTT AGC GCA GCA TGT GCA
      >.............................synthetic_gbnB.............................>
        t   p   v   a   p   e   r   q   e   s   g   y   r   a   v   s   a   a   c   a

1441  CGT CGT GTT CCG TGT CTG CCT GCA TAT GTT GTT CCG GTT GAT GCA ATT CCG TAT AGC CCG
      >.............................synthetic_gbnB.............................>
        r   r   v   p   c   l   p   a   y   v   v   p   v   d   a   i   p   y   s   p

1501  ACC GGT AAA GTT CGT GCA GCA CAG CTG CTG GAA AGC GTT CTG CCG CTG CTG GAT ATT CGT
      >.............................synthetic_gbnB.............................>
        t   g   k   v   r   a   a   q   l   l   e   s   v   l   p   l   l   d   i   r

1561  CCG GCA GAT GAA CGT GAA GCA GAA ACC GTT GGT GTT TAA
      >................synthetic_gbnB................>>
        p   a   d   e   r   e   a   e   t   v   g   v   -
```

Synthetic *sven_4181* (*S. venezuelae gbnC*) native sequence is NCBI Gene ID: 13819339

```
   1  ATG AAT AGC CAG AAT ACC TAT ACC CGT GGT GTT CTG CTG TGT CTG CTG GCA ACC GTT AGC
      >>.............................synthetic_gbnC.............................>
        m   n   s   q   n   t   y   t   r   g   v   l   l   c   l   l   a   t   v   s

  61  TGG GGT GCA ATG TTT CCG CTG ATG GAT AGC ACC CTG CAG CAT ATT GAT CCG TTT ACC TTT
      >.............................synthetic_gbnC.............................>
        w   g   a   m   f   p   l   m   d   s   t   l   q   h   i   d   p   f   t   f

 121  ACC GTT ATG CGT TAT ACC ATT GCC GGT GCC ATG TTT CTG GTT TTT CTG CGT ATG CGT GAA
      >.............................synthetic_gbnC.............................>
        t   v   m   r   y   t   i   a   g   a   m   f   l   v   f   l   r   m   r   e

 181  GGT CGC GAA GGT CTG CGT CTG AAA GGT GAA CGT ATT GGT CTG GCA TGG CTG TTT GGT ACA
      >.............................synthetic_gbnC.............................>
        g   r   e   g   l   r   l   k   g   e   r   i   g   l   a   w   l   f   g   t

 241  GCA GGT TTT GCC GGT TTT CAG TTT CTG GTG TTT TTT GGT CAG GAT CTG ATT GGT GCA CGT
      >.............................synthetic_gbnC.............................>
        a   g   f   a   g   f   q   f   l   v   f   f   g   q   d   l   i   g   a   r

 301  GGT GCA CTG AAT GCA AGC ATT ATG ATG GCA ACC ATG CCG ATG ATG GGT TTT CTG GTT AAT
      >.............................synthetic_gbnC.............................>
        g   a   l   n   a   s   i   m   m   a   t   m   p   m   m   g   f   l   v   n

 361  TGG GTG ATG AAA AAA GTT GTG CCT CCG AAA TTT AGC CTG GTG TTT ATT GCA ATG AGC TTT
      >.............................synthetic_gbnC.............................>
        w   v   m   k   k   v   v   p   p   k   f   s   l   v   f   i   a   m   s   f

 421  GTT GGC ACC ATT CTG GTT GTT AGC AAT GGT GAT ATT GGT AGC CTG ATT GCA AGC CCG AAA
      >.............................synthetic_gbnC.............................>
        v   g   t   i   l   v   v   s   n   g   d   i   g   s   l   i   a   s   p   k

 481  GAA GCA GGC GCA GAC GCA CTG CTG CTG TTT GCA GCA CTG TGT TGG GTT GTT TAT ACC AGC
      >.............................synthetic_gbnC.............................>
        e   a   g   a   d   a   l   l   l   f   a   a   l   c   w   v   v   y   t   s

 541  GGT GCA AGC TAT TTT CCG ACC TGG TCA CCG ATC AAA TAT ACC GCA ATT ACC ACC GTT CTG
      >.............................synthetic_gbnC.............................>
        g   a   s   y   f   p   t   w   s   p   i   k   y   t   a   i   t   t   v   l

 601  GGT CTG GCA AGC GCA ACC GTT ATT ACC GCA GTT ATT CTG GCA GCC GGT GGT GTT CCG GTT
      >.............................synthetic_gbnC.............................>
        g   l   a   s   a   t   v   i   t   a   v   i   l   a   a   g   g   v   p   v

 661  CCG ACC GCT GGT GCA GTT GGT ACA ATT CTG CCG CAG CTG GCA TAT ATG AGC GTT ATT GCA
      >.............................synthetic_gbnC.............................>
        p   t   a   g   a   v   g   t   i   l   p   q   l   a   y   m   s   v   i   a

 721  GGT TTT GTT GGT GTT CTG GGT TGG AAT TTT GGT AAT CGT TAT CTG GGT CCG CTG AAT GGT
      >.............................synthetic_gbnC.............................>
        g   f   v   g   v   l   g   w   n   f   g   n   r   y   l   g   p   l   n   g

 781  GTT CTG TTT ATG GAT GTT GTT CCG GTT ACC GCA TTT GTT ATT AGC GCA CTG ACC GGT GTT
      >.............................synthetic_gbnC.............................>
        v   l   f   m   d   v   v   p   v   t   a   f   v   i   s   a   l   t   g   v

 841  GTT CCG GCA GGC GTT CAG ATT GTT GGT GCA AGC CTG ACC GCA GCA GCA CTG GTT TTT AAC
      >.............................synthetic_gbnC.............................>
        v   p   a   g   v   q   i   v   g   a   s   l   t   a   a   a   l   v   f   n

 901  AAT CTG TAT CTG CGT CGT ATT GCA AAA GCA GCA CCG GCA GCA AGT CCG GCA CCG GCA GCT
      >.............................synthetic_gbnC.............................>
        n   l   y   l   r   r   i   a   k   a   a   p   a   a   s   p   a   p   a   a

 961  AGC GGT CCG AGC GTT GGT AGC AGC AGC GGT ACA AGC AGC GGC ACC GCA AGC GGT AGC GGT
      >.............................synthetic_gbnC.............................>
        s   g   p   s   v   g   s   s   s   g   t   s   s   g   t   a   s   g   s   g

1021  AGC GCA AGT GCA CCG GCA CCG GTT GGT AGC GGC ACC CGT AGC TAA
      >.....................synthetic_gbnC.....................>>
        s   a   s   a   p   a   p   v   g   s   g   t   r   s   -
```

Synthetic *seag_b2133* (*S. enterica gbnB*) native sequence is NCBI Gene ID: 6795111

```
   1  ATG ATT ACC CTG CGT CGT CTG AAT GAA ATT GCA ATT ACC CGT GGT AAC GAT ATC TGC ATC
      >>.................................seag_gbnB.................................>
        m   i   t   l   r   r   l   n   e   i   a   i   t   r   g   n   d   i   c   i

  61  ATT GAT AAA GAA CGT CAG TAT ACC TGG TAC GAT ATT ATT CGT CGT ACC GAA AGC CGT ATT
      >.................................seag_gbnB.................................>
        i   d   k   e   r   q   y   t   w   y   d   i   i   r   r   t   e   s   r   i

 121  GTT TTT CTG CGT CGT GCA TTT AAT CCG GAA CAG CTG CGT AGC GTT TGT TAT CTG AGC AAA
      >.................................seag_gbnB.................................>
        v   f   l   r   r   a   f   n   p   e   q   l   r   s   v   c   y   l   s   k

 181  AAT AGC GTT GAT CTG ATT TGT TGG CTG GCA GCA TTT GCA ACC CTG GGT ATT CCG GCA AAT
      >.................................seag_gbnB.................................>
        n   s   v   d   l   i   c   w   l   a   a   f   a   t   l   g   i   p   a   n

 241  GGT CTG GAT TAT AGC CTG CCG ATT GAA ACC CTG CGT GGT CTG CTG ATT AAA ATC AAT CCG
      >.................................seag_gbnB.................................>
        g   l   d   y   s   l   p   i   e   t   l   r   g   l   l   i   k   i   n   p

 301  GGT CTG ATG CTG GTT AGC TTT AGC CTG TAT AGT CCG GAT GAA CTG AAT AAA CTG CAT GTT
      >.................................seag_gbnB.................................>
        g   l   m   l   v   s   f   s   l   y   s   p   d   e   l   n   k   l   h   v

 361  CGT ACC ATT ACC ATG CTG GCA GTT GAT GCA CCG ACC GAT CCG GTT ATT GGT AGC ATT GGT
      >.................................seag_gbnB.................................>
        r   t   i   t   m   l   a   v   d   a   p   t   d   p   v   i   g   s   i   g

 421  GAA TTT CAT CAT CCG GAA CTG GAA AGC CTG CTG GCA ACC CAT ATT CCG GCA CCG TTT CGT
      >.................................seag_gbnB.................................>
        e   f   h   h   p   e   l   e   s   l   l   a   t   h   i   p   a   p   f   r

 481  AGC GTT AGC CTG ACC AGC GGC ACC AGC AGC GCA CCG AAA ATT GTT CTG CGT TAT AAT AGC
      >.................................seag_gbnB.................................>
        s   v   s   l   t   s   g   t   s   s   a   p   k   i   v   l   r   y   n   s

 541  TTT GAT GCC CGT CGT TTT GAT TGG TTT ACC CAG CGT TTT AAC TTT ACG CAT CAT GAT GGT
      >.................................seag_gbnB.................................>
        f   d   a   r   r   f   d   w   f   t   q   r   f   n   f   t   h   h   d   g

 601  TTT CTG CTG ATT CTG CCG CTG TAT CAT GCA GCA GGT AAT GGT TGG GCA CGT ATG TTT ATG
      >.................................seag_gbnB.................................>
        f   l   l   i   l   p   l   y   h   a   a   g   n   g   w   a   r   m   f   m

 661  GGT CTG GGT GCA AGC CTG CAT CTG GTT GAT CAG GAT GAT GAA AGC GCA CTG ATT CAG GCA
      >.................................seag_gbnB.................................>
        g   l   g   a   s   l   h   l   v   d   q   d   d   e   s   a   l   i   q   a

 721  CTG AGC CTG AAT AGC GTT AAA GCA ACC GTT ATG ACC CCG AAT CTG GTT AGC CGT CTG ACC
      >.................................seag_gbnB.................................>
        l   s   l   n   s   v   k   a   t   v   m   t   p   n   l   v   s   r   l   t

 781  AAA CTG GCA AGC GAA ACC GTT CTG CAT CAT TAT CTG CGT TGG GTT CTG GTT GGT GGT AGC
      >.................................seag_gbnB.................................>
        k   l   a   s   e   t   v   l   h   h   y   l   r   w   v   l   v   g   g   s

 841  TAT TTT CCG GTT AAA AGC AAA ATT GCA GCC TAT ACC CAT CTG GGC CAT ATC TTT AAT GAA
      >.................................seag_gbnB.................................>
        y   f   p   v   k   s   k   i   a   a   y   t   h   l   g   h   i   f   n   e

 901  TAC TAT GGT TGT ACC GAA ACC GGT GTT AAT GTT CTG AGC GAA AGC AGC GAT ATG TTT GAA
      >.................................seag_gbnB.................................>
        y   y   g   c   t   e   t   g   v   n   v   l   s   e   s   s   d   m   f   e

 961  TGT CCG GGT AGC GTT GGT CGT GCA TTT GAT GGT AAC AAA ATT CGC ATT CTG GAT GAA GAT
      >.................................seag_gbnB.................................>
        c   p   g   s   v   g   r   a   f   d   g   n   k   i   r   i   l   d   e   d

1021  AAT GTT CCG CTG AAA GCA GGT AAT CGT GGT CGT ATT GCA GTT GCA AGC TAT ATG CTG ATG
      >.................................seag_gbnB.................................>
        n   v   p   l   k   a   g   n   r   g   r   i   a   v   a   s   y   m   l   m

1081  GAT GAA TAT GGT GAT GGT AGC CGT CCG TTT ATT GAA ATT GAT AGC GAA CGC TAT TTT CTG
      >.................................seag_gbnB.................................>
        d   e   y   g   d   g   s   r   p   f   i   e   i   d   s   e   r   y   f   l
```

```
1141  ATG GCC GAT TAT GGT TAT CTG GAT GAT AAT GGT CGT CTG TTT CTG ATG AAT CGC AAC AGC
      >.................................seag_gbnB.................................>
        m   a   d   y   g   y   l   d   d   n   g   r   l   f   l   m   n   r   n   s

1201  GAA ATC AAA TGC GAA CAG GAT ATC TAT CAC ATC GAA GAA CAT CTG CGT GCA CTG CCG TGT
      >.................................seag_gbnB.................................>
        e   i   k   c   e   q   d   i   y   h   i   e   e   h   l   r   a   l   p   c

1261  ATT ACC GAT GTT GCA CTG ATT CCG ATT CGT CAG CAG AAC AAA GAT CAT ATT CGT TGC ATT
      >.................................seag_gbnB.................................>
        i   t   d   v   a   l   i   p   i   r   q   q   n   k   d   h   i   r   c   i

1321  TTC AGC GCC AAA TAC ATC AAC GAA GAT GAT GTG AGC TTT ATT ATG GAT GAA ATC AAA AAC
      >.................................seag_gbnB.................................>
        f   s   a   k   y   i   n   e   d   d   v   s   f   i   m   d   e   i   k   n

1381  AAA ATT AAC CAT ATT GGC GTG ACC GAT TTT ACC GCA CAT ATG GTT GAC AAA ATT CCG TAT
      >.................................seag_gbnB.................................>
        k   i   n   h   i   g   v   t   d   f   t   a   h   m   v   d   k   i   p   y

1441  AGC CCG AGC GGT AAA GTT CGT TTT AGC GAA ATT GTT CAG ACC CTG ACC GCA GCA TAA
      >.................................seag_gbnB.................................>>
        s   p   s   g   k   v   r   f   s   e   i   v   q   t   l   t   a   a   -
```

Synthetic *smugs5_01685* (*S. mutans* GS-5 *gbnB*) native sequence is NCBI Gene ID: 13299420

```
   1  ATG CTG GCA CTG GAA AAT CTG ATT CAG ATT CGT AAT CGC AAT CCG GAT AAA CTG ATT CTG
      >>................................smug_gbnB.................................>
        m   l   a   l   e   n   l   i   q   i   r   n   r   n   p   d   k   l   i   l

  61  ATT AGC GAC GAA AAA AGC TTC AGC TGG AAA GAA TAT ACC AAC CTG GTG ATT AAT AAC CTG
      >................................smug_gbnB.................................>
        i   s   d   e   k   s   f   s   w   k   e   y   t   n   l   v   i   n   n   l

 121  CGT AAT ACC ACC CTG CAG AGC GTT CTG AAT AAA ACC GAT CGT GCA ATT ATC ATC AGC GAG
      >................................smug_gbnB.................................>
        r   n   t   t   l   q   s   v   l   n   k   t   d   r   a   i   i   i   s   e

 181  AAT ACC TGG AAA GTG TTC ACC ATT TAT AGC TGT CTG AGC ACC CAG AAA ATT CCG TAT AGC
      >................................smug_gbnB.................................>
        n   t   w   k   v   f   t   i   y   s   c   l   s   t   q   k   i   p   y   s

 241  GGT ATT GAT TAC AGC ATG GAA GAT GAT AAA AAA GTG GCA GCC ATT AAC AAA AGC GGT GCA
      >................................smug_gbnB.................................>
        g   i   d   y   s   m   e   d   d   k   k   v   a   a   i   n   k   s   g   a

 301  AAT ACC GTG TTC TAT AGC AAA GAT CAG AAA CCG AGC CAG AAT CTG CGT AAT AGC CTG AAA
      >................................smug_gbnB.................................>
        n   t   v   f   y   s   k   d   q   k   p   s   q   n   l   r   n   s   l   k

 361  GGT GTT AGC TTT ATT AGC CTG GAT ATC CTG CAT GAT GAT ATT GAA GGT AGC GAT CTG AGC
      >................................smug_gbnB.................................>
        g   v   s   f   i   s   l   d   i   l   h   d   d   i   e   g   s   d   l   s

 421  GAC TTT AAT ATC AAA AAA CAC AGC GAT AGC ATC GTG AGC TTT GGT TTT ACC AGC GGC ACC
      >................................smug_gbnB.................................>
        d   f   n   i   k   k   h   s   d   s   i   v   s   f   g   f   t   s   g   t

 481  ACC GGT CTG CCG AAA TGT ATT TAT CGT GAT TAT AGC TTT GCC ACC GAA CGC ATG AAA GAA
      >................................smug_gbnB.................................>
        t   g   l   p   k   c   i   y   r   d   y   s   f   a   t   e   r   m   k   e

 541  CTG ACC AAA CTG TAT AAC TTT AAC GCG ACC GAT GTT TTT CTG GTT ACC ATG CCG TTT TAT
      >................................smug_gbnB.................................>
        l   t   k   l   y   n   f   n   a   t   d   v   f   l   v   t   m   p   f   y

 601  CAT GTT AGC GTT AAT GGT TGG GTT AAA CTG ACC CTG AAT AAT GGT GGT AGC GTT GTT CTG
      >................................smug_gbnB.................................>
        h   v   s   v   n   g   w   v   k   l   t   l   n   n   g   g   s   v   v   l
```

```
 661  GGT GAT TTT AAC AAT CCG ATT GAT CTG AGC AGC AAA ATC AAA CAG TAT GAC ATT ACC ACC
      >.................................smug_gbnB.................................>
        g   d   f   n   n   p   i   d   l   s   s   k   i   k   q   y   d   i   t   t

 721  ATG CTG ATT ACC CCT CCG GTT CTG AAA AGC CTG AAT TTT GTT CTG AAT CAG CAG GGC TTT
      >.................................smug_gbnB.................................>
        m   l   i   t   p   p   v   l   k   s   l   n   f   v   l   n   q   q   g   f

 781  ATC AAT AGC ACC GTT CGC TTT ATT ATG GTG GGT GGT AAA AAC TTT CCG CCT AAA CTG AAA
      >.................................smug_gbnB.................................>
        i   n   s   t   v   r   f   i   m   v   g   g   k   n   f   p   p   k   l   k

 841  GAA GAA ACC CAG AAC CTG TTT GGT AGC GTT CTG CAT GAA TAT TAT GGT AGC AGC GAA ACC
      >.................................smug_gbnB.................................>
        e   e   t   q   n   l   f   g   s   v   l   h   e   y   y   g   s   s   e   t

 901  GGT ATT AAT GTT CTG GCA AAT AGC AGC GAT ATG ATG CTG TAT CCG AGC AGC AGC GGT CGT
      >.................................smug_gbnB.................................>
        g   i   n   v   l   a   n   s   s   d   m   m   l   y   p   s   s   s   g   r

 961  GTT ATG AAA GGT AGT GAT GTT ATT ATC GTG GAT AGC GAC AAT CGC AAA ATT CCG AAT AAT
      >.................................smug_gbnB.................................>
        v   m   k   g   s   d   v   i   i   v   d   s   d   n   r   k   i   p   n   n

1021  CAT ATT GGT CGG ATT GCC ATC TAT AGC TAT CAG AAT GCA ACC GGC TAT ATT AAC CAG CCG
      >.................................smug_gbnB.................................>
        h   i   g   r   i   a   i   y   s   y   q   n   a   t   g   y   i   n   q   p

1081  CTG GAA AAA TTC AAT TAC CGC CAG AAA GAA TAC ATC CTG ACC AGC GAT TAT GGC TAT GTG
      >.................................smug_gbnB.................................>
        l   e   k   f   n   y   r   q   k   e   y   i   l   t   s   d   y   g   y   v

1141  AAT AAT GAA GGC TAT ATC TTT GTG GTG CAG CGC ATT CTG AAT CAC GAA AAC AAC AAA ATC
      >.................................smug_gbnB.................................>
        n   n   e   g   y   i   f   v   v   q   r   i   l   n   h   e   n   n   k   i

1201  ATC AAT GTG TTT CAG ATC GAA AAT CGC CTG CGC CTG ATT AAA GAT ATT GAT GAT GTT GCC
      >.................................smug_gbnB.................................>
        i   n   v   f   q   i   e   n   r   l   r   l   i   k   d   i   d   d   v   a

1261  ATC GTG CAG AAA AAT AAC CTG CTG CTG GTT AAC ATT AAA CTG AAA AAA ATC AGC GAA ATG
      >.................................smug_gbnB.................................>
        i   v   q   k   n   n   l   l   l   v   n   i   k   l   k   k   i   s   e   m

1321  AAA CGC AGC CTG GTT AAT GAT CTG GTT TGT TGG ATT CTG GAA AAA ACC AAA ATC CCG TAC
      >.................................smug_gbnB.................................>
        k   r   s   l   v   n   d   l   v   c   w   i   l   e   k   t   k   i   p   y

1381  GAT CTG AAA TAT ACC GAT GAG ATC CAT TAT TCC ATG AGC GGC AAA GTG AAA TAC ACC GAA
      >.................................smug_gbnB.................................>
        d   l   k   y   t   d   e   i   h   y   s   m   s   g   k   v   k   y   t   e

1441  GTG ATT AAT AGC GAA GGT CGT TAA
      >.........smug_gbnB.........>>
        v   i   n   s   e   g   r   -
```

Appendix 4 - Maps of pET-Duet and pACYC-Duet multiple cloning sites

pET-Duet:

His-tag is underlined; S-tag highlighted in blue

Primer binding sites for 'pET up', 'UP2', 'DOWN1' and 'T7_term' highlighted in red

```
   1  GTGATGTCGG CGATATAGGC GCCAGCAACC GCACCTGTGG CGCCGGTGAT GCCGGCCACG
      CACTACAGCC GCTATATCCG CGGTCGTTGG CGTGGACACC GCGGCCACTA CGGCCGGTGC

  61  ATGCGTCCGG CGTAGAGGAT CGAGATCGAT CTCGATCCCG CGAAATTAAT ACGACTCACT
      TACGCAGGCC GCATCTCCTA GCTCTAGCTA GAGCTAGGGC GCTTTAATTA TGCTGAGTGA

 121  ATAGGGGAAT TGTGAGCGGA TAACAATTCC CCTCTAGAAA TAATTTTGTT TAACTTTAAG
      TATCCCCTTA ACACTCGCCT ATTGTTAAGG GGAGATCTTT ATTAAAACAA ATTGAAATTC

                                                          BamHI
                                                         -+-----
 181  AAGGAGATAT ACCATGGGCA GCAGCCATCA CCATCATCAC CACAGCCAGG ATCCGAATTC
      TTCCTCTATA TGGTACCCGT CGTCGGTAGT GGTAGTAGTG GTGTCGGTCC TAGGCTTAAG

                              HindIII
                              -+-----
 241  GAGCTCGGCG CGCCTGCAGG TCGACAAGCT TGCGGCCGCA TAATGCTTAA GTCGAACAGA
      CTCGAGCCGC GCGGACGTCC AGCTGTTCGA ACGCCGGCGT ATTACGAATT CAGCTTGTCT

 301  AAGTAATCGT ATTGTACACG GCCGCATAAT CGAAATTAAT ACGACTCACT ATAGGGGAAT
      TTCATTAGCA TAACATGTGC CGGCGTATTA GCTTTAATTA TGCTGAGTGA TATCCCCTTA

 361  TGTGAGCGGA TAACAATTCC CCATCTTAGT ATATTAGTTA AGTATAAGAA GGAGATATAC
      ACACTCGCCT ATTGTTAAGG GGTAGAATCA TATAATCAAT TCATATTCTT CCTCTATATG

                                  FseI                     KpnI
                                 -------+-                 -----+
 421  ATATGGCAGA TCTCAATTGG ATATCGGCCG GCCACGCGAT CGCTGACGTC GGTACCCTCG
      TATACCGTCT AGAGTTAACC TATAGCCGGC CGGTGCGCTA GCGACTGCAG CCATGGGAGC

 481  AGTCTGGTAA AGAAACCGCT GCTGCGAAAT TTGAACGCCA GCACATGGAC TCGTCTACTA
      TCAGACCATT TCTTTGGCGA CGACGCTTTA AACTTGCGGT CGTGTACCTG AGCAGATGAT

 541  GCGCAGCTTA ATTAACCTAG GCTGCTGCCA CCGCTGAGCA ATAACTAGCA
      CGCGTCGAAT TAATTGGATC CGACGACGGT GGCGACTCGT TATTGATCGT
```

pACYC-Duet：

His-tag is underlined; S-tag highlighted in blue

Primer binding sites for ‘pACYCup’, ‘UP2’, ‘DOWN1’ and ‘T7_term’ in red

```
   1                          GCCATAC CGCGAAAGGT TTTGCGCCAT TCGATGGTGT
                              CGGTATG GCGCTTTCCA AAACGCGGTA AGCTACCACA

  61  CCGGGATCTC GACGCTCTCC CTTATGCGAC TCCTGCATTA GGAAATTAAT ACGACTCACT
      GGCCCTAGAG CTGCGAGAGG GAATACGCTG AGGACGTAAT ACTTTAATTA TGCTGAGTGA

 121  ATAGGGGAAT TGTGAGCGGA TAACAATTCC CCTCTAGAAA TAATTTTGTT TAACTTTAAG
      TATCCCCTTA ACACTCGCCT ATTGTTAAGG GGAGATCTTT ATTAAAACAA ATTGAAATTC

                                                          BamHI
                                                         -+-----
 181  AAGGAGATAT ACCATGGGCA GCAGCCATCA CCATCATCAC CACAGCCAGG ATCCGAATTC
      TTCCTCTATA TGGTACCCGT CGTCGGTAGT GGTAGTAGTG GTGTCGGTCC TAGGCTTAAG

                              HindIII
                              -+-----
 241  GAGCTCGGCG CGCCTGCAGG TCGACAAGCT TGCGGCCGCA TAATGCTTAA GTCGAACAGA
      CTCGAGCCGC GCGGACGTCC AGCTGTTCGA ACGCCGGCGT ATTACGAATT CAGCTTGTCT

 301  AAGTAATCGT ATTGTACACG GCCGCATAAT CGAAATTAAT ACGACTCACT ATAGGGGAAT
      TTCATTAGCA TAACATGTGC CGGCGTATTA GCTTTAATTA TGCTGAGTGA TATCCCCTTA

 361  TGTGAGCGGA TAACAATTCC CCATCTTAGT ATATTAGTTA AGTATAAGAA GGAGATATAC
      ACACTCGCCT ATTGTTAAGG GGTAGAATCA TATAATCAAT TCATATTCTT CCTCTATATG

                                  FseI                     KpnI
                              -------+-                   -----+
 421  ATATGGCAGA TCTCAATTGG ATATCGGCCG GCCACGCGAT CGCTGACGTC GGTACCCTCG
      TATACCGTCT AGAGTTAACC TATAGCCGGC CGGTGCGCTA GCGACTGCAG CCATGGGAGC

 481  AGTCTGGTAA AGAAACCGCT GCTGCGAAAT TTGAACGCCA GCACATGGAC TCGTCTACTA
      TCAGACCATT TCTTTGGCGA CGACGCTTTA AACTTGCGGT CGTGTACCTG AGCAGATGAT

 541  GCGCAGCTTA ATTAACCTAG GCTGCTGCCA CCGCTGAGCA ATAACTAGCA
      CGCGTCGAAT TAATTGGATC CGACGACGGT GGCGACTCGT TATTGATCGT
```

Appendix 5 - UHR-MS of GBLs identified in *S. coelicolor* M1152 extract

UHR-MS of SCB 1-8 identified in culture extract of *S. coelicolor* M1152 grown on AlaMM media

compound	*Molecular formula*	*Observed* m/z	*Calculated* m/z	*Assignment*
SCB8 (**193**)	$C_{11}H_{20}NaO_4$	239.1258	239.1254	$[M+Na]^+$
	$C_{11}H_{21}O_4$	217.1444	217.1434	$[M+H]^+$
	$C_{11}H_{19}O_3$	199.1335	199.1329	$[M-H_2O+H]^+$
	$C_{11}H_{17}O_2$	181.1227	181.1223	$[M-2H_2O+H]^+$
SCB4 (**194**)	$C_{12}H_{22}NaO_4$	253.1421	253.1410	$[M+Na]^+$
	$C_{12}H_{23}O_4$	231.1594	231.1591	$[M+H]^+$
	$C_{12}H_{21}O_3$	213.1487	213.1485	$[M-H_2O+H]^+$
	$C_{12}H_{19}O_2$	195.1377	195.1380	$[M-2H_2O+H]^+$
SCB5 (**195**)	$C_{12}H_{22}NaO_4$	253.1421	253.1410	$[M+Na]^+$
	$C_{12}H_{23}O_4$	231.1600	231.1591	$[M+H]^+$
	$C_{12}H_{21}O_3$	213.1495	213.1485	$[M-H_2O+H]^+$
	$C_{12}H_{19}O_2$	195.1385	195.1380	$[M-2H_2O+H]^+$
SCB6 (**196**)	$C_{12}H_{22}NaO_4$	253.1418	253.1410	$[M+Na]^+$
	$C_{12}H_{23}O_4$	231.1593	231.1591	$[M+H]^+$
	$C_{12}H_{21}O_3$	213.1496	213.1485	$[M-H_2O+H]^+$
	$C_{12}H_{19}O_2$	195.1388	195.1380	$[M-2H_2O+H]^+$
SCB1 (**197**)	$C_{13}H_{24}NaO_4$	267.1571	267.1567	$[M+Na]^+$
	$C_{13}H_{25}O_4$	245.1753	245.1747	$[M+H]^+$
	$C_{13}H_{23}O_3$	227.1644	227.1642	$[M-H_2O+H]^+$
	$C_{13}H_{21}O_2$	209.1541	209.1536	$[M-2H_2O+H]^+$
SCB2 (**198**)	$C_{13}H_{24}NaO_4$	267.1579	267.1567	$[M+Na]^+$
	$C_{13}H_{25}O_4$	245.1749	245.1747	$[M+H]^+$
	$C_{13}H_{23}O_3$	227.1643	227.1642	$[M-H_2O+H]^+$
	$C_{13}H_{21}O_2$	209.1540	209.1536	$[M-2H_2O+H]^+$
SCB3 (**199**)	$C_{14}H_{26}NaO_4$	281.1729	281.1723	$[M+Na]^+$
	$C_{14}H_{27}O_4$	259.1912	259.1904	$[M+H]^+$
	$C_{14}H_{25}O_3$	241.1801	241.1798	$[M-H_2O+H]^+$
	$C_{14}H_{23}O_2$	223.1699	223.1693	$[M-2H_2O+H]^+$
SCB7 (**200**)	$C_{14}H_{26}NaO_4$	281.1731	281.1723	$[M+Na]^+$
	$C_{14}H_{27}O_4$	259.1919	259.1904	$[M+H]^+$
	$C_{14}H_{25}O_3$	241.1800	241.1798	$[M-H_2O+H]^+$
	$C_{14}H_{23}O_2$	223.1696	223.1693	$[M-2H_2O+H]^+$

Appendix 6 - UHR-MS of AHFCAs identified in this study

UHR-MS of AHFCAs 5-8 identified in culture extract of *S. venezuelae gbnR* mutants grown on SMMS and AHFCA8E identified in the *E. coli*/pJDS8 culture extract

compound	*Molecular formula*	*Observed* m/z	*Calculated* m/z	*Assignment*
AHFCA5 (**201**)	$C_{11}H_{16}NaO_4$	235.0936	235.0941	$[M+Na]^+$
	$C_{11}H_{17}O_4$	213.1112	213.1121	$[M+H]^+$
	$C_{11}H_{15}O_3$	195.1006	195.1016	$[M-H_2O+H]^+$
AHFCA6 (**202**)	$C_{12}H_{18}NaO_4$	249.1089	249.1087	$[M+Na]^+$
	$C_{12}H_{19}O_4$	227.1251	227.1278	$[M+H]^+$
	$C_{12}H_{17}O_3$	209.1174	209.1172	$[M-H_2O+H]^+$
AHFCA7 (**203**)	$C_{12}H_{18}NaO_4$	249.1089	249.1087	$[M+Na]^+$
	$C_{12}H_{19}O_4$	227.1261	227.1278	$[M+H]^+$
	$C_{12}H_{17}O_3$	209.1158	209.1172	$[M-H_2O+H]^+$
AHFCA8 (**204**)	$C_{13}H_{20}NaO_4$	263.1242	263.1254	$[M+Na]^+$
	$C_{13}H_{21}O_4$	241.1420	241.1434	$[M+H]^+$
	$C_{13}H_{19}O_3$	223.1320	223.1329	$[M-H_2O+H]^+$
AHFCA8E (**214**)	$C_{13}H_{24}NaO_4$	263.1245	263.1254	$[M+Na]^+$
	$C_{13}H_{25}O_4$	241.1419	241.1434	$[M+H]^+$
	$C_{13}H_{23}O_3$	223.1321	223.1329	$[M-H_2O+H]^+$