

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/97197>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

**Elucidating molecular mechanisms of
actinobacterial polyketide alkaloid
biosynthesis**

by

Jade L. Ronan



Thesis submitted in partial fulfilment of the requirements for the
degree of
Doctor of Philosophy in Chemistry

**University of Warwick
Department of Chemistry**

July 2017

Contents

Contents	i
List of figures	viii
List of schemes.....	xvi
List of tables.....	xxii
Acknowledgements.....	xxiii
Declaration.....	xxiv
Abbreviations	xxv
Abstract.....	xxviii
1 INTRODUCTION.....	1
1.1 <i>Streptomyces</i>	2
1.2 <i>Streptomyces coelicolor</i> A3(2)	4
1.3 Discovery of coelimycin P1	5
1.3.1 Type I modular PKSs	7
1.3.1.1 Minimal catalytic domains	9
1.3.1.2 Accessory catalytic domains	11
1.3.1.3 Polyketide chain offloading	13
1.3.1.4 Predicted module and domain organisation of the coelimycin PKS.....	15
1.3.2 Precursor supply genes.....	17
1.3.3 Post-PKS tailoring genes	18
1.3.3.1 ω -Transaminases	18
1.3.3.2 Flavin-dependent oxidases, dehydrogenases and monooxygenases	20
1.3.3.3 Short-chain dehydrogenases.....	24
1.3.3.4 Isomerases	25
1.4 Proposed biosynthesis of coelimycin P1	25
1.5 Regulation of coelimycin P1 biosynthesis	28
1.6 Other polyketide alkaloids	30
1.7 Aims of the project	31
2 RESULTS AND DISCUSSION I: Nitrogen incorporation in coelimycin and related polyketide alkaloids	32
2.1 Biochemical characterisation of CpkG	33
2.1.1 Introduction	33

2.1.2	Overproduction, purification and characterisation of CpkG.....	33
2.1.3	<i>In vitro</i> assays.....	35
2.1.3.1	Amino acid and aldehyde detection	35
2.1.3.2	α -Keto acid and amine detection.....	38
2.1.3.3	Utilisation of synthetic substrate mimics	39
2.2	Crystal structures of CpkG	41
2.2.1	Crystallisation screens.....	41
2.2.2	Overall structure of CpkG.....	41
2.2.3	Binding of PMP	43
2.2.4	Binding of PLP.....	44
2.3	Investigations into CpkG and CpkC-TR binding	46
2.3.1	Native-PAGE analysis	46
2.3.2	Pull-down assays.....	47
2.4	Further polyketide alkaloid specialised metabolites	49
2.4.1	Identification of polyketide alkaloid biosynthetic gene clusters.....	49
2.4.2	Prediction of polyketide intermediates.....	49
2.4.3	Prediction of polyketide alkaloid metabolites.....	53
2.4.4	Attempted detection of putative polyketide alkaloid from <i>Streptomyces</i> <i>cyaneofuscatus</i> NRRL B-2570.....	54
2.4.4.1	Overexpression of a putative SARP.....	55
2.4.4.2	Inactivation of the PKS	57
2.4.4.3	Comparative metabolic profiling	59
2.4.4.4	Overproduction, purification and characterisation of putative polyketide alkaloid metabolite(s).....	60
2.4.5	Whole genome sequencing of <i>Streptomyces abikoensis</i> , a latumcidin producer.....	62
2.5	Summary	65
3	RESULTS AND DISCUSSION II: Characterisation of post-PKS tailoring enzymes in coelimycin biosynthesis	67
3.1	CpkI and CpkE	68
3.1.1	Introduction.....	68
3.1.2	Cloning, overproduction and purification of proteins.....	68
3.1.3	Preliminary <i>in vitro</i> assays	71

3.2 CpkD, ScF and CpkH	73
3.2.1 Introduction	73
3.2.2 Cloning, overproduction and purification of proteins.....	74
3.2.3 Examination of flavin binding and NAD(P)H utilisation	75
3.2.3.1 UV/Vis spectroscopic analysis to identify the flavin cofactor.....	75
3.2.3.2 Mass spectrometric analysis to determine the mode of binding	76
3.2.3.3 Prediction of sites of covalent FAD attachment	77
3.2.3.4 UV/Vis spectroscopic analysis of flavin reduction by NAD(P)H	78
3.2.4 Probing flavin-dependent dehydrogenase activity	78
3.2.4.1 Incubation of 1-aminododecane-5-ol with CpkD, ScF and CpkH.....	78
3.2.4.2 Synthesis of (<i>R,E</i>)- and (<i>S,E</i>)-1-aminododec-2-en-5-ol.....	80
3.2.4.3 Determination of absolute stereochemistry and enantiomeric purity of (<i>R,E</i>)- and (<i>S,E</i>)-1-aminododec-2-en-5-ol	85
3.2.4.4 Incubation of (<i>R,E</i>)- and (<i>S,E</i>)-1-aminododec-2-en-5-ol with CpkH, CpkE and CpkI.....	87
3.2.4.5 Synthesis of (<i>R</i>)- and (<i>S</i>)-1-aminododecane-5-ol.....	89
3.2.4.6 Incubation of (<i>R</i>)- and (<i>S</i>)-1-aminododecane-5-ol with CpkH	90
3.2.5 Probing flavin-dependent monooxygenase activity	91
3.2.5.1 Synthesis of ethyl (<i>2E,4E,6E</i>)-octa-2,4,6-trienoate	91
3.2.5.2 Incubation of ethyl (<i>2E,4E,6E</i>)-octa-2,4,6-trienoate with CpkD and ScF	92
3.2.5.3 Tandem mass spectrometry and ¹⁸ O ₂ incorporation experiments.....	96
3.3 Summary.....	100
4 CONCLUSIONS AND PERSPECTIVES	102
4.1 Biochemical and structural characterisation of CpkG.....	103
4.2 Identification of further polyketide alkaloid biosynthetic gene clusters.....	103
4.3 <i>In vitro</i> investigations with CpkE and CpkI	104
4.4 <i>In vitro</i> investigations with CpkD, ScF and CpkH.....	105
4.5 Updated biosynthetic pathway.....	106
5 EXPERIMENTAL	108
5.1 Instruments and equipment	109
5.2 Materials	110
5.2.1 General	110

5.2.2	Kits and enzymes	111
5.2.3	Growth and production media.....	111
5.2.4	Buffers.....	113
5.2.5	Antibiotics.....	114
5.2.6	Bacterial strains.....	114
5.2.7	Vectors	115
5.2.8	PCR Primers.....	115
5.3	Biological procedures.....	116
5.3.1	Growth, preparation and storage of microbial cells.....	116
5.3.1.1	General culture conditions	116
5.3.1.2	Preparation and storage of <i>Streptomyces</i> spore stock solutions.....	116
5.3.1.3	Preparation and storage of chemically competent <i>E. coli</i> cells.....	117
5.3.1.4	Preparation and storage of electrocompetent <i>E. coli</i> cells	117
5.3.2	DNA isolation and purification.....	117
5.3.2.1	Isolation of genomic DNA from <i>Streptomyces</i>	117
5.3.2.2	Isolation of plasmid DNA from <i>E. coli</i>	118
5.3.3	DNA manipulation.....	118
5.3.3.1	Polymerase chain reaction (PCR)	118
5.3.3.2	Agarose gel electrophoresis and gel extraction.....	120
5.3.3.3	Restriction digests	120
5.3.3.4	Ligations.....	121
5.3.4	DNA transfer.....	121
5.3.4.1	Transformation of chemically competent <i>E. coli</i> cells.....	121
5.3.4.2	Transformation of electrocompetent <i>E. coli</i> cells	121
5.3.4.3	DNA transfer from <i>E. coli</i> to <i>Streptomyces</i> via conjugation	122
5.3.5	Protein overexpression, purification and analysis.....	123
5.3.5.1	IPTG-induced gene expression	123
5.3.5.2	Recombinant fusion protein purification	123
5.3.5.3	SDS-PAGE gel electrophoresis.....	123
5.3.5.4	Determining protein concentration	124
5.3.5.5	Determining protein molecular weight	124
5.3.5.6	CD spectroscopic analysis of proteins	125
5.3.5.7	Tryptic digestion of proteins	125
5.3.5.8	Analysis of putative protein-protein interactions.....	125

5.3.6	Attempted detection of putative polyketide alkaloid metabolites.....	126
5.3.6.1	Extraction of metabolites from agar.....	126
5.3.6.2	Extraction of metabolites from liquid media	126
5.3.6.3	Sample preparation for LC-MS.....	127
5.3.7	Overproduction and purification of valinomycin fragments.....	127
5.3.7.1	Characterisation data.....	128
5.4	Biochemical assays	129
5.4.1	<i>In vitro</i> investigations with CpkG.....	129
5.4.1.1	UV-Vis spectroscopic analysis of PLP binding to CpkG	129
5.4.1.2	Spectrophotometric analysis of the substrate specificity of CpkG	129
5.4.1.3	Determination of the absolute stereochemistry of alanine.....	130
5.4.1.4	Detection of aldehyde	131
5.4.1.5	Detection of pyruvate and amine	131
5.4.2	<i>In vitro</i> investigations with CpkE and CpkI.....	132
5.4.2.1	Incubation of <i>trans</i> -octen-2-al with CpkE	132
5.4.2.2	Incubation of dodecane-1,5-diol with CpkI.....	132
5.4.3	<i>In vitro</i> investigations with CpkD, ScF and CpkH	132
5.4.3.1	UV-Vis spectroscopic analysis of flavin binding.....	132
5.4.3.2	UV-Vis spectroscopic analysis of NAD(P)H oxidation by flavoprotein	133
5.4.3.3	CpkH-catalysed oxidation of 1-aminododecan-5-ol	133
5.4.3.4	Incubation of (<i>R,E</i>)- or (<i>S,E</i>)-1-aminododec-2-en-5-ol with coelimycin biosynthetic enzymes	133
5.4.3.5	CpkD-catalysed epoxidation of ethyl (<i>2E,4E,6E</i>)-octa-2,4,6- trienoate.....	134
5.5	Crystallisation of CpkG.....	134
5.6	Chemical procedures	135
5.6.1	Synthesis of putative CpkI substrate and CpkE standard	135
5.6.1.1	Dodecane-1,5-diol (76)	135
5.6.1.2	<i>cis</i> -Octen-3-al (79)	135
5.6.2	Synthesis of racemic CpkG and CpkH substrates and standards.....	136
5.6.2.1	5-Hydroxydodecanal (59)	136
5.6.2.2	1-Aminododecan-5-ol (60).....	137
5.6.2.3	<i>tert</i> -Butyl (5-hydroxydodecyl)carbamate (82).....	138

5.6.2.4	<i>tert</i> -Butyl (5-oxododecyl)carbamate (83)	138
5.6.2.5	6-Heptyl-2,3,4,5-tetrahydropyridine (81)	139
5.6.3	Synthesis of enantiopure CpkH substrates	140
5.6.3.1	Methyl-3-oxodecanoate (87)	140
5.6.3.2	Methyl (<i>R</i>)-3-hydroxydecanoate (88)	140
5.6.3.3	Methyl (<i>S</i>)-3-hydroxydecanoate (88)	141
5.6.3.4	Methyl (<i>R</i>)-3-((<i>tert</i> -butyldimethylsilyl)oxy)decanoate (89)	141
5.6.3.5	Methyl (<i>S</i>)-3-((<i>tert</i> -butyldimethylsilyl)oxy)decanoate (89)	142
5.6.3.6	(<i>R</i>)-3-((<i>tert</i> -Butyldimethylsilyl)oxy)decanal (90)	142
5.6.3.7	(<i>S</i>)-3-((<i>tert</i> -Butyldimethylsilyl)oxy)decanal (90)	143
5.6.3.8	Ethyl (<i>R,E</i>)-5-((<i>tert</i> -butyldimethylsilyl)oxy)dodec-2-enoate (91)	143
5.6.3.9	Ethyl (<i>S,E</i>)-5-((<i>tert</i> -butyldimethylsilyl)oxy)dodec-2-enoate (91)	144
5.6.3.10	(<i>R,E</i>)-5-((<i>tert</i> -Butyldimethylsilyl)oxy)dodec-2-en-1-ol (92)	144
5.6.3.11	(<i>S,E</i>)-5-((<i>tert</i> -Butyldimethylsilyl)oxy)dodec-2-en-1-ol (92)	145
5.6.3.12	(<i>R,E</i>)- <i>tert</i> -Butyl((1-chlorododec-2-en-5-yl)oxy)dimethylsilane (93) ...	145
5.6.3.13	(<i>S,E</i>)- <i>tert</i> -Butyl((1-chlorododec-2-en-5-yl)oxy)dimethylsilane (93) ...	146
5.6.3.14	(<i>R,E</i>)-2-(5-((<i>tert</i> -Butyldimethylsilyl)oxy)dodec-2-en-1-yl) isoindoline-1,3-dione (94)	147
5.6.3.15	(<i>S,E</i>)-2-(5-((<i>tert</i> -Butyldimethylsilyl)oxy)dodec-2-en-1-yl) isoindoline-1,3-dione (94)	147
5.6.3.16	(<i>R,E</i>)-2-(5-Hydroxydodec-2-en-1-yl)isoindoline-1,3-dione (95)	148
5.6.3.17	(<i>S,E</i>)-2-(5-Hydroxydodec-2-en-1-yl)isoindoline-1,3-dione (95)	149
5.6.3.18	(<i>R,E</i>)-1-Aminododec-2-en-5-ol (84)	149
5.6.3.19	(<i>S,E</i>)-1-Aminododec-2-en-5-ol (84)	149
5.6.3.20	(<i>R</i>)-2-(5-Hydroxydodecyl)isoindoline-1,3-dione (98)	150
5.6.3.21	(<i>S</i>)-2-(5-Hydroxydodecyl)isoindoline-1,3-dione (98)	150
5.6.3.22	(<i>R</i>)-1-Aminododecan-5-ol (60)	151
5.6.3.23	(<i>S</i>)-1-Aminododecan-5-ol (60)	151
5.6.4	Verification of absolute stereochemistry of substrate intermediates	152
5.6.4.1	(<i>R,E</i>)-1-(1,3-Dioxoisindolin-2-yl)dodec-2-en-5-yl (<i>R</i>)-3,3,3- trifluoro-2-methoxy-2-phenylpropanoate (96)	152
5.6.4.2	(<i>R,E</i>)-1-(1,3-Dioxoisindolin-2-yl)dodec-2-en-5-yl (<i>S</i>)-3,3,3- trifluoro-2-methoxy-2-phenylpropanoate (96)	153

5.6.4.3	(<i>S,E</i>)-1-(1,3-Dioxoisindolin-2-yl)dodec-2-en-5-yl (<i>R</i>)-3,3,3-trifluoro-2-methoxy-2-phenylpropanoate (96).....	153
5.6.4.4	(<i>S,E</i>)-1-(1,3-Dioxoisindolin-2-yl)dodec-2-en-5-yl (<i>S</i>)-3,3,3-trifluoro-2-methoxy-2-phenylpropanoate (96).....	154
5.6.5	Synthesis of CpkD substrate	154
5.6.5.1	Ethyl (<i>2E,4E,6E</i>)-octa-2,4,6-trienoate (99).....	154
6	REFERENCES	156
7	APPENDICES	169

List of figures

Figure 1.1: The life cycle of *Streptomyces coelicolor* A3(2). Spore germination results in the formation of germ tubes, which extend and branch to produce the substrate mycelium. Hyphae subsequently grow into the air to generate the aerial mycelium, which develops and matures into long chains of spores. Spores are then released to start a new cycle.

Figure 1.2: Selected bioactive compounds produced by *Streptomyces*, including antibacterials (1-3), insecticides (4), herbicides (5), anticancer agents (6), immunosuppressants (7) and antifungals (8)

Figure 1.3: *Streptomyces coelicolor* specialised metabolites, discovered prior to the availability of the genome sequence through a classical bioassay-guided approach

Figure 1.4: Selected *Streptomyces coelicolor* specialised metabolites, discovered through a genomics-driven approach

Figure 1.5: The organisation of the 59 kb-long *cpk* gene cluster directing coelimycin biosynthesis in *Streptomyces coelicolor* M145. The proposed functions of the proteins encoded by the genes are as follows: Purple: precursor supply; red: PKS component; blue: post-PKS tailoring; green: regulatory; gold: export; and grey: hypothetical.

Figure 1.6: Structures of the characterised shunt metabolite, coelimycin P1 **18**, and the putative true metabolite, coelimycin A **19**, of the *cpk* biosynthetic pathway

Figure 1.7: The structures of riboflavin, FMN and FAD. The isoalloxazine ring and ribityl side chain are highlighted in purple and blue, respectively.

Figure 1.8: The different oxidation states of flavin, in protonated (top) and unprotonated (bottom) forms

Figure 1.9: The structure of NADH (R = H) and NADPH (R = PO₃²⁻). The nicotinamide and ADP ribose groups are highlighted in purple and blue, respectively.

Figure 1.10: Regulation of coelimycin biosynthesis, involving *S. coelicolor* butenolide (SCB) signalling molecules and interactions with receptors and ligands from other pathways (grey). Activatory and inhibitory steps are indicated by arrows or lines ending with a bar, respectively. Additional binding sites of ScbR2 (intergenic regions of *accaI-scF*, *cpkA-cpkD*, *cpkI-cpkJ* and the promoter of *cpkN*) have been omitted for clarity. Adapted from Liu and coworkers.

Figure 1.11: Structures of polyketide alkaloids from Actinobacteria

Figure 2.1: (A) Nickel affinity purification of His₆-CpkG. L1: Molecular weight marker; L2: elution with 20 mM imidazole; L3: elution with 50 mM imidazole; L4: elution with 100 mM imidazole; L5: elution with 200 mM imidazole; L6: elution with 300 mM imidazole; L7: flow through from column; L8: soluble protein; L9: insoluble protein. (B) Deconvoluted mass spectrum of His₆-CpkG. The measured and calculated molecular weights of His₆-CpkG are 60648 Da and 60646 Da, respectively.

Figure 2.2: UV/Vis spectra and structures of PLP (light grey) and the aldimine adduct of PLP with CpkG (dark grey). The λ_{\max} values of the two peaks are 410 nm and 418 nm, respectively.

Figure 2.3: Amino donor and acceptor specificities of CpkG as determined by the CuSO₄/MeOH staining method. The specificities for the different substrate combinations are shown relative to the specificity of CpkG for octylamine and pyruvate as amino donor and acceptor. Data are represented as means \pm 1 standard deviation from triplicate experiments.

Figure 2.4: Extracted ion chromatograms at $m/z = 232.1366 \pm 0.005$ (corresponding to the [M+H]⁺ ion of the D-cysteine derivative of octanal) from LC-MS analyses of the CpkG-catalysed reaction of octylamine with pyruvate. From top to bottom: enzymatic reaction; control from which the enzyme was omitted; and authentic standard.

Figure 2.5: Extracted ion chromatograms at $m/z = 342.1044 \pm 0.005$ (corresponding to the [M+H]⁺ ion of Marfey's-derivatised alanine) from LC-MS analyses of the CpkG-catalysed reaction of octylamine with pyruvate. From top to bottom: Marfey's derivatisation of the product of the reaction; Marfey's derivatisation of the control reaction from which enzyme was omitted; Marfey's derivatisation of an L-alanine standard; Marfey's derivatisation of a D-alanine standard.

Figure 2.6: (A) Detection of pyruvate formation from L-alanine using a coupled assay with LDH and NADH. The consumption of NADH was determined by monitoring the decrease in absorbance at 340 nm after CpkG addition (dark grey) and without CpkG addition (light grey). (B) Extracted ion chromatograms at $m/z = 130.1590 \pm 0.005$ (corresponding to the [M+H]⁺ ion for octylamine) from LC-MS analyses of the CpkG-catalysed reaction of octanal with L-alanine. From top to bottom: enzymatic reaction; control reaction from which the enzyme was omitted; and authentic standard.

Figure 2.7: (A) Extracted ion chromatograms at $m/z = 202.2165 \pm 0.005$ (corresponding to the [M+H]⁺ ion for 1-aminododecan-5-ol) from LC-MS analyses of the CpkG-catalysed reaction of 5-hydroxydodecanal with L-alanine. From top to bottom:

enzymatic reaction; control reaction from which the enzyme was omitted; and authentic standard. (B) Extracted ion chromatograms at $m/z = 304.1941 \pm 0.005$ (corresponding to the $[M+H]^+$ ion for D-cysteine-derivatised 5-hydroxydodecanal) from LC-MS analyses of the CpkG-catalysed reaction of 1-aminododecan-5-ol with pyruvate. From top to bottom: enzymatic reaction; control reaction from which the enzyme was omitted; and authentic standard.

Figure 2.8: A crystal obtained from a PACT premier™ screen

Figure 2.9: (A) A cartoon display of the CpkG homodimer. Monomer units are highlighted in violet and teal, while PMP ligands are depicted as spheres. (B) A cartoon display of the domain architecture of a CpkG monomer. The large domain, small domain and additional domain are highlighted in blue, green and magenta, respectively, while the PMP ligand is depicted as sticks in CPK colouring. (C) The predicted substrate binding tunnel positioned behind the primary amino group of PMP, displayed in surface view.

Figure 2.10: (A) Polar interactions of PMP with surrounding residues. (B) Face-to-edge π - π stacking interaction of PMP with Phe209. The carbon chains of PMP and residues are depicted in grey and teal, respectively, while other atoms are shown in CPK colouring.

Figure 2.11: (A) Polar interactions of PLP with surrounding residues. (B) An overlay of CpkG_ala (teal) with CpkG_ala_lactol (violet). Residues that have undergone significant conformational alternations are depicted. Nitrogen, oxygen and phosphorous atoms are shown in CPK colours.

Figure 2.12: (A) SDS-PAGE analysis of His₆-CpkC-TR (L2) and His₆-CpkG (L3), relative to a protein molecular weight marker (L1). The molecular weights of His₆-CpkC-TR and His₆-CpkG are 46.0 and 60.6 kDa, respectively. (B) Native-PAGE analysis of His₆-CpkC-TR (L1), His₆-CpkG (L2) and a 1:1 mixture of His₆-CpkC-TR and His₆-CpkG (L3).

Figure 2.13: Left: SDS-PAGE analysis of (A) CpkG and (B) CpkC-TR, following cleavage of hexahistidine tags with TEV protease. L1: Protein molecular weight marker; L2: His₆-tagged protein standard; L3: Protein collected from wash fraction (20 mM imidazole); L4: Protein collected from elution fraction (300 mM imidazole). Right: SDS-PAGE analysis of fractions collected from pull-down assays, following incubation of (C) His₆-CpkC-TR with CpkG, and (D) His₆-CpkG with CpkC-TR. L1: Protein molecular weight marker; L2: protein collected from wash fraction; L3: protein collected from elution fraction.

Figure 2.14: A partial display of a multiple sequence alignment of CpkG and homologues encoded within putative polyketide alkaloid biosynthetic gene clusters, produced using T-Coffee and BoxShade software. Darker regions indicate higher levels of homology between proteins. The additional domain of CpkG (residues 142-189) appears to contain fewer amino acids, which may have been necessary for interactions with CpkC-TR.

Figure 2.15: Overview of gene clusters harbouring PKS genes and homologues of *cpkC-TR* and *cpkG*

Figure 2.16: Base peak chromatograms from LC-MS analyses of *S. cyaneofuscatum* metabolites extracted from ISP2 (top), R3 (middle) and R5 (bottom) agar

Figure 2.17: Plasmid map of pOSV556t. An integrase gene (*int_pSAM2*), origins of replication (ColE1) and transfer (*oriT*), ampicillin and hygromycin resistance genes (*amp^r* and *hyg^r*) and restriction sites (*XhoI* and *HindIII*) used for cloning are highlighted.

Figure 2.18: (A) Gel from agarose electrophoresis following PCR amplification of *IF17_RS0112435*; (B) plasmid map with restriction sites used for analysis displayed; (C) gel from agarose electrophoresis following restriction digest with *HindIII* and *XhoI*; and (D) gel from agarose electrophoresis following restriction digest with *NotI*.

Figure 2.19: Plasmid map of pKC1132. The *lacZ α* gene, an apramycin resistance gene (*am^R*), origins of replication (*rep^{PUc}*) and transfer (*oriT*) and restriction sites used for cloning (*HindIII* and *XbaI*) are highlighted.

Figure 2.20: (A) Gel from agarose electrophoresis following PCR amplification of *IF17_RS0112385*; (B) plasmid map with restriction sites used for analysis displayed; (C) gel from agarose electrophoresis following restriction digest with *HindIII* and *XbaI*; and (D) gel from agarose electrophoresis following restriction digest with *SmaI*.

Figure 2.21: Base peak chromatograms from LC-MS analyses of *S. cyaneofuscatum* metabolites extracted from R3 agar. Black trace: wild-type *S. cyaneofuscatum*; red trace: *S. cyaneofuscatum*PKS::*am*; green trace: *S. cyaneofuscatum*::*SARP*; blue trace: *S. cyaneofuscatum*::*SARP*-PKS::*am*.

Figure 2.22: (A) Characterised (**64** and **65**) and predicted (**67**) structures of *S. cyaneofuscatum*::*SARP* metabolite fragments; and (B) structure of valinomycin **66** with isolated and predicted fragments highlighted.

Figure 2.23: Comparison of predicted (top) and known (bottom) valinomycin biosynthetic gene clusters in *Streptomyces cyaneofuscatum* NRRL B-2570 and *Streptomyces tsusimaensis* ATCC 15141, respectively (Dr. Emmanuel de los Santos, University of Warwick). Homologous genes are coloured similarly, and are proposed to

encode a type II thioesterase (green), NRPS modules 1 and 2 (blue), NRPS modules 3 and 4 (purple) and a transposase (pink).

Figure 2.24: (A) Base peak chromatograms from LC-MS analyses of *S. cyaneofuscatus* metabolites extracted from AM1 liquid media. Black trace: wild-type *S. cyaneofuscatus*; green trace: *S. cyaneofuscatus*::*SARP*. (B) Observed $[M+Na]^+$ for metabolite eluting at 26.5 min, likely corresponding to valinomycin (calculated $[M+Na]^+ = 1133.6204$).

Figure 2.25: The organisation of a gene cluster within *Streptomyces abikoensis* DSM 40831, predicted to direct the biosynthesis of latumcidin **35**. The proposed functions of the proteins encoded by the genes are as follows. Purple: precursor supply; red: PKS component (with TR domain highlighted in yellow); blue: post-PKS tailoring; green: regulatory; grey: hypothetical.

Figure 2.26: Structures of the predicted polyketide aldehyde **68** and latumcidin **35**, with carbon atoms numbered accordingly

Figure 3.1: Map of the pET151/D-TOPO vector used for the assembly of protein expression constructs. Key features (clockwise): T7 promoter; *lac* operator (*lacO*); ribosome binding site (RBS); start codon (ATG); *N*-terminal His₆-tag; V5 epitope; TEV recognition site; TOPO® cloning site; *SacI* restriction site; T7 transcription termination region; ampicillin resistance gene (β -lactamase); pBR322 origin of replication (*ori*); ROP gene; and *lacI* gene.

Figure 3.2: Gels from agarose electrophoresis following PCR amplifications of (A) *cpkI*; and (B) *cpkE*, at selected annealing temperatures. The expected sizes of the PCR products are labelled.

Figure 3.3: (A) Plasmid maps with restriction sites used for analysis; and (B) gels from agarose electrophoresis following restriction digests, of pET151::*cpkI* (top) and pET151::*cpkE* (bottom). The expected sizes of the DNA fragments are labelled.

Figure 3.4: SDS-PAGE analysis of purified His₆-CpkI (L2) and His₆-CpkE (L3) alongside a protein molecular weight marker (L1)

Figure 3.5: Measured (left) and deconvoluted (right) mass spectra of His₆-CpkI (top) and His₆-CpkE (bottom). The calculated molecular weights of the proteins are 31214 and 34813 Da, respectively.

Figure 3.6: (A) Extracted ion chromatograms at $m/z = 304.1941 \pm 0.05$ (corresponding to the $[M+H]^+$ ion for (4*S*)-2-(4-hydroxyhexyl)thiazolidine-4-carboxylic acid). From top to bottom: reaction of His₆-CpkI with dodecane-1,5-diol and NAD⁺; reaction of His₆-CpkI

with dodecane-1,5-diol and NADP⁺; and D-cysteine derivatised authentic standard. (B) Extracted ion chromatograms at $m/z = 230.1751 \pm 0.05$ (corresponding to the $[M+H]^+$ ion for tris-derivatised octenal). From top to bottom: reaction of *trans*-octen-2-al with His₆-CpkE and tris; control reaction from which the enzyme has been omitted; and authentic standard of tris-derivatised *cis*-octen-3-al. Experiments were performed in duplicate.

Figure 3.7: (A) Gels from agarose electrophoresis following PCR amplifications at selected annealing temperatures; (B) plasmid maps with restriction sites used for analysis; and (C) gels from agarose electrophoresis following restriction digests of pET151 constructs harbouring *cpkD* (top), *scF* (middle) and *cpkH* (bottom). The expected sizes of the DNA fragments are labelled.

Figure 3.8: SDS-PAGE analysis of purified His₆-CpkD (L2), His₆-ScF (L3) and His₆-CpkH (L4) alongside a protein molecular weight marker (L1)

Figure 3.9: UV/Vis spectra of flavoproteins: (A) His₆-CpkD; (B) His₆-ScF; and (C) His₆-CpkH

Figure 3.10: Measured (top) and deconvoluted (bottom) mass spectra of (A) His₆-CpkD; (B) His₆-ScF; and (C) His₆-CpkH. The calculated protein molecular weights are 51807 Da, 60363 Da (+ FAD = 61148 Da) and 58827 Da (+ FAD = 59611 Da), respectively.

Figure 3.11: Extracted ion chromatograms at $m/z = 784.1499 \pm 0.05$ (top trace, corresponding to the $[M-H]^-$ ion of FAD) and $m/z = 455.0973 \pm 0.05$ (bottom trace, corresponding to the $[M-H]^-$ ion of FMN)

Figure 3.12: Amino acid residues involved in the covalent binding of FAD to TamL. Note: Only the isoalloxazine moiety of FAD is shown.

Figure 3.13: A partial view of a multiple sequence alignment of TamL, ScF and CpkH, produced using T-Coffee and BoxShade software. Darker regions indicate higher levels of homology. Residues implicated in covalent FAD binding are highlighted with a purple box.

Figure 3.14: UV/Vis monitoring at A₃₄₀ to detect the oxidation of NADH (dark grey) or NADPH (light grey), by (A) His₆-CpkD; (B) His₆-ScF; and (C) His₆-CpkH

Figure 3.15: Proposed substrate mimic of 31

Figure 3.16: (A) Extracted ion chromatograms at $m/z = 182.1903 \pm 0.05$ (corresponding to the $[M+H]^+$ ion for 6-heptyl-2,3,4,5-tetrahydropyridine). From top to bottom: reaction of 1-aminododecan-5-ol with CpkD, ScF, CpkH and without enzyme. (B) Measured m/z and predicted molecular formula for the compound eluting at 15.4 min.

Figure 3.17: Extracted ion chromatograms at $m/z = 182.1903 \pm 0.05$ (corresponding to the $[M+H]^+$ ion for 6-heptyl-2,3,4,5-tetrahydro-pyridine). From top to bottom: reaction of 1-aminododecan-5-ol with CpkH; control reaction from which the enzyme was omitted; and authentic standard.

Figure 3.18: Two different representations of (A) the (*R,R*)-diastereomer; (B) the (*R,S*)-diastereomer; (C) the (*S,R*)-diastereomer; and (D) the (*S,S*)-diastereomer of Mosher's-derivatised **95**, adapted from Hoye and coworkers. The shielding effect exerted by the phenyl group is indicated with a purple arrow.

Figure 3.19: Structure of **96** with protons labelled

Figure 3.20: A comparison of ^1H NMR spectra from 4.2–5.8 ppm of (A) (*R,R*)-**96**; (B) (*R,S*)-**96**; (C) (*S,R*)-**96**; and (D) (*S,S*)-**96**. Protons 3, 4, 5 and 7 in spectra (B) and (C) are shifted upfield compared to those of (A) and (D).

Figure 3.21: Extracted ion chromatograms at $m/z = 182.1903 \pm 0.05$ (corresponding to the $[M+H]^+$ ion for 6-heptyl-2,3,4,5-tetrahydropyridine). From top to bottom: reaction of (*R*)-1-aminododecan-5-ol with His₆-CpkH (ion count = 45640 a.u.); reaction of (*S*)-1-aminododecan-5-ol with His₆-CpkH (ion count = 1864 a.u.); control reaction from which the enzyme was omitted; and authentic standard. Independent experiments were performed in triplicate.

Figure 3.22: Proposed substrate mimic of **29**

Figure 3.23: Extracted ion chromatograms at (A) $m/z = 324.1240 \pm 0.05$ (corresponding to the $[M+Na]^+$ ion for the NAC-derivatised mono-epoxide) and (B) 340.1189 ± 0.05 (corresponding to the $[M+Na]^+$ ion for the NAC-derivatised bis-epoxide), Brown trace: reaction with His₆-ScF; blue trace: reaction with His₆-CpkD; purple trace: reaction with His₆-CpkD, supplemented with additional FAD; green trace: reaction with His₆-CpkD, supplemented with additional FMN. Independent experiments were performed in triplicate.

Figure 3.24: Possible structures of a NAC ring-opened mono-epoxide (A) and bis-epoxide (B). Other structures may include regioisomers (e.g. epoxidation at carbons 6-7 and NAC addition at carbon 7), Michael-addition products (e.g. epoxidation at carbons 4-5 and NAC addition at carbon 9) and related diastereomers.

Figure 3.25: Extracted ion chromatograms at (A) $m/z = 324.1240 \pm 0.05$ (corresponding to the $[M+Na]^+$ ion for the NAC-derivatised mono-epoxide) and (B) 340.1189 ± 0.05 (corresponding to the $[M+Na]^+$ ion for the NAC-derivatised bis-epoxide) arising from the incubation of His₆-CpkD with triene ester **99**, FMN, NADH and NAC. From top to

bottom: reaction without substrate; reaction without NAC; reaction without His₆-CpkD; reaction with boiled His₆-CpkD; and reaction without NADH. Independent experiments were performed in triplicate.

Figure 3.26: CD spectra of non-denatured (grey) and boiled (orange) His₆-CpkD

Figure 3.27: Extracted ion chromatograms at $m/z = 259.0545 \pm 0.005$ (corresponding to the $[M+Na]^+$ ion for diNAC). Green trace: incubation of His₆-CpkD with triene ester **99**, FMN, NADH and NAC; purple trace: control reaction from which FMN was omitted; and brown trace: control reaction from which NAC was omitted.

Figure 3.28: Extracted ion chromatograms from LC-MS analysis of reactions of CpkD, FMN, triene ester **99** and NAC, in air (A) and under ¹⁸O₂ (B). The extracted ion chromatograms at $m/z = 324.1240 \pm 0.05$ (green) and $m/z = 326.1240 \pm 0.05$ (purple) correspond to the ¹⁶O and ¹⁸O isotopes of the mono-epoxide, respectively. Independent assays were performed in duplicate.

Figure 3.29: Extracted ion chromatograms from LC-MS analysis of reactions of CpkD, FMN, triene ester **99** and NAC, in air (A) and under ¹⁸O₂ (B). The extracted ion chromatograms at $m/z = 340.1189 \pm 0.05$ (green), $m/z = 342.1189 \pm 0.05$ (purple) and $m/z = 344.1189 \pm 0.05$ (brown) correspond to the ¹⁶O, ¹⁸O₁ and ¹⁸O₂ isotopes of the bis-epoxide, respectively. Independent assays were performed in duplicate.

Figure 3.30: Histograms depicting the relative contributions of isotopes to the A+2, B+2 and B+4 peaks, following reaction of triene ester **99** with His₆-CpkD, FMN and NAC under an ¹⁸O₂-enriched atmosphere. For simplicity, contributions of naturally occurring isotopes are annotated as ¹³C₂ and ¹³C₄ responses. Calculations and data display are adapted from Johnson and coworkers.

List of schemes

Scheme 1.1: The organisation of the DEBS system, responsible for the assembly of 6-deoxyerythronolide B **20**. Subsequent modification by post-PKS tailoring enzymes affords erythromycin A **21**.

Scheme 1.2: Mechanism of reaction of apo-ACP with coenzyme A to form holo-ACP. B = general base.

Scheme 1.3: Mechanism of reaction of an AT domain with a starter unit (e.g. propionyl-CoA, R = H) or extender unit (e.g. methylmalonyl-CoA, R = CO₂⁻), and the subsequent transfer to a downstream holo-ACP. The catalytic histidine residue is indicated. M_x = module *x*.

Scheme 1.4: Mechanism of reaction of a KS domain with an upstream acyl-ACP, followed by a decarboxylative Claisen condensation with a methylmalonyl unit tethered to a downstream ACP. Catalytic histidine residues are indicated. R = polyketide chain; M_x = module *x*.

Scheme 1.5: Mechanisms of (A) reduction; and (B) epimerisation of a β-ketothioester, catalysed by a KR domain. The serine, tyrosine and lysine residues within the catalytic triad are indicated. R = polyketide chain; A = general acid; B = general base; M_x = module *x*.

Scheme 1.6: Mechanism of dehydration of a β-hydroxythioester, catalysed by a DH domain. Catalytic histidine and aspartate residues are indicated. R = polyketide chain; M_x = module *x*.

Scheme 1.7: Mechanism of reduction of an α,β-unsaturated thioester, catalysed by an ER domain. R = polyketide chain; A = general acid; M_x = module *x*.

Scheme 1.8: Mechanisms of chain release of a fully assembled polyketide chain by a TE domain, to yield a carboxylic acid (top) or macrolactone (bottom). The serine, histidine and aspartate residues within the catalytic triad are indicated.

Scheme 1.9: The PKS module and domain organisation of CpkABC and the proposed structures of ACP-bound intermediates. The cross indicates that the DH domain in module 4 is likely inactive.

Scheme 1.10: Mechanism of NADH-mediated reductive chain release of **22** by CpkC-TR. A = general acid.

Scheme 1.11: The conversion of pyruvate **25** to malonyl-CoA **27**, catalysed by a pyruvate-dependent ferredoxin reductase and an acetyl-CoA carboxylase. Relevant cofactors are highlighted.

Scheme 1.12: Catalytic cycle of a TA. The amino donor (top left) transfers an amino group to enzyme-bound pyridoxal-5'-phosphate (E-PLP), and the newly-formed enzyme-bound pyridixamine-5'-phosphate (E-PMP) transfers it on to the amino acceptor (bottom right). For α -TAs, $R^I = R^{III} = \text{CO}_2\text{H}$. For ω -TAs, if $R^I = \text{CO}_2\text{H}$, then $R^{III} \neq \text{CO}_2\text{H}$.

Scheme 1.13: Reaction mechanism of a TA. In the first half of the reaction, an amino group is transferred from an α -amino acid to enzyme-bound PLP, to form an α -keto acid and PMP. In the second half of the reaction, the amino group is transferred from PMP to an aldehyde, to form a primary amine and enzyme-bound PLP. Note: reactions are reversible, but arrows are unidirectional for clarity.

Scheme 1.14: The catalytic cycle of flavoenzymes

Scheme 1.15: The oxidation of an alcohol by a flavin-dependent oxidase *via* a direct hydride transfer. The oxidised-form of the cofactor is subsequently regenerated by molecular oxygen (or an alternative electron acceptor in the case of a dehydrogenase). A = general acid; B = general base.

Scheme 1.16: The epoxidation of an alkene by a flavin-dependent monooxygenase. The C(4a)-hydroperoxy species is formed by reaction of reduced flavin with molecular oxygen. Nucleophilic attack of an alkene at the electrophilic oxygen yields the epoxide product, and a C(4a)-hydroxyflavin adduct which eliminates water to form oxidised flavin. The reduced cofactor is subsequently regenerated by direct hydride transfer from NAD(P)H. A = general acid; B = general base.

Scheme 1.17: The dehydrogenation of an alcohol by a short-chain dehydrogenase, using NAD(P)⁺ as a cofactor

Scheme 1.18: The isomerisation of *trans*-2-decenoic acid to *cis*-3-decenoic acid by β -hydroxydecanoyl thioester isomerase. A = general acid. Revised from Bugg, Schwab and Klassen.

Scheme 1.19: The proposed biosynthesis of coelimycin A **19** and coelimycin P1 **18**

Scheme 1.20: The biosynthesis of *S. coelicolor* butenolides (SCBs, **13**) 1-8. An ACP-bound β -keto thioester from a fatty acid synthase (FAS) is condensed with dihydroxyacetone phosphate (DHAP) by ScbA and the double bond reduced by ScbB. A dephosphorylation (by an unknown enzyme) and reduction of the keto group by ScbC afford the SCBs.

Scheme 2.1: Proposed reaction catalysed by CpkG in coelimycin biosynthesis

Scheme 2.2: The general transamination reaction catalysed by CpkG.

Scheme 2.3: Assay employed to detect amino acids produced during the transamination reaction. Amines (10 mM, **45-49**) and α -keto acids (10 mM, **25** and **50**) tested as substrates for CpkG (20 μ M) are shown.

Scheme 2.4: Assay employed for the detection and derivatisation of octanal **57** with D-cysteine (400 μ M), following reaction of pyruvate **25** (200 μ M) and octylamine **49** (200 μ M) with CpkG (90 μ M).

Scheme 2.5: Assay employed to establish the stereochemistry of alanine **51** produced, following reaction of octylamine **49** (10 mM) and pyruvate **25** (10 mM) with CpkG (20 μ M).

Scheme 2.6: Assay employed to detect octylamine **49** and pyruvate **25** production, following reaction of octanal **57** (200 μ M) and alanine **51** (200 μ M) with CpkG (90 μ M). Octylamine was observed by LC-MS comparison to an authentic standard, while pyruvate was detected by monitoring the decrease in absorbance at 340 nm after addition of LDH (1 μ M) and NADH (200 μ M).

Scheme 2.7: Left: Synthesis of 5-hydroxydodecanal **59** by reduction of δ -dodecalactone **61** (1.0 eq.) with DIBAL-H (1.0 eq.). Right: Synthesis of 1-aminododecan-5-ol **60** by reaction of **59** (1.0 eq.) with benzylamine (1.3 eq.), Pd/C (catalytic) and H₂ (1 atm).

Scheme 2.8: General assay for the detection of 1-aminododecan-5-ol **60** (forward reaction) or 5-hydroxydodecanal **59** (reverse reaction), from incubation of the appropriate substrates (200 μ M) with CpkG (90 μ M). For the reverse reaction, **59** was derivatised with D-cysteine (400 μ M) prior to LC-MS analysis.

Scheme 2.9: Possible biosynthetic pathways leading to the formation of the amine intermediate. Grey: original proposed pathway; black: alternative pathway in which CpkC-TR and CpkG interact with one another, resulting in the direct conversion of the aldehyde to the amine. Note: the postulated isomerisation catalysed by CpkE has been omitted for clarity.

Scheme 2.10: Proposed universal biosynthesis of actinobacterial polyketide alkaloids

Scheme 2.11: Proposed biosynthesis of polyketide alkaloid intermediates in *Streptomyces cyaneofuscatus* NRRL B-2570

Scheme 2.12: A proposed biosynthesis of latumcidin **35**, based on the predicted functions of genes within cluster 43 of *Streptomyces abikoensis* DSM 40831 and the pathways proposed for streptazone E **42** and coelimycin P1 **18** biosynthesis

Scheme 3.1: Proposed reactions catalysed by CpkI and CpkE in coelimycin biosynthesis

Scheme 3.2: Synthesis of dodecane-1,5-diol **76** by reduction of δ -dodecalactone **61** (1.0 eq.) with LiAlH₄ (1.5 eq.)

Scheme 3.3: Assay employed to probe the possible oxidation of dodecane-1,5-diol **76** (200 μ M) by CpkI (90 μ M) and NAD(P)⁺ (200 μ M), followed by derivatisation with D-cysteine (400 μ M)

Scheme 3.4: Assay employed to probe the possible isomerisation of *trans*-octen-2-al **78** (200 μ M) by CpkE (100 μ M), and derivatisation with tris (20 mM)

Scheme 3.5: Synthesis of *cis*-octen-3-al **79** by oxidation of *cis*-octen-3-ol **80** (1.0 eq.) with DMP (1.3 eq.)

Scheme 3.6: Proposed reactions catalysed by CpkD, CpkH and ScF in coelimycin biosynthesis

Scheme 3.7: Proposed oxidation reaction catalysed by CpkD, CpkH or ScF

Scheme 3.8: Assay employed to probe flavin-dependent dehydrogenase activity, by incubation of 1-aminododecane-5-ol (200 μ M) with CpkD, ScF or CpkH (20 μ M)

Scheme 3.9: Synthesis of 6-heptyl-2,3,4,5-tetrahydropyridine **81** from 1-aminododecan-5-ol **60**; 18 % yield over 5 steps from δ -dodecalactone **61**

Scheme 3.10: Proposed CpkH-catalysed oxidation of (*R*)-**84** or (*S*)-**84**

Scheme 3.11: Synthetic routes to (*R*)-**84** and (*S*)-**84**. **a**) CDI, monomethyl malonate potassium salt, MgCl₂, THF, r.t., o/n; **b**) [Ru(OAc)₂((*R*)-BINAP)] or [Ru(OAc)₂((*S*)-BINAP)], MeOH, HCl, H₂ (49 atm), 50 °C, o/n; **c**) *tert*-butyldimethylsilyl chloride, imidazole, DMF, 0 °C \rightarrow r.t., o/n; **d**) DIBAL-H, DCM, -78 °C, 1 h; **e**) ethyl(triphenylphosphoranylidene)acetate, THF, r.t., o/n; **f**) DIBAL-H, DCM, -78 °C, 3 h; **g**) (*R*)-enantiomer: triphenylphosphine, *N*-chlorosuccinamide, THF, r.t., 3.5 h; (*S*)-enantiomer: triethylamine, tosyl chloride, DCM, 0 °C \rightarrow r.t., o/n; **h**) potassium phthalimide, DMF, r.t., 72 h; **i**) (1*S*)-(+)-10-camphorsulfonic acid, MeOH, DCM, r.t., 1.5 h; **j**) 40 % aq. MeNH₂, EtOH, r.t., 2 h.

Scheme 3.12: Catalytic cycle for the asymmetric hydrogenation of β -keto esters by [Ru(Cl)₂BINAP)], generated *in situ* from [Ru(OAc)₂BINAP)] and methanolic HCl. In the first step, the active Ru^{II} monohydride complex is formed from reaction of the dichloride

precatalyst with H₂. The β-keto ester substrate then chelates to the complex, displacing two solvent ligands. The keto oxygen is then protonated, increasing the electrophilicity at the carbonyl carbon and thereby facilitating intramolecular hydride transfer. The chiral β-hydroxy ester product is then displaced by incoming solvent ligands. The ruthenium complex then reacts with H₂ to regenerate the monohydride, thus completing the catalytic cycle. L = solvent; P∧P = BINAP.

Scheme 3.13: Enantio-determining transition states during reaction of a β-keto ester with [RuCl₂(*R*)-BINAP] and H₂, adapted from Noyori et al. The phenyl rings projecting forward and backward are depicted as shaded and colourless hexagons, respectively. BINAP rings have been omitted for clarity. In TS_R, the *R* group occupies an open quadrant within the chiral template. In TS_S, there is a significant nonbonding repulsion between the *R* group and the equatorial phenyl ring, indicated by the double-headed arrow. Thus, TS_R, and therefore the (*R*)-β-hydroxy-ester product, are favoured.

Scheme 3.14: Reaction of aldehyde **90** with ethyl(triphenylphosphoranylidene)acetate to form (*E*)-**91** as the major product. Unfavourable 1,2-steric interactions between the *R* group and the ester moiety are minimised in the *trans*-oxaphosphetane transition state.

Scheme 3.15: Reaction of vinyl alcohol (*S*)-**92** with TsCl and Et₃N to form a tosylate intermediate. The leaving group is subsequently displaced by a chloride ion to afford (*S*)-**93**.

Scheme 3.16: Synthesis of Mosher's esters of (*R*)- and (*S*)-**95**. Reaction conditions: (*R*)- or (*S*)-**95** (1.0 eq.), (*R*)- or (*S*)-Mosher's acid (MTPA: α-methoxy-α-trifluoromethyl-phenylacetic acid, 1.1 eq.), EDC (1.1 eq.), DMAP (0.1 eq.), DCM, r.t..

Scheme 3.17: Assays employed to investigate the hypothetical conversion of (*R,E*)- or (*S,E*)-1-aminododec-2-en-5-ol **84** (200 μM) to 6-heptyl-2,3-dihydropyridine **97** or 6-heptyl-2,3,4,5-tetrahydropyridine **81**, catalysed by CpkH (20 μM) and either CpkE (20 μM) or CpkI (20 μM), respectively.

Scheme 3.18: Left: Reduction of (*R,E*)- and (*S,E*)-**95** (1.0 eq.) with Pd/C (catalytic) and H₂ (1 atm), to form (*R*)- and (*S*)-**98**. Right: Deprotection of (*R*)- and (*S*)-**98** (1.0 eq.) with MeNH₂ (excess), to form (*R*)- and (*S*)-**60**.

Scheme 3.19: Assay employed to investigate the oxidation of (*R*)- or (*S*)-**60** (200 μM), catalysed by CpkH (20 μM)

Scheme 3.20: Proposed epoxidation reactions catalysed by CpkD and/or ScF

Scheme 3.21: Attempts to synthesise ethyl (*2E,4E,6E*)-octa-2,4,6-trienoate **99** from *trans,trans*-2,4-hexadienal **100**

Scheme 3.22: Putative epoxidation of **99** (200 μM) by CpkD/ScF (20 μM), FMN/FAD (40 μM) and NADH (40 μM), followed by trapping with NAC (400 μM)

Scheme 3.23: Mechanism of FMN reduction by NADH (A) and NAC (B)

Scheme 3.24: Putative epoxidation of **99** (200 μM) by CpkD (20 μM) and FMN (40 μM), followed by derivatisation with NAC (400 μM), to yield NAC-adducts of a mono- and bis-epoxide (possible structures are shown). Under an $^{18}\text{O}_2$ -atmosphere, one or two atoms of ^{18}O would be expected to be incorporated into the mono and bis-epoxide derivatives, respectively, resulting in an increase in intensity of peaks of 326 and 344 Da.

Scheme 4.1: Updated proposed biosynthesis of coelimycin A **19** and coelimycin P1 **18**. [O] represents an unknown dehydrogenase.

List of tables

Table 2.1: Actinobacterial strain (left), predicted aldehyde intermediate (middle) and predicted corresponding metabolite (right)

Table 3.1: Observed and relative intensities of peaks corresponding to isotopes of mono-epoxide ($m/z = 324$ and 326) and bis-epoxide ($m/z = 340$, 342 and 344) derivatives, from reactions in air and under $^{18}\text{O}_2$

Table 3.2: The corrected relative intensities of peaks corresponding to mono- and bis-epoxide derivatives, following subtraction of isotopic contributions (e.g. $^{13}\text{C}_2$, ^{34}S , $^{13}\text{C}_4$, $^{34}\text{S}_2$, $^{13}\text{C}_2^{34}\text{S}$) from unlabelled and singly-labelled products

Table 5.1: Antibiotics used within this project

Table 5.2: Bacterial strains used within this project

Table 5.3: Vectors used within this project

Table 5.4: PCR primers used within this project

Table 5.5: Volumes of reagents used for 20 μL PCR reactions

Table 5.6: Volumes of reagents used for the preparation of 8 %, 10 % and 12 % gels

Table 5.7: Elution conditions used for LC-MS analyses of purified recombinant proteins

Table 5.8: Elution conditions used for LC-MS detection of putative polyketide alkaloid metabolites *in vivo*, and to assess enzymatic activity *in vitro*

Table 5.9: Elution conditions used for HPLC purification detection **64** and **65**

Table 5.10: Elution conditions used for HPLC purification of 1-aminododecan-5-ol **60**

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Gregory Challis, for giving me the opportunity to work on such a varied and exciting project. His enthusiasm, good humour and guidance have made these last four years an enjoyable and rewarding experience.

Secondly, I would like to thank all of the members of the Challis group, past and present; I really couldn't imagine being around a lovelier group of people: Matt Beech, Patrick Capel, Joshua Cartwright, Marianne Costa, Chris Fage, Angelo Gallo, Chang-An Geng, Richard Gibson, Daniel Griffiths, Christian Hobson, Chuan Huang, Gideon Idowu, Yuki Inahashi, Matthew Jenner, Xinyun Jian, Simone Kosol, Chris Perry, Vincent Poon, Slawomir Potocki, Panward Prasongpholchai, Hannah Pugh, Matias Rey-Corizo, Douglas Roberts, Miriam Rodriguez-Garcia, Rebin Salih, Jake Sargeant, Rakesh Saroay, Helen Smith, Alma Svatos, David Withall and Daniel Zabala-Alvarez. With special thanks to Shanshan Zhou, my wonderful and quirky lab and office neighbour, and Joleen Masschelein, Emmanuel de los Santos, Lona Alkhalaf, Yousef Dashti and Lijiang Song, who have been enormously helpful and a pleasure to collaborate with. I would also like to thank Ruby Awodi, for her previous work on the project, Nick Barker, for his endless positivity and advice, and Anne Smith, Magda Mos and Sheila Hope, for their invaluable technical and organisational support.

Finally, I would like to thank my amazing family, especially my mum, dad and sister, for their love and encouragement, and Daniel Ronan, for listening to my rants and for making me smile. Thanks are also owed to the BBSRC and Chancellor's Scholarship for funding.

Declaration

The experimental work reported in this thesis is original research carried out by the author, unless otherwise stated, in the Department of Chemistry, University of Warwick, between September 2013 and June 2017. No material has been submitted for any other degree, or at any other institution.

Parts of this thesis have been published by the author:

U. R. Awodi, J. L. Ronan, J. Masschelein, E. L. C. de los Santos and G. L. Challis, *Chem. Sci.*, 2017, **8**, 411–415.

Results from other authors are referenced in the usual manner throughout the text.

Jade L. Ronan

Date: _____

Abbreviations

°C	Degrees Celsius
ACP	Acyl carrier protein
ADP	Adenosine diphosphate
AM1	Alkaloid medium 1
AT	Acyltransferase (domain)
ATP	Adenosine triphosphate
BINAP	2,2'-Bis(diphenylphosphino)-1,1'-binaphthyl
bp	Base pairs
Bn	Benzyl
CD	Circular dichromism
CDI	1'-carbonyldiimidazole
CoA	Coenzyme A
COSY	Correlation spectroscopy
CSA	Camphorsulfonic acid
DCM	Dichloromethane
DH	Dehydratase (domain)
DHAP	Dihydroxyacetone phosphate
DIBAL-H	Diisobutylaluminium hydride
DMAP	4-Dimethylaminopyridine
DMF	<i>N,N</i> -Dimethylformamide
DMP	Dess-Martin periodinane
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted ion chromatogram
ER	Enoyl reductase (domain)
ESI	Electrospray ionisation
Et	Ethyl
eq	Equivalents

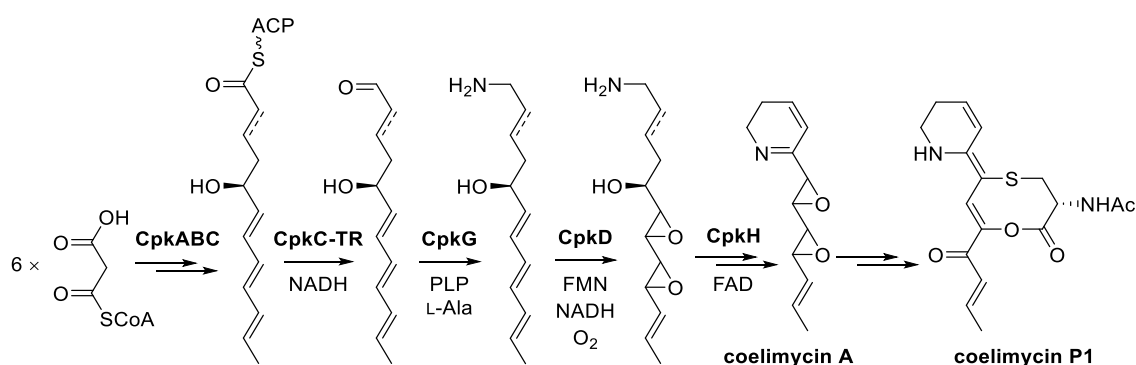
FA	Formic acid
FAD	Flavin adenine dinucleotide
FAS	Fatty acid synthase
Fd	Ferredoxin
FMN	Flavin mononucleotide
FPLC	Fast protein liquid chromatography
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HMBC	Heteronuclear multiple bond coherence
HMQC	Heteronuclear single quantum coherence
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
IMAC	Immobilised metal-affinity chromatography
IPTG	Isopropyl- β -D-thiogalactoside
ISP2/4	International <i>Streptomyces</i> project medium 2/4
KR	Ketoreductase (domain)
KS	Ketosynthase (domain)
LB	Luria-Bertani (medium)
LC-MS	Liquid chromatography-mass spectrometry
LDH	Lactate dehydrogenase
LRMS	Low resolution mass spectrometry
Me	Methyl
MS	Mass spectrometry
MTPA	α -Methoxy- α -trifluoromethylphenylacetic acid
m/z	Mass to charge ratio
NAC	<i>N</i> -acetyl cysteamine
NAD ⁺ / NADH	β -Nicotinamide adenine dinucleotide (ox/red)
NADP ⁺ / NADPH	β -Nicotinamide adenine dinucleotide 2'-phosphate (ox/red)
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthetase
NTA	Nitriloacetic acid
OAT	Ornithine aminotransferase
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis

PCR	Polymerase chain reaction
Phth	Phthalimide
PKS	Polyketide synthase
PLP	Pyridoxal 5'-phosphate
PMP	Pyridoxamine 5'-phosphate
PMSF	Phenylmethylsulfonyl fluoride
PVDF	Polyvinylidene fluoride
rt	Room temperature
SARP	<i>Streptomyces</i> antibiotic regulatory protein
SCB	<i>Streptomyces coelicolor</i> butenolide
SDH	Short-chain dehydrogenase
SDS	Sodium dodecyl sulfate
SFM	Soya flour mannitol (medium)
SMM	Supplemented minimal medium
STE	Sodium chloride-TRIS-EDTA (buffer)
TA	Transaminase
TAT	Twin arginine translocase
TBAF	Tetrabutylammonium fluoride
TE	Thioesterase (domain)
TEMED	Tetramethylethylenediamine
TES	2-[(2-Hydroxy-1,1-bis(hydroxymethyl)ethyl)amino]-ethanesulfonic acid
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TMS	Trimethyl silyl
TPP	Thiamine pyrophosphate
TR	Thioester reductase (domain)
Tris	Tris(hydroxymethyl)aminomethane
TSB	Tryptone soya broth (medium)
UV	Ultraviolet
Vis	Visible

Abstract

Coelimycin P1 is a yellow-pigmented alkaloid with a unique 1,5-oxathiocane ring, produced by *Streptomyces coelicolor* M145. Recently, it was demonstrated that a type I modular polyketide synthase (PKS) with a C-terminal thioester reductase (TR) domain is involved in its assembly. Due to its unusual biosynthetic origin, unprecedented structure and the antibiotic activity likely associated with coelimycin A, its biosynthesis is of great interest.

To investigate the roles of the putative post-PKS tailoring enzymes (CpkG, CpkH, CpkD, ScF, CpkE and CpkI), the recombinant proteins were overproduced in *E. coli*, purified and incubated with various commercial and synthetic substrates. CpkG was characterised as an (*S*)-selective ω -transaminase with a broad substrate tolerance, responsible for the incorporation of nitrogen into the six-membered ring of coelimycin P1. Crystal structures of CpkG revealed a rare tri-domain architecture, key active site residues and provided insight into the transamination mechanism. CpkH, CpkD and ScF were all subsequently characterised as flavoproteins. Specifically, CpkH was demonstrated to catalyse an (*R*)-specific FAD-dependent dehydrogenation, while CpkD was shown to catalyse two FMN-dependent epoxidations. The roles of ScF, CpkE and CpkI in coelimycin biosynthesis remain to be elucidated.



Scheme 1: The revised proposed biosynthesis of coelimycin A and coelimycin P1

Bioinformatics searches identified 22 additional actinobacterial gene clusters, which also encode modular PKSs with a TR domain and a homologue of CpkG. These have been predicted to direct the biosynthesis of both known and novel polyketide alkaloids, suggesting that reductive chain release and transamination constitute a conserved mechanism for the biosynthesis of such metabolites.

1 INTRODUCTION

1.1 *Streptomyces*

Streptomyces are a class of GC-rich, Gram-positive bacteria, with a filamentous form resembling that of fungi. Together, they constitute the largest genus of Actinobacteria.¹ They are ubiquitous in the soil, with their characteristic earthy odour emanating from the production of the volatile metabolite, geosmin.² They are also widespread in marine and freshwater sediments, and in decaying vegetation. *Streptomyces* are essential in these environments due to their involvement in a range of metabolic processes and biotransformations, including the decomposition of biological polymers. Consequently, these bacteria play a fundamental role in carbon and nitrogen cycling.³

Streptomyces have a complex developmental life cycle, which begins with a free spore. Under favourable germination conditions, one or two germ tubes can emerge and grow through tip extension and branching to form a network of multigenomic hyphae, known as the substrate mycelium. This stage often coincides with the onset of specialised metabolite production.⁴ Due to nutrient limitation and other stress signals, aerial hyphae can form and grow into the air, then differentiate into long chains of unigenomic pre-spore compartments. Following septation, the pre-spores can undergo further maturation, involving cell-wall thickening and the production of a grey spore pigment. The dormant spores are then ready for dispersal at the end of the cycle (figure 1.1).^{5,6}

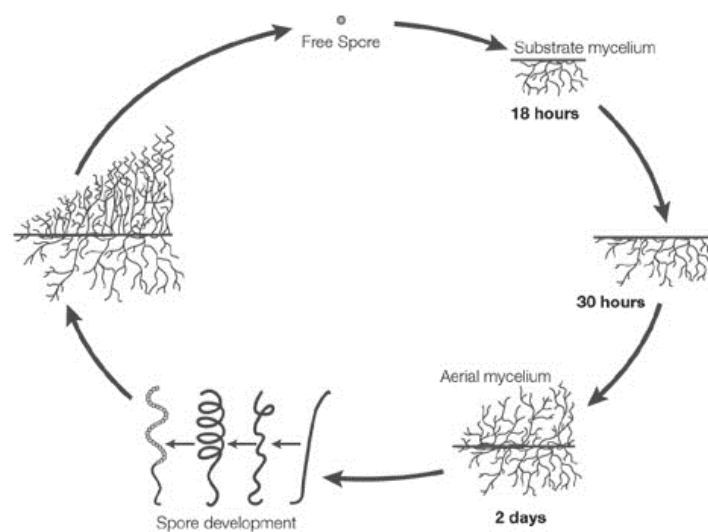


Figure 1.1: The life cycle of *Streptomyces coelicolor* A3(2).⁵ Spore germination results in the formation of germ tubes, which extend and branch to produce the substrate mycelium. Hyphae subsequently grow into the air to generate the aerial mycelium, which develops and matures into long chains of spores. Spores are then released to start a new cycle.

Perhaps most importantly, *Streptomyces* are prolific producers of specialised metabolites. Approximately two-thirds of naturally-occurring antibiotics have been isolated from the genus, including many clinically-approved drugs such as daptomycin **1**, chloramphenicol **2** and tetracycline **3**.^{7,8} In addition to antibiotics, *Streptomyces* produce a diverse array of other bioactive compounds with enormous medicinal and agrochemical potential. Renowned examples include the insecticide avermectin B_{1b} **4**, the herbicide bialaphos **5**, the anticancer agent daunorubicin **6**, the immunosuppressant tacrolimus **7** and the fungicide amphotericin B **8** (figure 1.2).

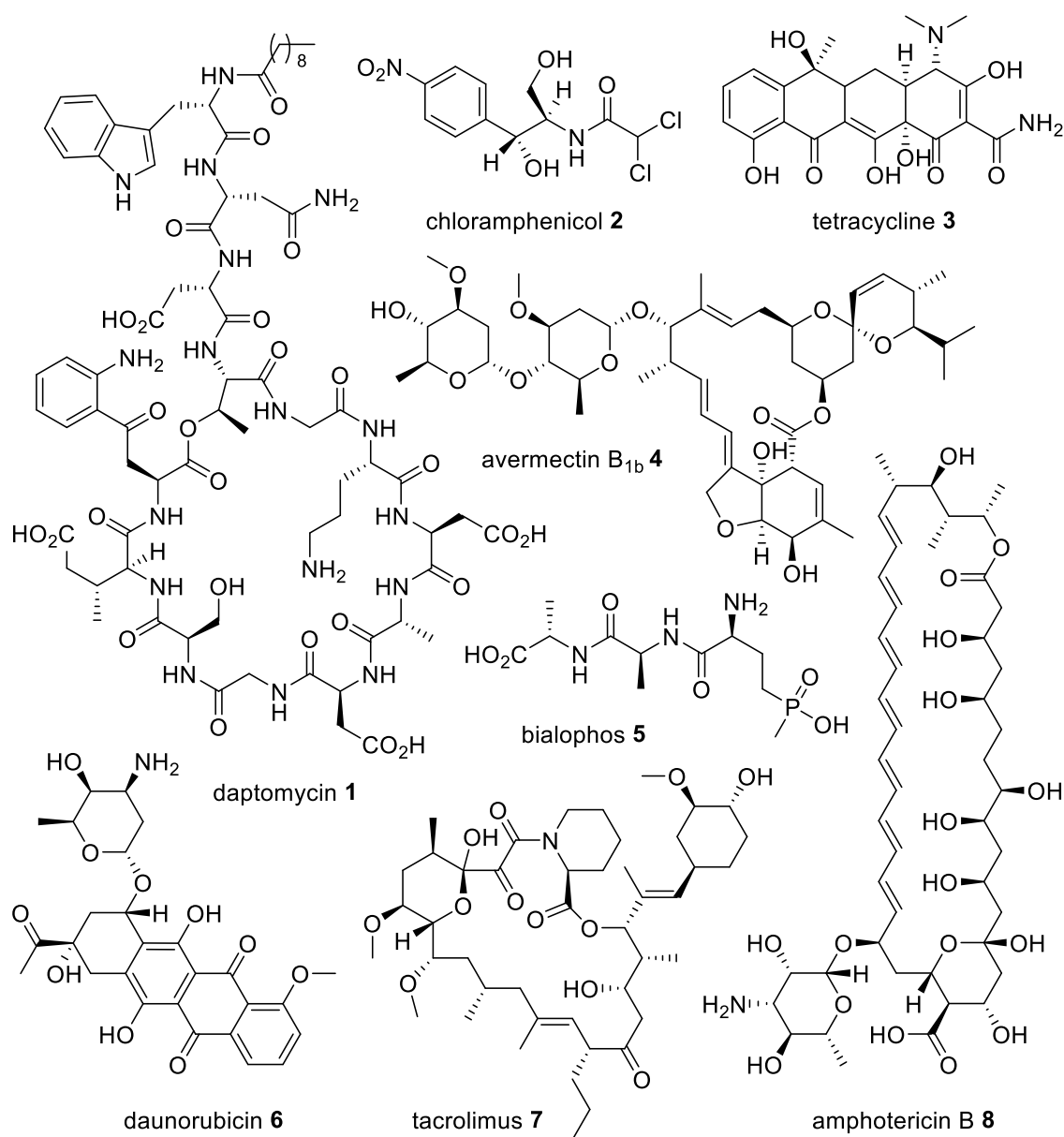


Figure 1.2: Selected bioactive compounds produced by *Streptomyces*, including antibacterials (**1-3**), insecticides (**4**), herbicides (**5**), anticancer agents (**6**), immunosuppressants (**7**) and antifungals (**8**)

1.2 *Streptomyces coelicolor* A3(2)

Streptomyces coelicolor A3(2) is a particular strain of *Streptomyces* that is considered to be a model representative of the genus.⁹ This is due to its genetic tractability and aptitude to produce strongly pigmented antibiotics, such as the actinorhodins **9** (blue) and prodiginines **10-11** (red), which can function as useful genetic markers (figure 1.3).^{10,11}

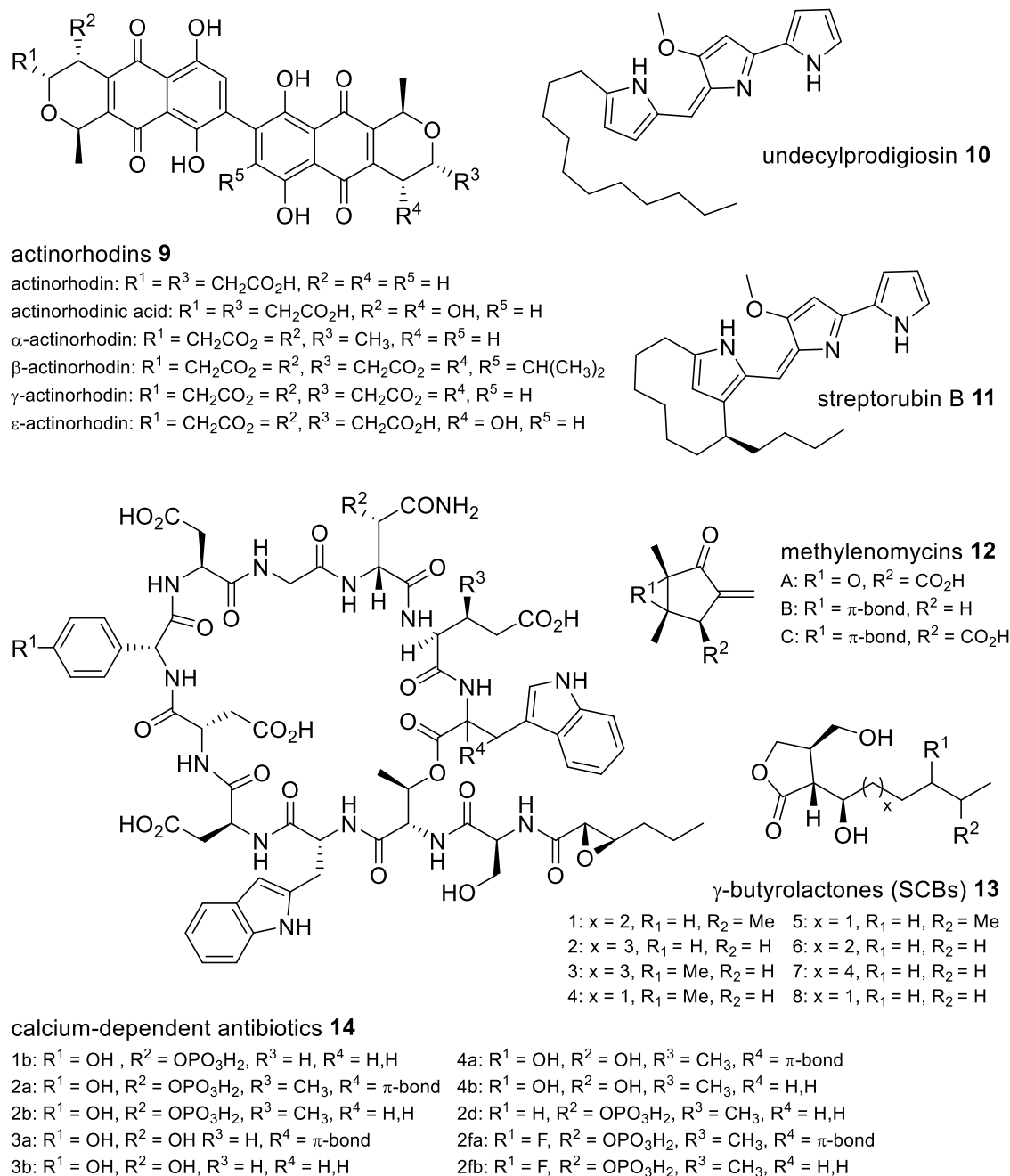


Figure 1.3: *Streptomyces coelicolor* specialised metabolites, discovered prior to the availability of the genome sequence through a classical bioassay-guided approach

In the early 2000s, the complete genome sequence of *S. coelicolor* A3(2) was published.^{7,12,13} It is comprised of an 8.7 Mb-long linear chromosome, a 356 kb linear plasmid, SCP1, and a 31 kb circular plasmid, SCP2. Before this, gene clusters for only six groups of specialised metabolites had been characterised: the actinorhodins **9**, prodiginines **10-11**, methylenomycins **12**, γ -butyrolactones **13**, calcium-dependent antibiotics **14** (figure 1.3) and grey spore pigment.^{14,15,16,17,18,19} Detailed bioinformatic analysis of the genome sequence unveiled a further 16 gene clusters, encoding proteins normally associated with complex natural product biosynthesis, whose metabolic products were unknown.²⁰ This prompted the investigation of these cryptic gene clusters, and thereafter the discovery of several novel metabolites, including germicidin C **15**, coelichelin **16** and methylenomycin furan 1 **17** (figure 1.4).^{21,22,23} In addition, a wealth of information regarding biosynthetic mechanisms was revealed; for example, unprecedented type III polyketide synthase (PKS) enzymology and atypical non-ribosomal peptide synthetase (NRPS) biochemistry in germicidin **15** and coelichelin **16** biosynthesis, respectively.^{21,22}

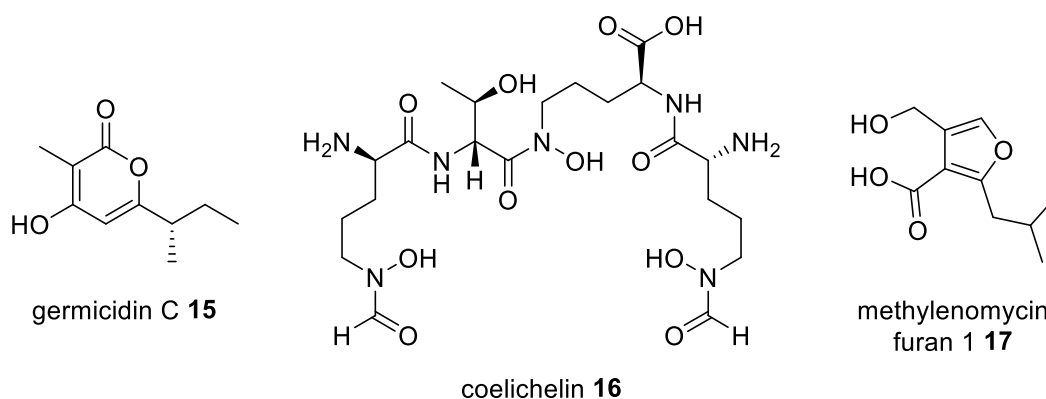


Figure 1.4: Selected *Streptomyces coelicolor* specialised metabolites, discovered through a genomics-driven approach

1.3 Discovery of coelimycin P1

Another cryptic gene cluster uncovered through the *S. coelicolor* A3(2) genome sequencing project was predicted to encode a type I modular PKS.⁷ It was thereafter named the *cpk* (*coelicolor* polyketide) biosynthetic gene cluster (figure 1.5, also see appendix 1). DNA fragments encoding acyltransferase (AT) and ketosynthase (KS) domains within the PKS were first identified in the mid-1990s by Kuczek and co-workers,

who realised that they were not associated with the production of any known specialised metabolites.^{24,25}

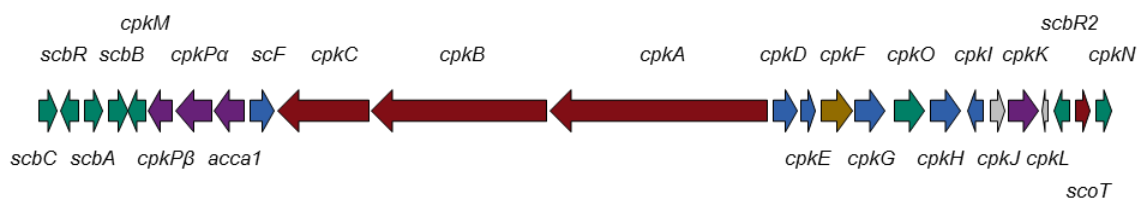


Figure 1.5: The organisation of the 59 kb-long *cpk* gene cluster directing coelimycin biosynthesis in *Streptomyces coelicolor* M145. The proposed functions of the proteins encoded by the genes are as follows: Purple: precursor supply; red: PKS component; blue: post-PKS tailoring; green: regulatory; gold: export; and grey: hypothetical.

In 2010, Gottelt and Pawlik independently reported the polyketide product to be a diffusible yellow pigment, although no structural information was provided.^{26,27} Gottelt also observed that production of the pigment was associated with antibiotic activity.²⁶

In 2012, Gomez-Escribano and coworkers attempted to increase the metabolic flux through the *cpk* pathway using two genetic engineering strategies. Firstly, gene clusters responsible for the production of three classes of antibiotics (the actinorhodins **9**, prodiginines **10-11** and calcium-dependent antibiotics **14**) were deleted from *S. coelicolor* M145, a derivative of A3(2) lacking its two plasmids, to reduce competition for biosynthetic precursors. Secondly, spontaneous mutations in *rpoB*, encoding the β -subunit of RNA polymerase (RNAP), were selected for with rifampicin. These mutations are known to trigger antibiotic production, presumably because the resulting RNAP mutant has an altered conformation resembling that of a guanosine tetraphosphate (ppGpp)-bound state, which promotes the transcription of specialised metabolite gene clusters.^{28,29} One rifampicin-resistant strain, named *Streptomyces coelicolor* M1157, produced large quantities of the yellow pigment.

Comparative metabolic profiling of *S. coelicolor* M1157 with a functionally equivalent strain harbouring a different *rpoB* mutation (*S. coelicolor* M1152) resulted in the identification of a compound with an absorbance maximum at 360 nm and a molecular formula of $C_{17}H_{20}N_2O_4S$. Structural elucidation of the compound by NMR revealed that it was an alkaloid containing an unprecedented 1,5-oxathiocane ring (figure 1.6).³⁰

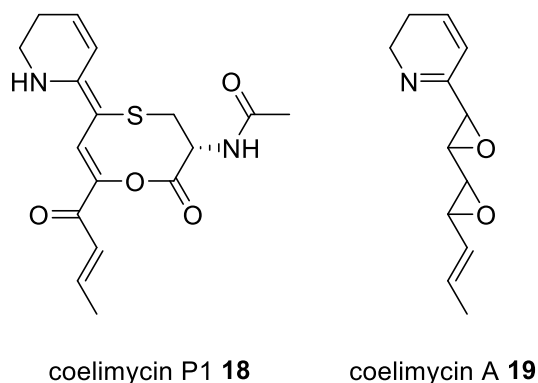


Figure 1.6: Structures of the characterised shunt metabolite, coelimycin P1 **18**, and the putative true metabolite, coelimycin A **19**, of the *cpk* biosynthetic pathway

Interestingly, while the culture supernatants possessed antibiotic activity, the purified compound, named coelimycin P1 **18**, did not. It was therefore suggested that coelimycin P1 **18** may be a shunt metabolite of the pathway, derived from reaction of the putative antibacterial metabolic product, coelimycin A **19** (figure 1.6), with *N*-acetylcysteine in the culture medium.

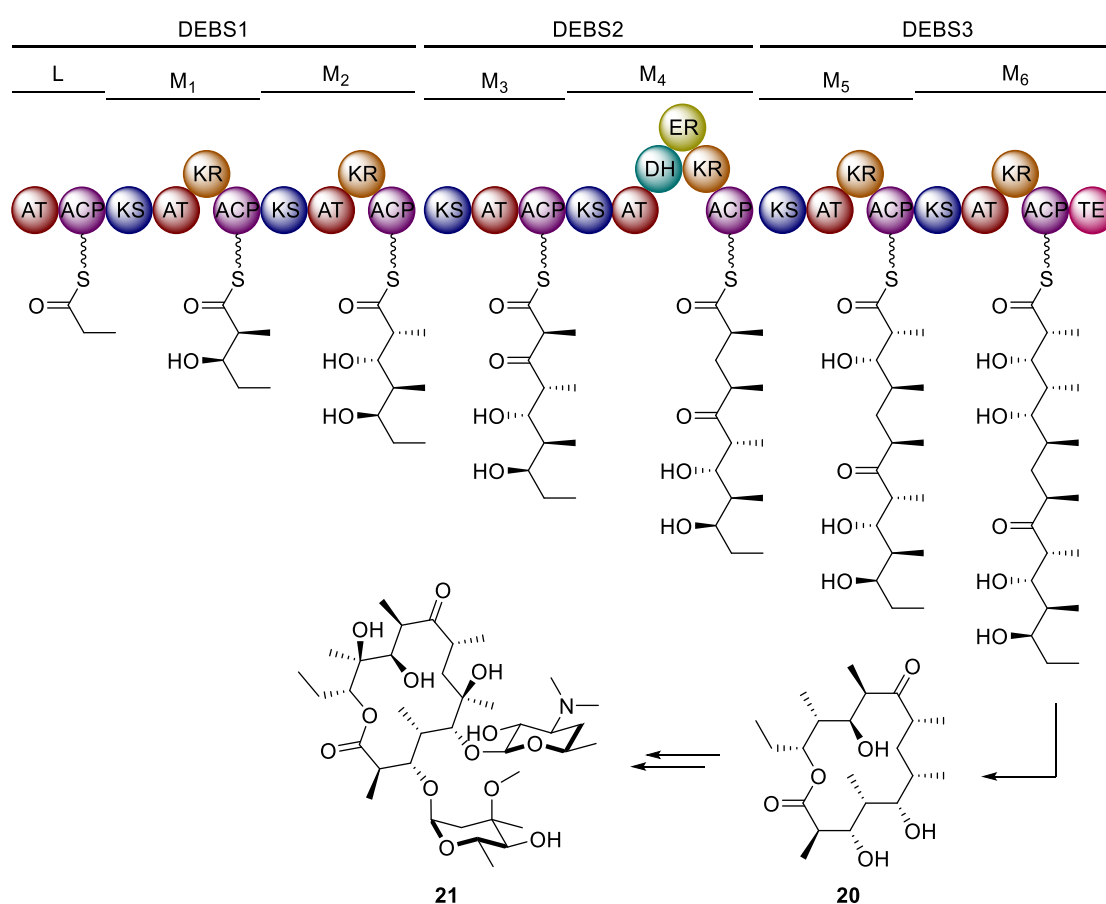
Notably, the yellow pigment observed by Gottelt and coworkers²⁶ was overproduced in glutamate-supplemented medium, and had a different molecular formula ($C_{17}H_{22}N_2O_6$) corresponding to a putative glutamate adduct of coelimycin A **19**. It was later named coelimycin P2.³⁰

1.3.1 Type I modular PKSs

Type I modular PKS assembly lines catalyse a series of condensations and subsequent modifications of acyl-CoA thioesters to form diverse polyketide scaffolds.^{31,32} Type I PKSs differ from their type II and III counterparts, in that they are multifunctional polypeptide complexes harbouring different catalytic domains. Type II PKSs are dissociable complexes containing discrete, monofunctional, iteratively-acting enzymes, while type III PKSs are smaller, homodimeric enzymes with a single active site and an acyl carrier protein (ACP)-independent mechanism.^{33,34}

Type I PKSs can be further subdivided depending on whether they act in an iterative or modular fashion. Type I iterative PKSs have their catalytic domains housed within a single module, meaning that each domain is used repeatedly during polyketide chain

assembly. In contrast, modular PKSs have their catalytic domains organised within several sequential modules, with each module usually associated with a single round of chain elongation and β -carbon processing.³⁵ In these systems, accurate predictions of the polyketide product can be made, based on knowledge of the module and domain order. However, non-canonical examples of modular PKSs have also been reported, in which modules can be bypassed (e.g. in mupirocin biosynthesis)³⁶ or used more than once (e.g. in borrelidin biosynthesis)³⁷. These events have been termed ‘module skipping’ and ‘module stuttering’, respectively, and result in a loss of the co-linearity normally associated with these enzymes.



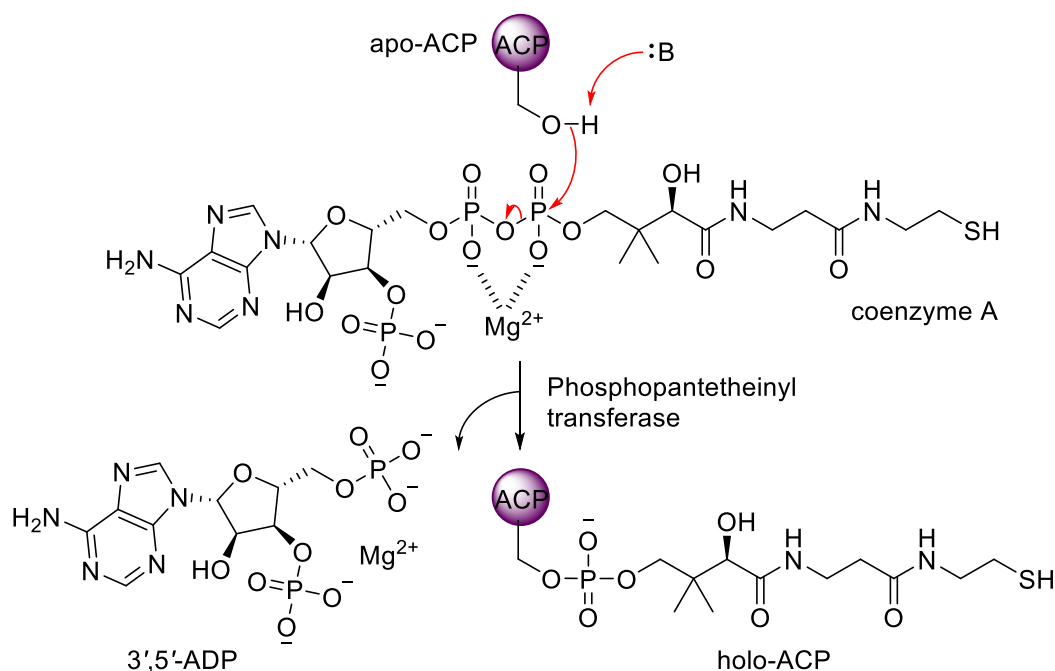
Scheme 1.1: The organisation of the DEBS system, responsible for the assembly of 6-deoxyerythronolide B **20**. Subsequent modification by post-PKS tailoring enzymes affords erythromycin A **21**.

The majority of type I modular PKSs contain three minimal catalytic domains within their chain extension modules: the acyl carrier protein (ACP), acyltransferase (AT), and ketosynthase (KS), which result in the incorporation of one acyl-CoA thioester building block into the nascent polyketide chain. Some chain extension modules also contain

accessory domains, such as the ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER), which are involved in β -carbon processing. A terminal thioesterase (TE) is usually employed for the release of the full-length polyketide, *via* hydrolysis or macrolactonisation.³⁸ The 6-deoxyerythronolide B synthase (DEBS) is the prototypical example of a type I modular PKS, comprised of three proteins and seven modules. In this system, one propionyl-CoA starter unit and six methylmalonyl-CoA extender units are condensed, processed and released to yield 6-deoxyerythronolide B **20**, the precursor to the antibiotic erythromycin A **21** (scheme 1.1).³⁹

1.3.1.1 Minimal catalytic domains

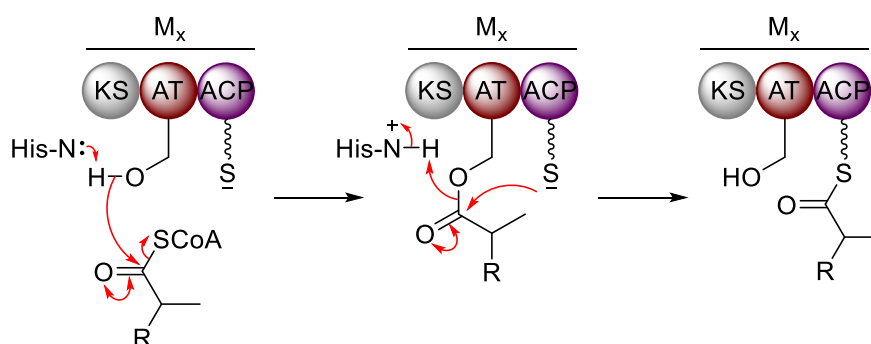
The ACP domain is a small, four-helical bundle, located at the C-terminus of each module. Before polyketide chain assembly can occur, a conserved serine residue within the ACP must first be post-translationally modified by reaction with coenzyme A, catalysed by a phosphopantetheinyl transferase (scheme 1.2). This converts the inactive apo-form of ACP to its active holo-form, in which the newly-attached, flexible, phosphopantetheinyl arm can shuttle growing polyketide chains between active sites of the PKS, through the formation of thioester bonds.⁴⁰



Scheme 1.2: Mechanism of reaction of apo-ACP with coenzyme A to form holo-ACP. B = general base.

The AT domain is responsible for catalysing the transfer of particular starter and extender units to the downstream ACPs with high specificity, and is thus often referred to as the gatekeeper domain.⁴¹ Typical starter units include acetyl- and propionyl-CoA, while extender units comprise their carboxylated analogues: malonyl- and methylmalonyl-CoA, respectively (see section 1.3.2). However, many additional starter and extender units have been identified, including branched, hydroxylated, chlorinated, cyclic and aromatic derivatives, in addition to amino acids, as part of PKS/NRPS hybrid systems.^{42,43}

The transacylation reaction occurs by a ping-pong mechanism, in which the catalytic serine, activated by a neighbouring histidine residue, first attacks the thioester of a CoA-derivatised building block (e.g. propionyl- or methylmalonyl-CoA).⁴⁴ The AT-tethered intermediate is then transferred to the phosphopantetheinyl arm of a downstream holo-ACP (scheme 1.3), prior to being processed by the subsequent module.⁴¹

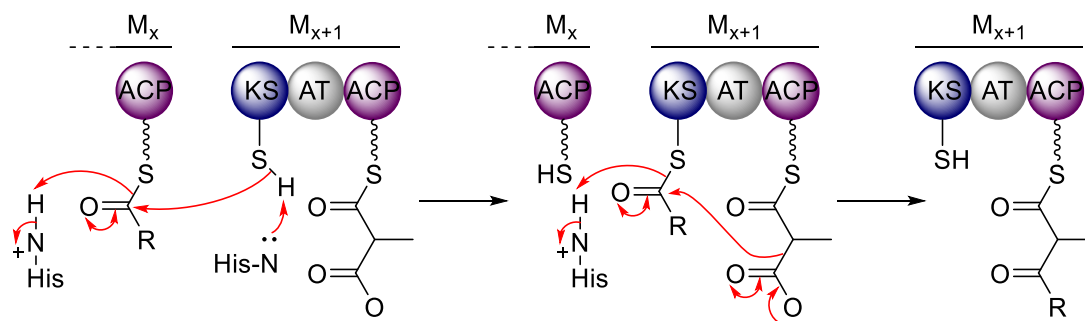


Scheme 1.3: Mechanism of reaction of an AT domain with a starter unit (e.g. propionyl-CoA, R = H) or extender unit (e.g. methylmalonyl-CoA, R = CO₂⁻), and the subsequent transfer to a downstream holo-ACP. The catalytic histidine residue is indicated. M_x = module *x*.

While most PKS assembly lines contain AT domains that are integrated within modules (*cis*-AT PKSs), other AT domains are encoded by discrete genes and expressed as stand-alone polypeptides (*trans*-AT PKSs). In these systems, the AT is capable of acting iteratively, and transferring extender units to multiple ACP domains within the PKS.⁴⁵

The KS domain is responsible for catalysing intermodular translocation and extension of the polyketide chain, through the formation of carbon-carbon bonds. Initially, the conserved nucleophilic cysteine residue attacks the upstream ACP-bound thioester to form a KS-bound thioester. The downstream ACP, loaded with an extender unit, then docks to the acylated KS, and a decarboxylative Claisen condensation ensues, resulting

in the elongation of the polyketide backbone by two carbon units (scheme 1.4). Two conserved histidine residues have recently been reported to mediate these reactions through general acid and base catalysis.⁴⁶

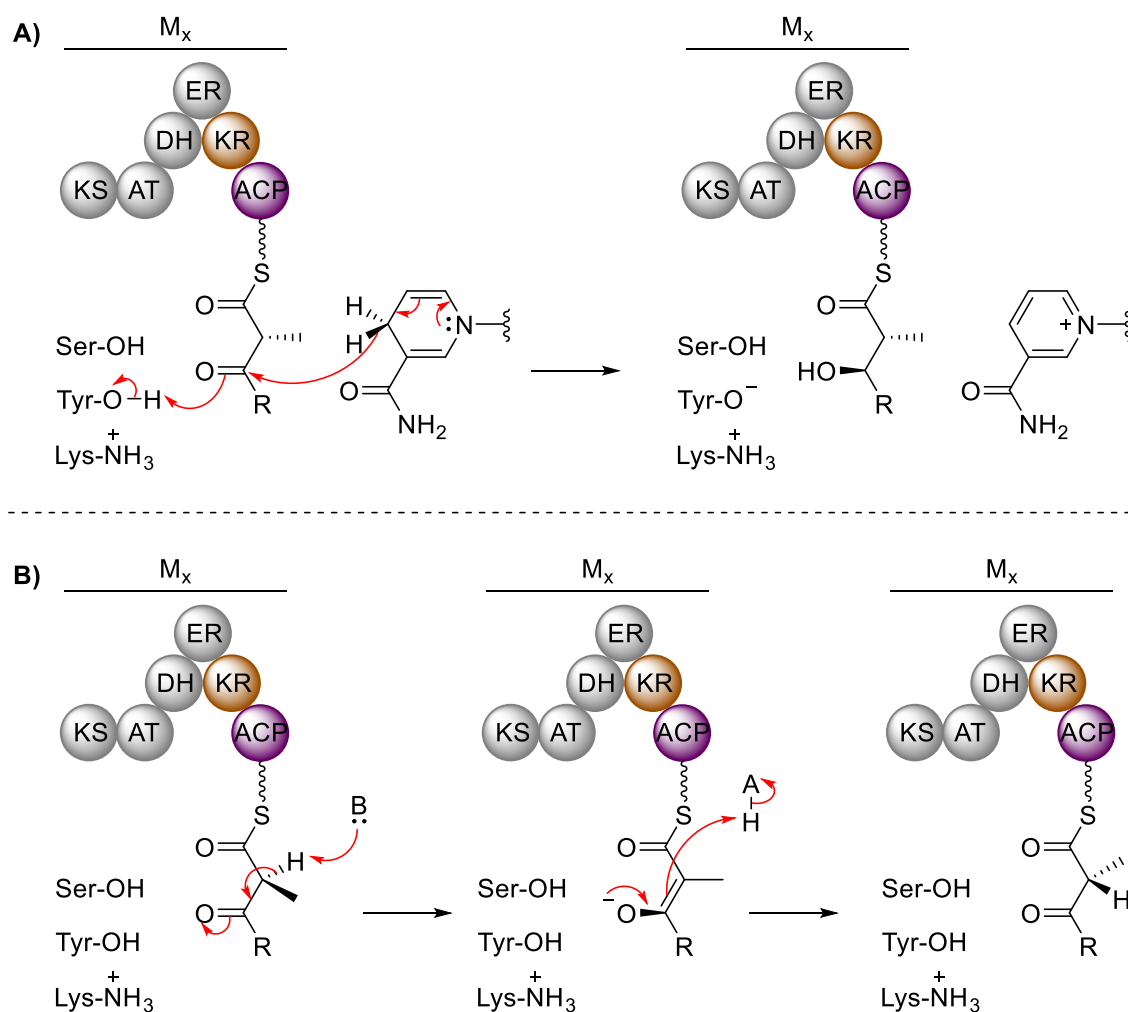


Scheme 1.4: Mechanism of reaction of a KS domain with an upstream acyl-ACP, followed by a decarboxylative Claisen condensation with a methylmalonyl unit tethered to a downstream ACP. Catalytic histidine residues are indicated. R = polyketide chain; M_x = module x .

Sometimes, modular PKSs have a modified KS domain in the loading module, which contains a mutation of the active site cysteine to a glutamine. In these loading modules, the AT domain recruits a malonyl-CoA, rather than an acetyl-CoA, and the so-called KS^Q solely catalyses its decarboxylation.⁴⁷

1.3.1.2 Accessory catalytic domains

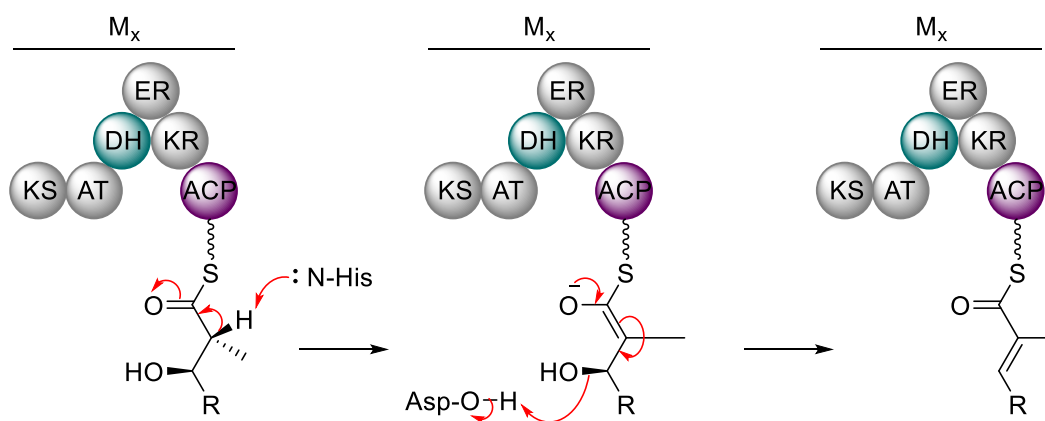
A number of accessory domains can then act to modify the nascent polyketide chain. The KR domain is responsible for reducing the β -keto group of the ACP-bound thioester to a β -hydroxyl group, and controlling the configuration at both the α - and β -stereogenic centres. During catalysis, a conserved tyrosine and serine residue activate the β -keto group to receive the *pro*-(4*S*) hydride of a nicotinamide adenine dinucleotide phosphate (NADPH) cofactor. The oxygen anion then abstracts a proton from the same catalytic tyrosine residue, which is stabilised by a neighbouring lysine (scheme 1.5a). Based on their activity and stereospecificity, KR domains can be labelled as one of three types: A- and B-type KRs produce a β -hydroxyl group with (*S*)- or (*R*)-stereochemistry, respectively (according to polyketide convention in which the α -substituent is given higher priority than the γ -substituent), while C-type KRs are reductase-inactive. In addition, KRs may be selective for substrates with either an (*S*)- or (*R*)-configured α -substituent, or may catalyse an epimerisation at that position (scheme 1.5b).^{48,40}



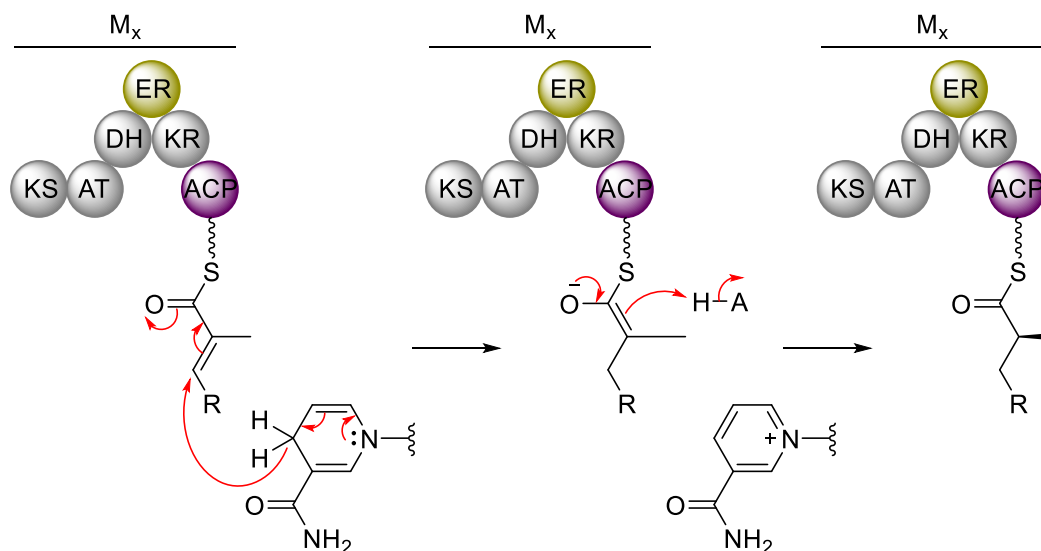
Scheme 1.5: Mechanisms of (A) reduction; and (B) epimerisation of a β -ketothioester, catalysed by a KR domain. The serine, tyrosine and lysine residues within the catalytic triad are indicated. R = polyketide chain; A = general acid; B = general base; M_x = module x .

After reduction of the β -ketothioester to a β -hydroxythioester, a DH domain can catalyse a dehydration reaction to generate a double bond between the α - and β -carbons. A catalytic histidine and aspartate have been proposed to abstract the α -proton, and protonate the β -hydroxyl group, respectively, to promote the *syn* elimination of water (scheme 1.6). This converts (*R*)- or (*S*)-configured β -hydroxyl groups to the corresponding *trans*- or *cis*- α,β -unsaturated thioesters.^{40,49}

Following the dehydration reaction, an ER domain can reduce the α,β -unsaturated double bond, setting the stereochemistry of α -substituents. During catalysis, a hydride is transferred from an NADPH cofactor and a proton is delivered to the resulting carbanion (scheme 1.7). Interestingly, the transfer of a *pro*-(4*S*) or *pro*-(4*R*) hydride appears to be organism-dependent.^{50,51} Active site residues involved in catalysis are as of yet unclear.



Scheme 1.6: Mechanism of dehydration of a β -hydroxythioester, catalysed by a DH domain. Catalytic histidine and aspartate residues are indicated. R = polyketide chain; M_x = module x .

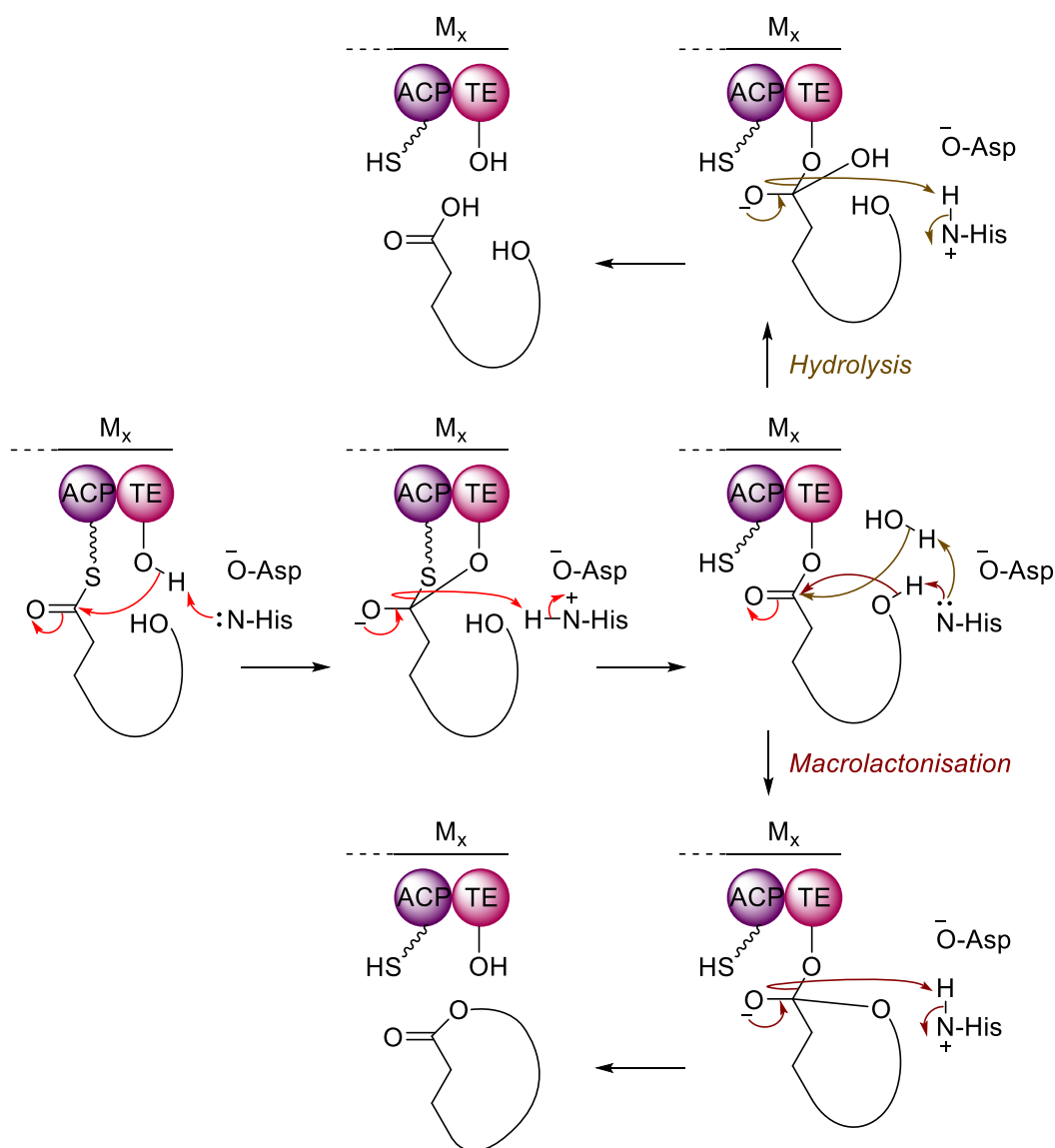


Scheme 1.7: Mechanism of reduction of an α,β -unsaturated thioester, catalysed by an ER domain. R = polyketide chain; A = general acid; M_x = module x .

Occasionally, *O*- and *C*-methyltransferase (MT) domains also reside within modules of type I PKSs.^{52,53,54,55} Rather than acting as “reductive” domains, they yield β -methoxy or α -alkylated products.

1.3.1.3 Polyketide chain offloading

Following the assembly of a full-length polyketide chain, the product is offloaded, usually by a TE domain. This can occur *via* hydrolysis or intramolecular cyclisation, to yield a linear carboxylic acid or macrolactone, respectively (scheme 1.8).



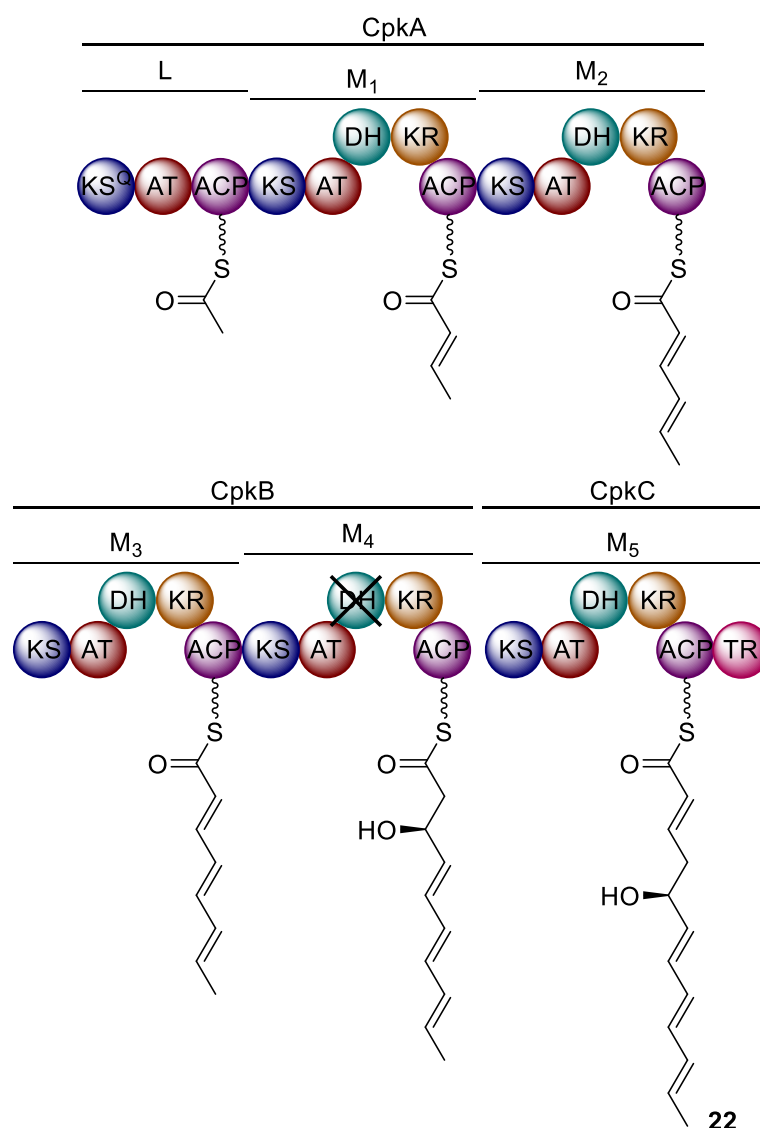
Scheme 1.8: Mechanisms of chain release of a fully assembled polyketide chain by a TE domain, to yield a carboxylic acid (top) or macrolactone (bottom). The serine, histidine and aspartate residues within the catalytic triad are indicated.

In both cases, a conserved histidine residue first deprotonates the nucleophilic serine, which then attacks the ACP-bound thioester. The tetrahedral intermediate then collapses with protonation of the phosphopantetheinyl leaving group by the same histidine residue, to yield a TE-bound ester. Subsequently, either a water molecule, or a hydroxyl moiety within the polyketide chain, can attack, through a similar addition-elimination mechanism.⁵⁶ Other mechanisms of chain release are also prevalent, which result in the generation of different structural classes of specialised metabolites. For example, standalone *N*-acyltransferase (NAT), α -oxamine synthase (OAS) and KS domains can catalyse polyketide chain offloading to yield macrolactams, prodiginines and tetronates, respectively.⁵⁷

Commonly, PKS and NRPS gene clusters also encode a discrete type II thioesterase, which has a fundamental ‘editing’ role; incorrect starter units and non-reactive ACP-bound acyl residues which would otherwise block the assembly line can be hydrolytically released. The responsible enzyme encoded within the *cpk* cluster is ScoT, which has been demonstrated to hydrolyse acetyl, propionyl and butyryl units.⁵⁸ More recently, ScoT has been shown to be essential for coelimycin P2 production.⁵⁹

1.3.1.4 Predicted module and domain organisation of the coelimycin PKS

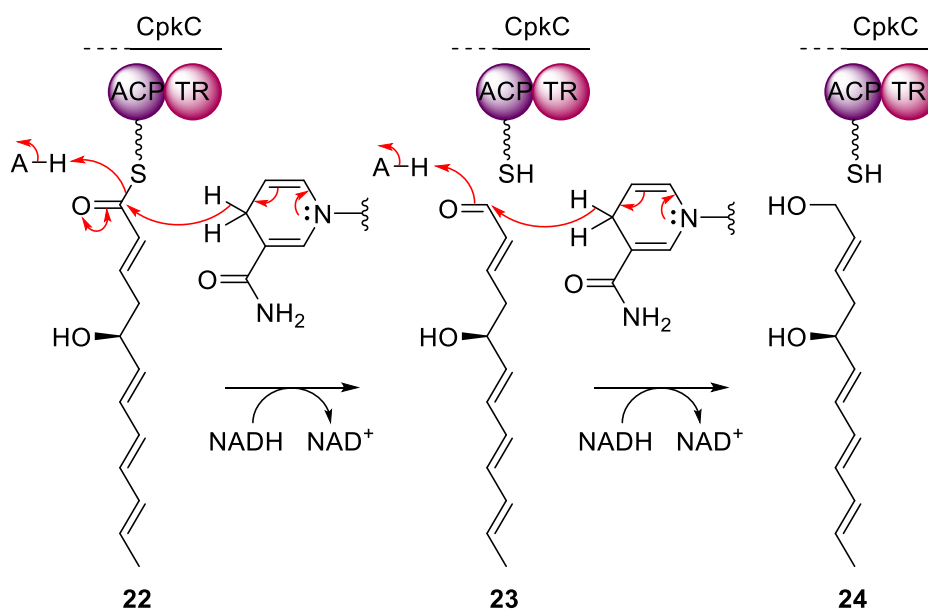
The PKS involved in coelimycin biosynthesis is comprised of three proteins (CpkA, CpkB and CpkC) and six modules (scheme 1.9).



Scheme 1.9: The PKS module and domain organisation of CpkABC and the proposed structures of ACP-bound intermediates. The cross indicates that the DH domain in module 4 is likely inactive.

All chain extension modules are predicted to contain a KR and DH domain, in addition to the minimal catalytic domains. In contrast to the DEBS PKS (scheme 1.1), not all domains are active; the DH domain in module 4 is likely non-functional. Furthermore, no ER domains appear present, suggesting that the nascent polyketide chain is not fully reduced at any stage. One further difference is that all AT domains are predicted to recruit malonyl-CoA, rather than its methylated analogue. Assuming that the number of modules corresponds to the number of malonyl-CoA units used during chain assembly, a 5-hydroxy-2,6,8,10-dodecatetraenoyl thioester **22** is likely assembled (scheme 1.9).

Interestingly, the final PKS module contains a thioester reductase (TR) domain, rather than a TE domain, which has been proposed to catalyse reductive cleavage of the full-length polyketide chain. To examine this hypothesis, purified recombinant CpkC-TR was incubated with octanoyl-CoA as a mimic of the proposed ACP-bound thioester **22**, and either NADH or NADPH as a cofactor (Ruby Awodi, University of Warwick). UV-Vis and GC-MS analyses revealed that CpkC-TR is capable of catalysing two NADH-mediated two-electron reductions of octanoyl-CoA to octanol.⁵⁷ This suggests that CpkC-TR can fully reduce thioester **22** to alcohol **24** during coelimycin biosynthesis (scheme 1.10).



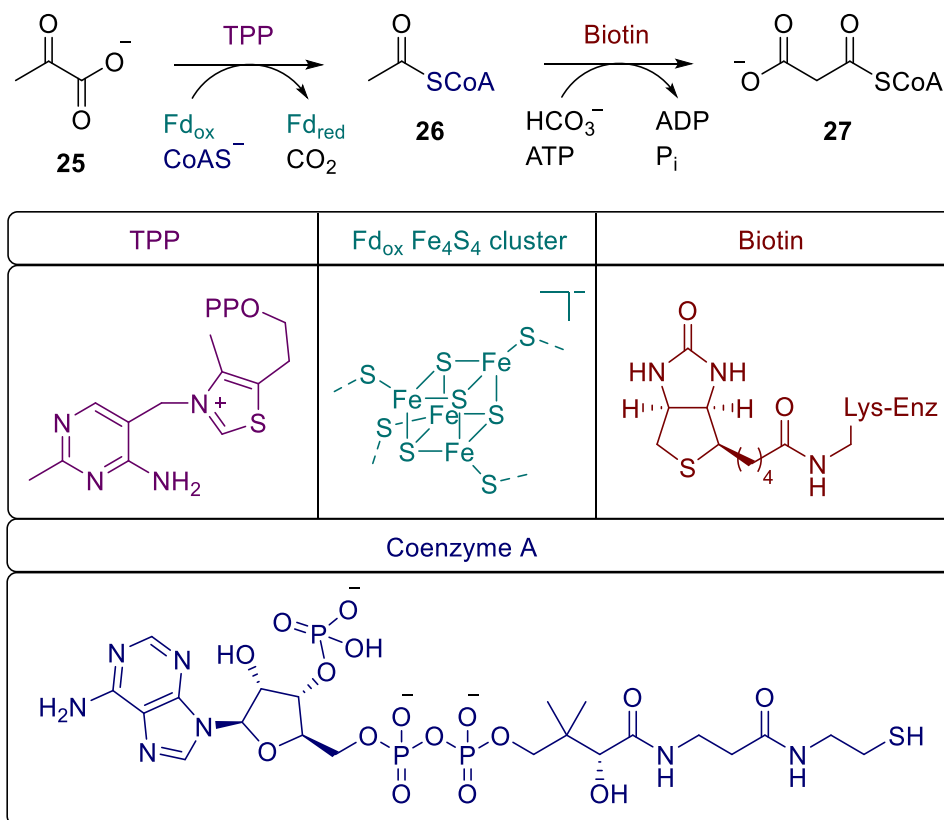
Scheme 1.10: Mechanism of NADH-mediated reductive chain release of **22** by CpkC-TR. A = general acid.

Notably, while several instances of TR-mediated reductive chain release have been reported within NRPSs, hybrid NRPS-PKSs, fatty acid synthases (FASs) and type I iterative PKSs,^{60,61,62,63} this is the first example within a type I modular PKS.³⁰

1.3.2 Precursor supply genes

Four genes within the *cpk* cluster (*cpkPa*, *cpkPβ*, *accA1* and *cpkK*) are proposed to encode enzymes which generate polyketide precursors. Typical precursors include acetyl- and propionyl-CoA, and their carboxylated equivalents, malonyl- and methylmalonyl-CoA. Six units of malonyl-CoA are predicted to be utilised by CpkA, CpkB and CpkC during the assembly of polyketide thioester **22**.

cpkPa and *cpkPβ* show sequence similarity to the α - and β -subunits of α -ketoacid-dependent ferredoxin reductases. These are thiamine pyrophosphate (TPP)-, coenzyme A- and ferredoxin (Fd)-dependent enzymes which catalyse the oxidative decarboxylation of α -keto acids to acyl-CoAs (e.g. pyruvate **25** to acetyl-CoA **26**, scheme 1.11).⁶⁴



Scheme 1.11: The conversion of pyruvate **25** to malonyl-CoA **27**, catalysed by a pyruvate-dependent ferredoxin reductase and an acetyl-CoA carboxylase. Relevant cofactors are highlighted.

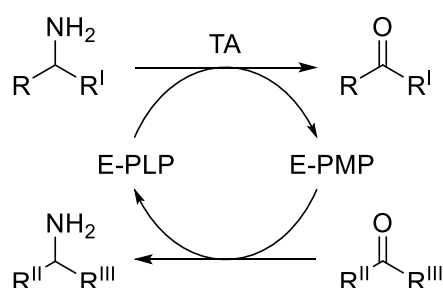
accA1 and *cpkK* show sequence similarity to the α - and β -subunits of acyl-CoA carboxylases. These are biotin-dependent enzymes which catalyse a two-step reaction. Firstly, bicarbonate attacks adenosine triphosphate (ATP) to form CO₂, which then reacts with biotin to afford carboxybiotin. Secondly, the carboxyl group is transferred from carboxybiotin to an acyl-CoA to yield a carboxylated derivative (e.g. acetyl-CoA **26** to malonyl-CoA **27**, scheme 1.11).⁶⁵

1.3.3 Post-PKS tailoring genes

Following the assembly and release of polyketide thioester **22**, a series of transformations are necessary for its conversion to coelimycin P1 **18**. The responsible post-PKS tailoring enzymes are likely encoded by the genes *cpkG*, *cpkD*, *scF*, *cpkH*, *cpkI* and *cpkE*.

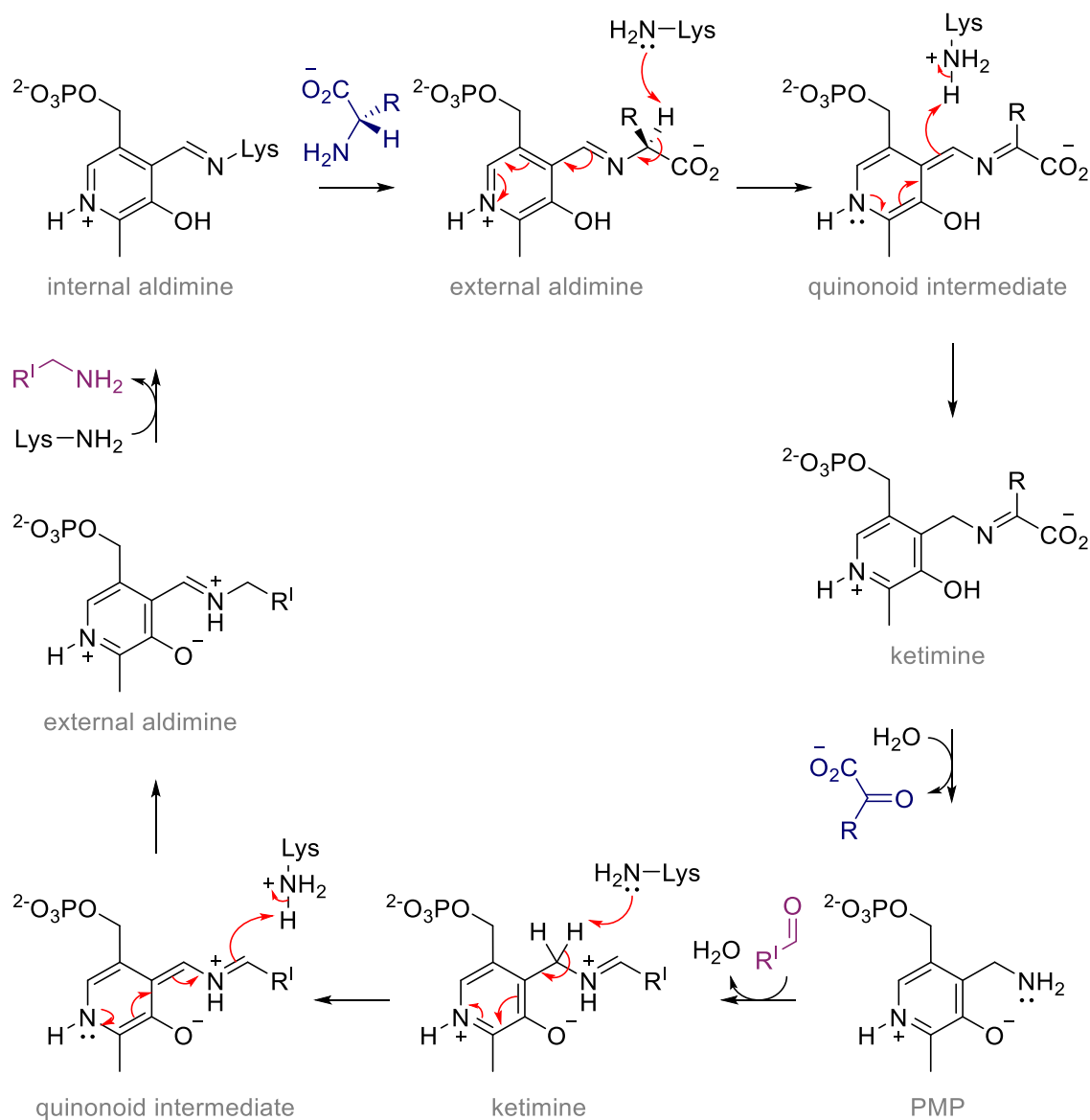
1.3.3.1 ω -Transaminases

cpkG has been demonstrated to encode a pyridoxal 5'-phosphate (PLP)-dependent enzyme with transaminase activity towards selected aldehydes and amines (Ruby Awodi, University of Warwick). Transaminases (TAs), also known as aminotransferases, are enzymes that mediate the transfer of an amino group between a suitable amino donor and acceptor. TAs that catalyse the transfer of amino groups at the α -carbon position with respect to a carboxylate moiety are given the prefix ' α ', while TAs that perform the reaction at a β -, γ - or more distal position, or on substrates lacking a carboxylate moiety, are labelled as ' ω '. Thus, in ω -TAs, at least one of the two substrates cannot be an α -amino acid or α -keto acid (scheme 1.12).⁶⁶



Scheme 1.12: Catalytic cycle of a TA. The amino donor (top left) transfers an amino group to enzyme-bound pyridoxal-5'-phosphate (E-PLP), and the newly-formed enzyme-bound pyridoxamine-5'-phosphate (E-PMP) transfers it on to the amino acceptor (bottom right). For α -TAs, R^I = R^{III} = CO₂H. For ω -TAs, if R^I = CO₂H, then R^{III} \neq CO₂H.

TAs utilise PLP, a derivative of vitamin B₆, as a cofactor, which acts as an effective molecular shuttle for protons, electrons and ammonia. The reaction catalysed by TAs can be divided into two halves: i) oxidative deamination of the amino donor and formation of pyridoxamine 5'-phosphate (PMP); and ii) reductive amination of the amino acceptor and reformation of PLP.⁶⁷ This is often referred to as a 'ping-pong bi-bi' mechanism, as the second substrate cannot bind until the first product leaves.⁶⁸



Scheme 1.13: Reaction mechanism of a TA. In the first half of the reaction, an amino group is transferred from an α -amino acid to enzyme-bound PLP, to form an α -keto acid and PMP. In the second half of the reaction, the amino group is transferred from PMP to an aldehyde, to form a primary amine and enzyme-bound PLP. Note: reactions are reversible, but arrows are unidirectional for clarity.

In the catalytic cycle, PLP is initially tethered to an ϵ -amino group of a conserved lysine residue within the active site, through an internal aldimine linkage. The substrate amino donor (e.g. an α -amino acid) then displaces the lysine residue in a transaldimination reaction, forming an external aldimine. Deprotonation of the α -hydrogen by the same lysine residue affords a quinonoid intermediate, which is reprotonated near the pyridine ring to form a ketimine intermediate. Subsequent hydrolysis releases the corresponding carbonyl compound (e.g. an α -keto acid) and generates PMP (scheme 1.13).

In the second half of the reaction, a reverse transamination occurs, in which the free amino group of PMP first attacks the amino acceptor (e.g. an aldehyde) to form a ketimine intermediate. Analogous isomerisations afford the quinonoid and external aldimine, which is subsequently displaced by the lysine residue, releasing the corresponding amine compound (e.g. a primary amine). At the same time, the PLP-form of the enzyme, or the internal aldimine, is regenerated (scheme 1.13).^{69,70}

Based on their 3D structures, ω -TAs can be further subdivided into fold type I or fold type IV classes. Members of these classes are typically selective for (*S*)- and (*R*)-configured substrates, respectively.⁷¹ The highly stereoselective nature of ω -TAs makes them promising biocatalysts for the generation of chiral amines and unnatural amino acids, which are important synthons in pharmaceutical, fine chemical and agrochemical industries.^{72,73}

1.3.3.2 Flavin-dependent oxidases, dehydrogenases and monooxygenases

cpkD, *scF* and *cpkH* are proposed to encode flavoenzymes. These are proteins that bind and utilise a form of flavin, usually flavin adenine dinucleotide (FAD), or flavin adenine mononucleotide (FMN). These cofactors are derivatives of riboflavin, or vitamin B₂, and comprise a tricyclic isoalloxazine moiety attached to a ribityl chain at the N₁₀ position (figure 1.7). Flavins are fundamental in many aspects of biology, including cell and drug metabolism, immune defence, DNA repair, cell signalling, protein folding and apoptosis.^{74,75}

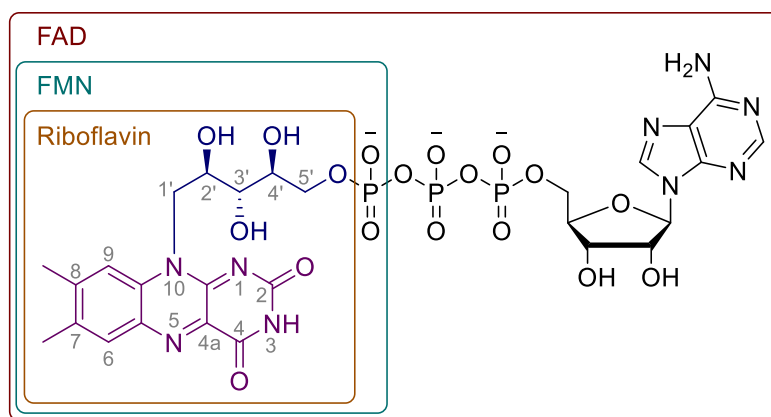


Figure 1.7: The structures of riboflavin, FMN and FAD. The isoalloxazine ring and ribityl side chain are highlighted in purple and blue, respectively.

Flavins are normally bound very tightly within the protein, and in approximately 11 % of cases, binding is covalent, with attachment of cysteine, histidine or tyrosine residues to the C-6 atom of the isoalloxazine ring and/or the 8 α -methyl group.⁷⁶ Flavins can exist in three oxidation states: oxidised (quinone), 1-electron reduced (semiquinone) or 2-electron reduced (hydroquinone), in both protonated and unprotonated forms (figure 1.8). As such, the cofactors are able to perform both 1- and 2-electron transfers.

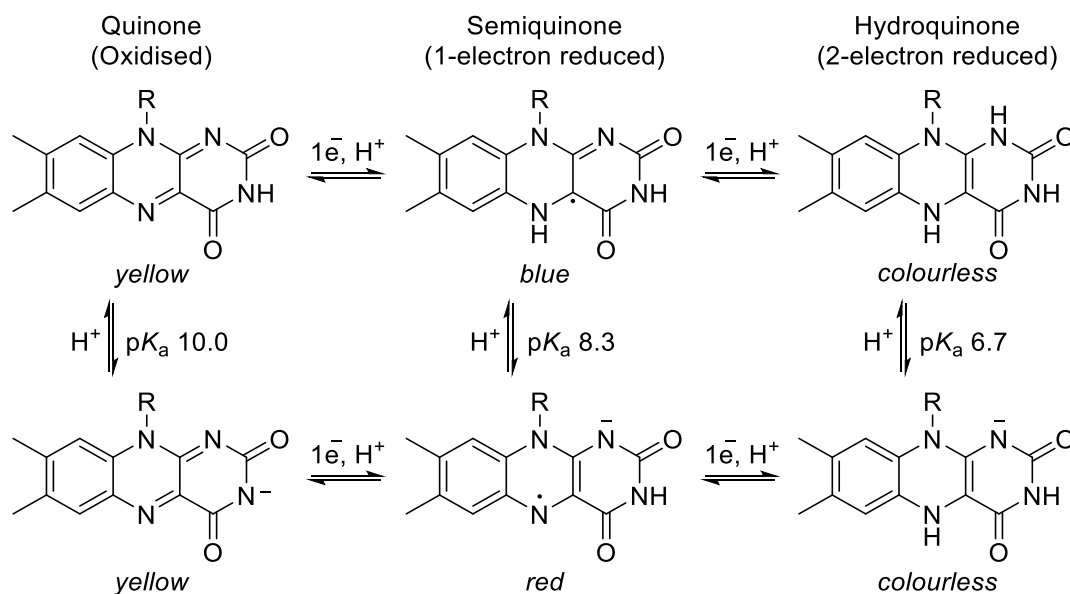
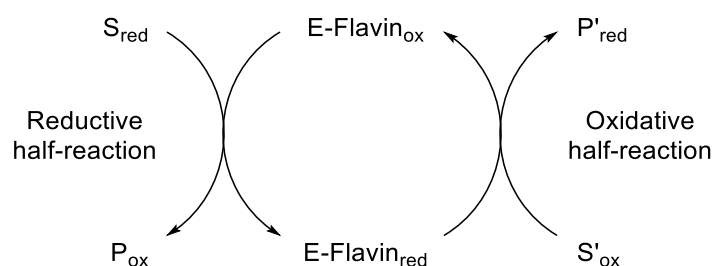


Figure 1.8: The different oxidation states of flavin, in protonated (top) and unprotonated (bottom) forms

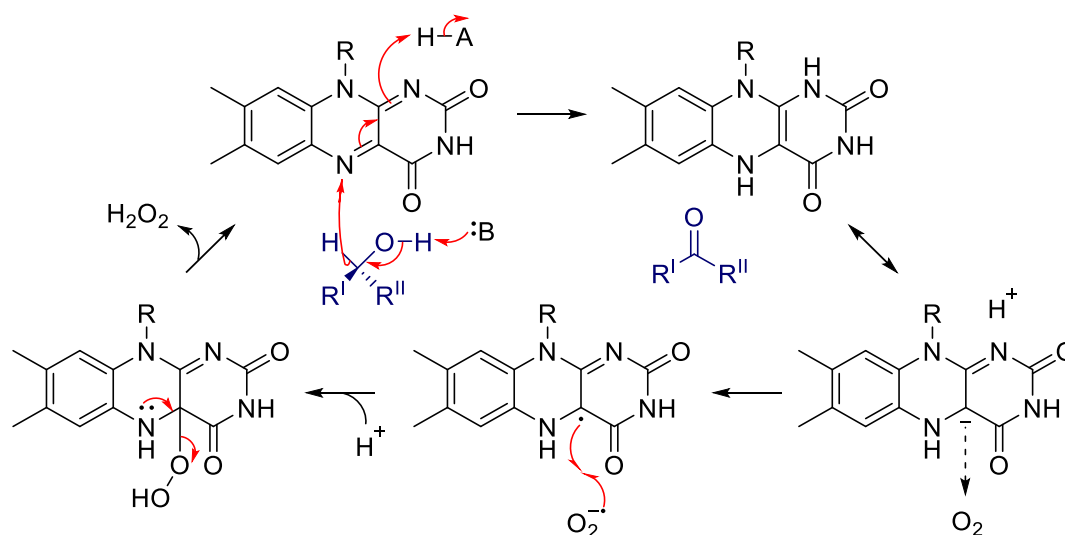
The reactions catalysed by flavoenzymes can be divided into reductive and oxidative half reactions. In the first half, a substrate reduces the oxidised cofactor, while in the second half, a separate substrate re-oxidises the reduced cofactor (scheme 1.14). In oxidases, this

second substrate is molecular oxygen, which is converted to hydrogen peroxide. In dehydrogenases, an alternative electron acceptor is employed.⁷⁷



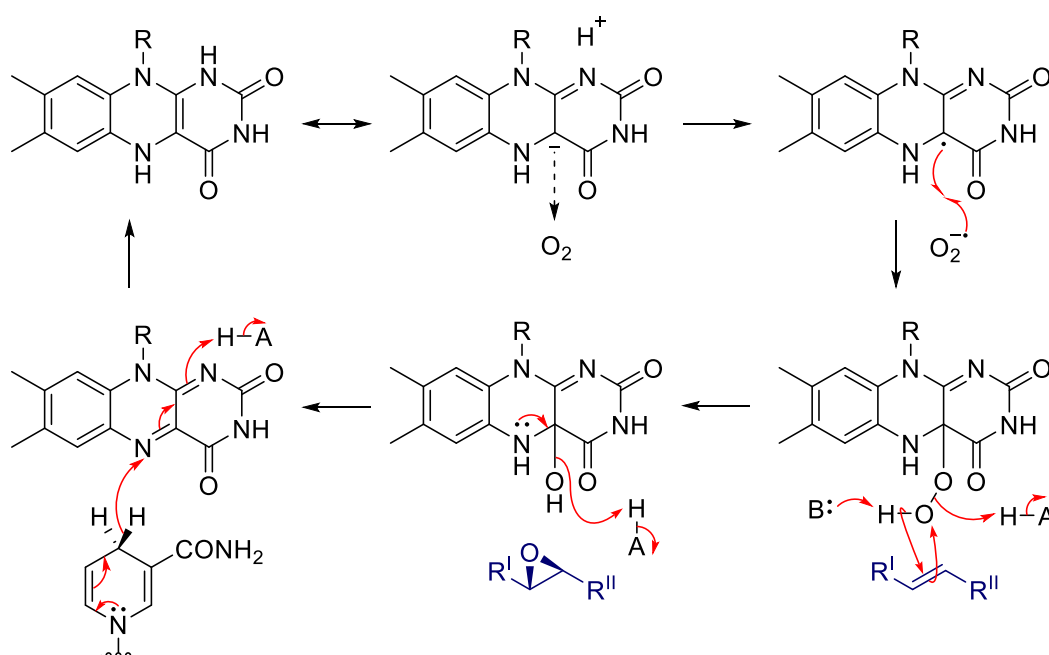
Scheme 1.14: The catalytic cycle of flavoenzymes

Flavin-dependent oxidases and dehydrogenases can catalyse the oxidation of α,β -carbon bonds in carboxylic acids or thioesters, and carbon-heteroatom bonds in substrates including alcohols, amines, amino acids and hydroxyacids.^{78,79} Mechanisms of oxidation of these substrates have been debated for several years, with proposals including: i) direct hydride transfer; ii) nucleophilic attack of a substrate at the C-4a position; ii) nucleophilic attack of a carbanion at the N-5 position; and iv) single electron transfer. However, it is generally accepted that the oxidation of alcohols and amines proceeds *via* a direct hydride transfer (scheme 1.15).⁸⁰



Scheme 1.15: The oxidation of an alcohol by a flavin-dependent oxidase *via* a direct hydride transfer. The oxidised-form of the cofactor is subsequently regenerated by molecular oxygen (or an alternative electron acceptor in the case of a dehydrogenase). A = general acid; B = general base.

Flavin-dependent monooxygenases can catalyse a series of oxygenations, including hydroxylations, epoxidations, Baeyer-Villiger oxidations, halogenations and *S*- or *N*-oxidations. These enzymes transfer a single atom of molecular oxygen into the substrate, while the other atom is reduced to water.⁸¹ In the catalytic cycle, flavin is first reduced to its hydroquinone form by nicotinamide adenine dinucleotide (NADH) or its 2'-phosphorylated analogue, NADPH (see section 1.3.3.3). Reaction with molecular oxygen then affords a C(4a)-hydroperoxy species, which donates an oxygen atom to the substrate. Elimination of water from the resulting C(4a)-hydroxyflavin adduct allows for the regeneration of oxidised flavin (scheme 1.16).⁸²



Scheme 1.16: The epoxidation of an alkene by a flavin-dependent monooxygenase. The C(4a)-hydroperoxy species is formed by reaction of reduced flavin with molecular oxygen. Nucleophilic attack of an alkene at the electrophilic oxygen yields the epoxide product, and a C(4a)-hydroxyflavin adduct which eliminates water to form oxidised flavin. The reduced cofactor is subsequently regenerated by direct hydride transfer from NAD(P)H. A = general acid; B = general base.

Flavin-dependent monooxygenases can be categorised into eight groups, according to their fold, function, key sequence motifs and the electron donor. Groups A and B employ NAD(P)H as an external electron donor, while groups C–F require a flavin reductase partner protein instead. Groups G and H are mechanistically dissimilar, and directly utilise the substrate to reduce the flavin cofactor.⁸³

1.3.3.3 Short-chain dehydrogenases

cpkI is proposed to encode a short-chain dehydrogenase (SDH). These are enzymes which catalyse a variety of catabolic and biosynthetic redox reactions, utilising NADH or NADPH as an electron carrier. These cofactors consist of an adenosine diphosphate (ADP) ribose group attached to a nicotinamide heterocyclic ring (figure 1.9). When oxidised to NAD(P)⁺, the 1,4-dihydropyridine moiety is converted to its pyridinium equivalent. In contrast to flavins, nicotinamide cofactors are exclusively bound non-covalently and used in stoichiometric quantities.⁷⁷

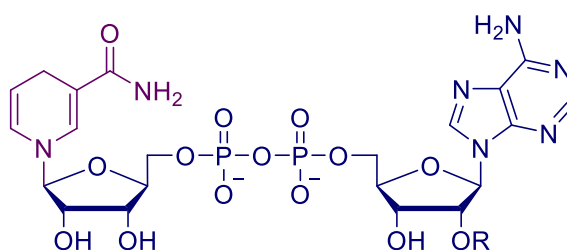
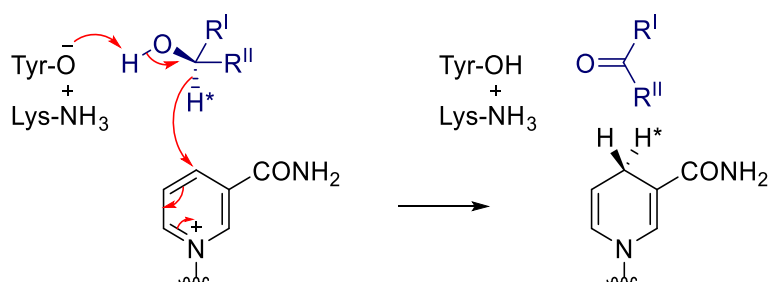


Figure 1.9: The structure of NADH (R = H) and NADPH (R = PO₃²⁻). The nicotinamide and ADP ribose groups are highlighted in purple and blue, respectively.

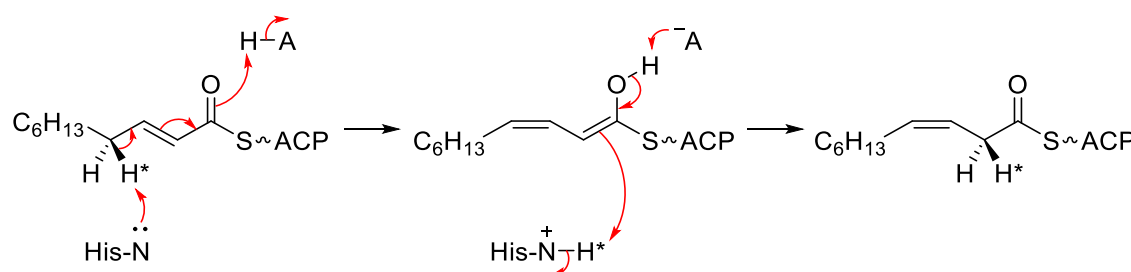
SDHs generally catalyse the oxidation or reduction of carbon-carbon, carbon-oxygen and carbon-nitrogen single- or double-bonds, respectively.⁸⁴ Studies of SDHs from *Drosophila*, humans and bacteria indicate that a conserved tyrosine hydroxyl residue, activated by a nearby lysine, abstracts or donates a proton from/to the substrate, as appropriate (scheme 1.17).^{85,86,87,88} An asparagine residue also participates in the proton relay, while a conserved serine acts to stabilise and polarise the substrate. Contrary to medium-chain dehydrogenases (MDH), it is the *pro*-(4*S*) hydride that is transferred between the nicotinamide cofactor and the substrate.⁸⁴



Scheme 1.17: The dehydrogenation of an alcohol by a short-chain dehydrogenase, using NAD(P)⁺ as a cofactor

1.3.3.4 Isomerases

cpkE is tentatively proposed to encode an isomerase. These are enzymes that catalyse intramolecular rearrangements, such as racemisations, epimerisations, tautomerisations and *cis-trans* isomerisations, often *via* a cofactor-independent mechanism.⁸⁹ Examples include aspartate racemase, methylmalonyl-CoA epimerase, phosphoglucose isomerase and β -hydroxydecanoyl thioester isomerase.^{90,91,92,93} Such enzymes are biologically important in primary and secondary metabolism, particularly in transformations involving carbohydrates, terpenoids and polyketides.⁹⁴ Isomerases generally operate *via* acid-base chemistry, in which a substrate is deprotonated at one position, and reprotonated at another (scheme 1.18).



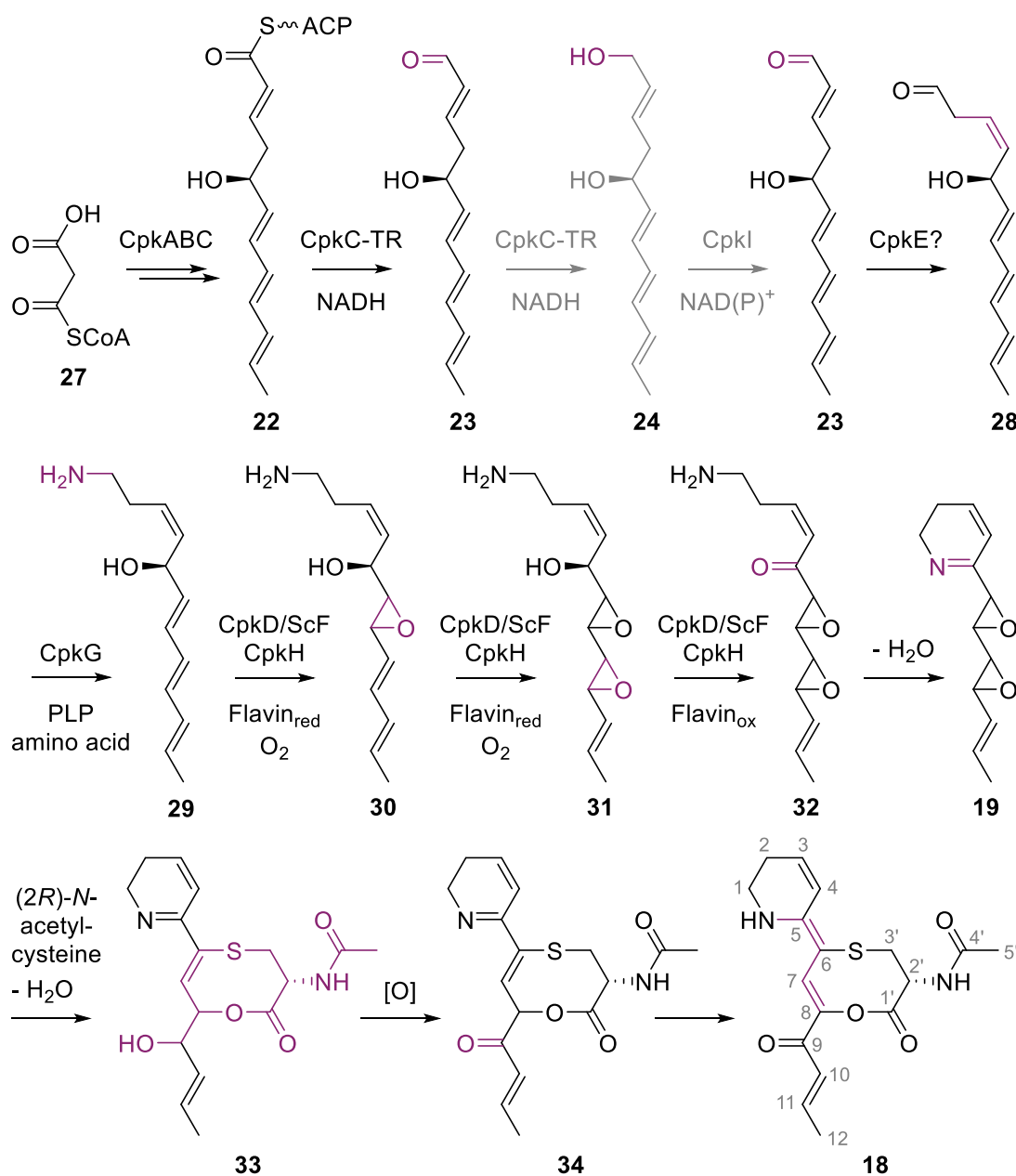
Scheme 1.18: The isomerisation of *trans*-2-decenoic acid to *cis*-3-decenoic acid by β -hydroxydecanoyl thioester isomerase. A = general acid. Revised from Bugg, Schwab and Klassen.^{89,93}

1.4 Proposed biosynthesis of coelimycin P1

Based on the predicted functions of the enzymes encoded by the *cpk* cluster, a biosynthetic pathway was proposed by Gomez-Escribano and coworkers.³⁰

Initially, six units of malonyl-CoA **27**, biosynthesised by CpkP α , CpkP β , Acca1 and CpkK, are likely condensed together and processed by PKS enzymes CpkA, CpkB and CpkC to form thioester **22**. This is then reductively cleaved by the TR domain of CpkC to afford aldehyde **23**, which may be further reduced to the corresponding alcohol **24** and re-oxidised by CpkI. The *E*-configured 2,3-double bond of **23** may then be isomerised to a *Z*-configured 3,4-alkene by CpkE, to allow cyclisation to occur at a later stage. Subsequently, aldehyde **28** is hypothesised to undergo a CpkG-catalysed reductive amination to afford amine **29**, which is then subject to a series of oxidative transformations by flavoenzymes CpkD, ScF and CpkH. These include epoxidations

across the C-6/C-7 and C-8/C-9 double bonds to form mono- and bis-epoxides **30** and **31**, respectively, and an oxidation of the C-4 hydroxyl group, to form ketone **32**. Spontaneous intramolecular imine formation would then afford coelimycin A **19**; the predicted true metabolic product of the *cpk* gene cluster. This is then likely attacked by *N*-acetyl cysteine (or glutamate, in the case of coelimycin P2) with loss of water to produce the 1,5-oxathiocane skeleton. Oxidation of the C-9 hydroxyl group of **33** by molecular oxygen would afford intermediate **34**, which could tautomerise to yield coelimycin P1 **18** (scheme 1.19).



Scheme 1.19: The proposed biosynthesis of coelimycin A **19** and coelimycin P1 **18**

Notably, CpkH and ScF are both substrates of the twin-arginine-translocase (TAT) secretion system,⁹⁵ indicating that they are translocated across the cytoplasmic membrane in their folded states.⁹⁶ Since they are secreted, this suggests that they may catalyse late-stage biosynthetic reactions. CpkF, a putative transmembrane efflux protein, has been proposed to facilitate the export of intermediate **29** before it is further processed.

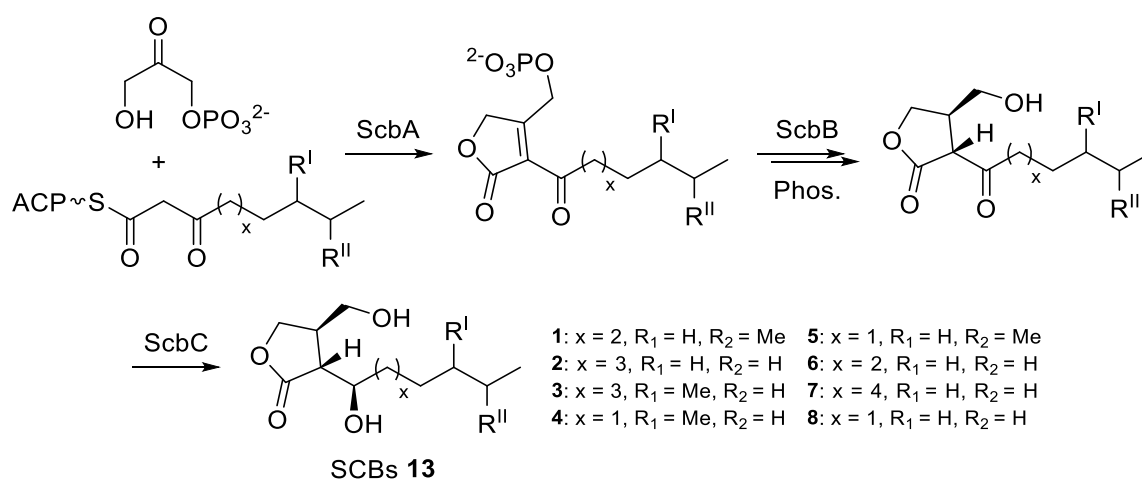
Isotope incorporation experiments have provided some evidence for the proposed pathway. After culturing *S. coelicolor* M1157 in an ¹⁸O₂ atmosphere, it was demonstrated that a single atom of labelled oxygen was specifically incorporated into the C-9 keto group of coelimycin P1 **18**, as anticipated. Furthermore, separate feeding experiments with *N*-(acetyl-*d*₃)-cysteine enantiomers showed the incorporation of deuterium atoms (and therefore *N*-acetylcysteine) into coelimycin P1 in both instances. However, only (*2R*)-*N*-acetylcysteine had the same retention time as unlabelled coelimycin P1 from chiral LC-MS analysis, proving its stereochemistry at that position.³⁰

Moreover, *in vivo* gene knockouts of *cpkD*, *scF* and *cpkH* have provided insight into the roles of the oxidative enzymes and the order of pathway intermediates (Gomez-Escribano, Song, Bibb and Challis, unpublished data). All mutants lost their ability to produce coelimycin P1, demonstrating that the corresponding enzymes are essential for its biosynthesis. Comparative metabolic profiling of *S. coelicolor* M1152Δ*cpkH* with the wild-type strain indicated that CpkH may catalyse the dehydrogenation of alcohol **31** to ketone **32**; molecular ions corresponding to intermediates **23/28** and **29-31**, but not **19**, could be detected. From LC-MS analysis of *S. coelicolor* M1152Δ*cpkD* and *S. coelicolor* M1152Δ*scF* supernatants, ions consistent with the molecular masses of **30** or **31**, respectively (and later stage intermediates), appeared absent, suggesting that CpkD and ScF may catalyse epoxidation reactions. However, further experiments are necessary to establish the structures represented by these ions.

While preliminary studies are suggestive of the functions of some of the biosynthetic enzymes, further investigations are necessary for their full characterisation. Deciphering the biosynthesis of coelimycin P1 could allow for re-engineering of the pathway, for the generation of stable and novel analogues that may still retain antibiotic activity.

1.5 Regulation of coelimycin P1 biosynthesis

The expression of the *cpk* cluster is controlled by several regulatory genes (*scbA*, *scbB*, *scbC*, *scbR*, *scbR2*, *cpkO*, *cpkN* and *cpkM*). For example, *scbA* and *scbB* encode orthologues of AfsA and BprA from *Streptomyces griseus*, which function as a butenolide synthase and a butenolide phosphate reductase, respectively, while *scbC* encodes a homologue of an SDH (it is worth noting that *scbB* and *scbC* are sometimes named inversely).^{97,98} Together, these enzymes have been shown to be essential for the biosynthesis of γ -butyrolactone signalling molecules;^{17,99,98} ScbA and ScbB likely condense dihydroxyacetone phosphate (DHAP) with a β -keto thioester intermediate and reduce the resulting butenolide phosphate, respectively,¹⁰⁰ while ScbC has been demonstrated to catalyse an NADPH-mediated reduction of the keto moiety (scheme 1.20).⁹⁸ At sufficiently high concentrations, these signalling molecules, termed *S. coelicolor* butenolides (SCBs, **13**, scheme 1.20) bind to ScbR; a repressor protein with a TetR-like *N*-terminal helix-turn-helix domain.¹⁰¹ This likely causes a change in conformation of ScbR,¹⁰² resulting in the dissociation of the protein from its target DNA sites and leading to derepression of transcription (figure 1.10). The sites at which ScbR binds include the operator in the bidirectional *scbA/scbR* promoter and the promoters for *cpkO* and *cpkM*.^{103,104} Another repressor protein, ScbR2, also represses transcription of *scbA*, *cpkO* and *cpkM*, but does not bind SCBs.^{104,105,106} Deletion of ScbR2 was found to induce coelimycin P2 production.²⁶



Scheme 1.20: The biosynthesis of *S. coelicolor* butenolides (SCBs, **13**) 1-8. An ACP-bound β -keto thioester from a fatty acid synthase (FAS) is condensed with dihydroxyacetone phosphate (DHAP) by ScbA and the double bond reduced by ScbB. A dephosphorylation (by an unknown enzyme) and reduction of the keto group by ScbC afford the SCBs.^{100,98}

Two further regulatory genes, *cpkO* and *cpkN*, encode *Streptomyces* Antibiotic Regulatory Proteins (SARPs); a family of transcriptional activators that recruit RNA polymerase to its target promoters.¹⁰⁷ However, while *cpkO* is essential for coelimycin biosynthesis, *cpkN* is not.^{108,109} A final regulatory gene, *cpkM*, encodes a two-component histidine kinase. These proteins are usually involved in signal transduction in numerous cellular processes.¹¹⁰

In addition to local regulatory effects, a range of global/pleiotropic regulators have been reported to bind to various promoters in the *cpk* cluster (figure 1.10), influencing the transcription of upstream genes. For example, the two-component systems DraR/K and AfsQ1/Q2 directly repress the transcription of *cpkO* and activate the transcription of *cpkABC*, respectively.^{111,112,113} Furthermore, DasR (involved in *N*-acetyl glucosamine metabolism),¹¹⁴ NdgR (involved in controlling nitrogen source-dependent growth)¹¹⁵ and PhoP (involved in phosphate metabolism)¹¹⁶ all bind to the *scbA/scbR* promoter region, suggesting a connection between primary and secondary metabolism through the SCB system.⁹⁸ Another global regulator, ArgA (involved in mediating arginine biosynthesis), has also been implicated in the downregulation of several *cpk* genes, although its target DNA sites are not known.¹¹⁷

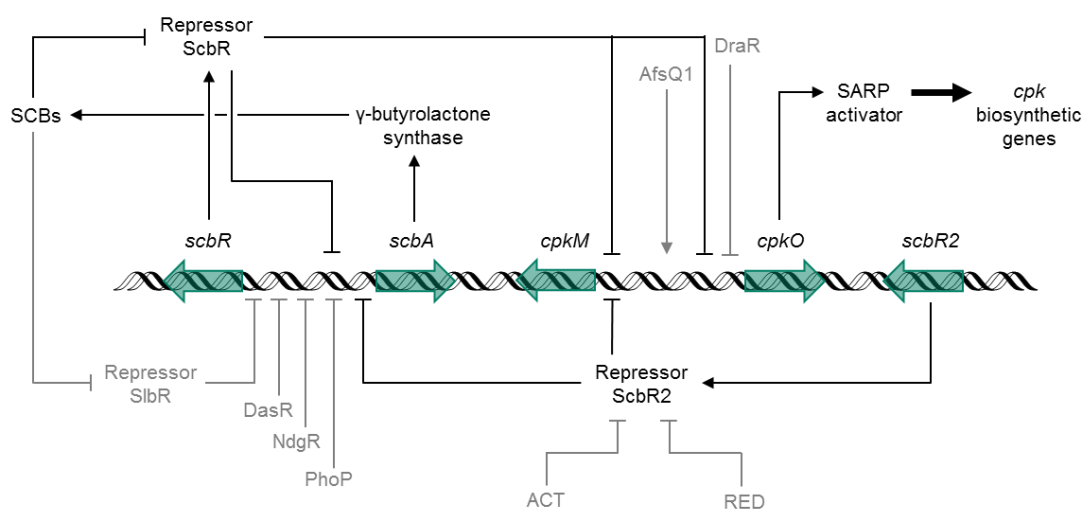


Figure 1.10: Regulation of coelimycin biosynthesis, involving *S. coelicolor* butenolide (SCB) signalling molecules and interactions with receptors and ligands from other pathways (grey). Activatory and inhibitory steps are indicated by arrows or lines ending with a bar, respectively. Additional binding sites of ScbR2 (intergenic regions of *accaI-scF*, *cpkA-cpkD*, *cpkI-cpkJ* and the promoter of *cpkN*)¹⁰⁴ have been omitted for clarity. Adapted from Liu and coworkers.¹⁰⁸

Further transcriptional regulation of the *cpk* cluster can occur *via* crosstalk with the *act* and *red* biosynthetic pathways. For instance, it has been demonstrated that actinorhodin **9** and undecylprodigiosin **10** can bind ScbR2 and release it from its promoter (figure 1.10), resulting in the derepression of transcription of *cpkO*.¹⁰⁵ Likewise, addition of SCBs **13** or inactivation of *scbA* alters the expression of the *act* and *red* pathways.^{118,17}

1.6 Other polyketide alkaloids

Coelimycin P1 **18** was the first actinobacterial alkaloid to be conclusively characterised as the product of a type I modular PKS. Several other actinobacterial alkaloids also appear to be of polyketide origin, as demonstrated by incorporation experiments with isotope-labelled acetate and propionate units. Examples include latumcidin **35** (also known as abikoviromycin), nigrifactin **36**, pyrindicin **37**, streptazolin **38** and cyclizidine **39**.^{119,120,121,122,123} The similarly-structured alkaloids JBIR-102 **40**, 4-hydroxy-(2-penta-1,3-dienyl)piperidine **41**, streptazone E **42**, indolizomycin **43** and iminimycin A **44** may also derive from polyketide precursors (figure 1.11).^{124,125,126,127,128} Many of these metabolites display interesting biological properties, including antibacterial, antifungal, cytotoxic and beta-blocker activities.

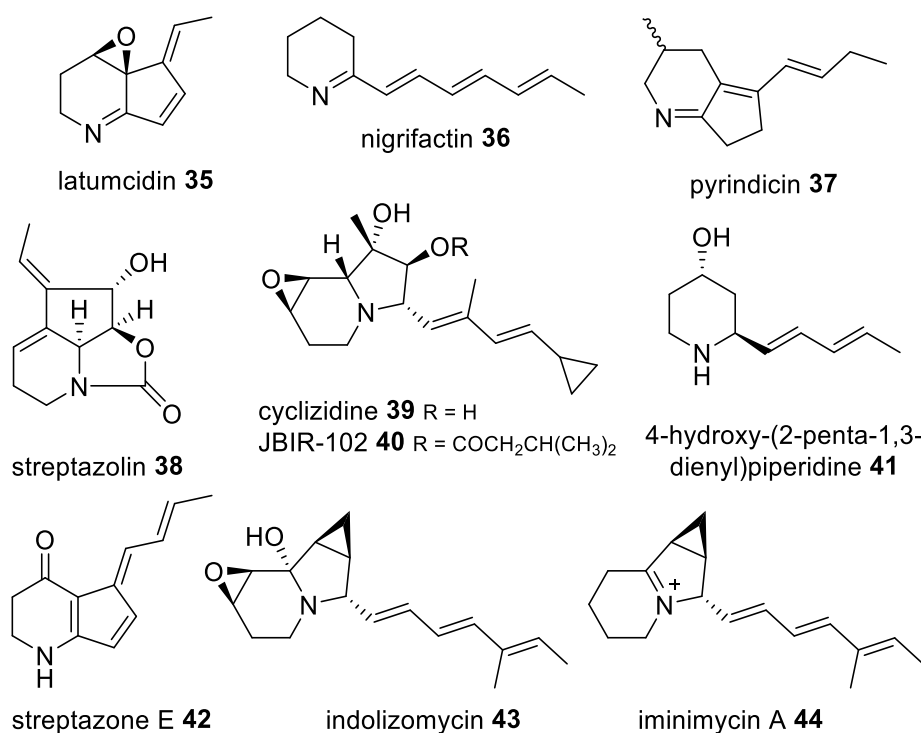


Figure 1.11: Structures of polyketide alkaloids from Actinobacteria

At the start of this project, no genetic information was available to link these metabolites to polyketide clusters. Understanding their biosynthesis could uncover new enzymology and allow for the engineering or discovery of novel polyketide alkaloids with potentially wide-ranging activities.

1.7 Aims of the project

The main aim of this project was to investigate the biosynthetic reactions and enzymatic machinery involved in actinobacterial polyketide alkaloid formation, with primary focus on coelimycin P1. This would initially involve further investigations of CpkG, to determine factors such as the substrate scope, preference and stereospecificity of the transamination reaction, and to attempt to obtain a crystal structure of the protein. Following this, the aim was to clone the remaining post-PKS tailoring genes (*cpkI*, *cpkE*, *cpkD*, *scF* and *cpkH*) and purify the resulting recombinant enzymes. A range of putative substrates and standards would then be synthesised to establish the activities of the enzymes and to elucidate the *cpk* biosynthetic pathway *in vitro*.

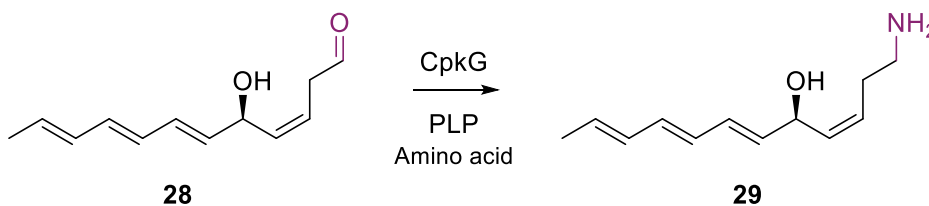
An additional aim was to explore the biosynthesis of further polyketide alkaloids through bioinformatics-guided and genomics-based approaches, in the hope of establishing whether general mechanistic principles underlie the assembly of this class of metabolite.

2 RESULTS AND DISCUSSION I: Nitrogen incorporation in coelimycin and related polyketide alkaloids

2.1 Biochemical characterisation of CpkG

2.1.1 Introduction

CpkG, a pyridoxal-5'-phosphate (PLP)-dependent ω -transaminase (TA), is proposed to catalyse the conversion of aldehyde **28** to amine **29** during coelimycin biosynthesis (scheme 2.1).



Scheme 2.1: Proposed reaction catalysed by CpkG in coelimycin biosynthesis³⁰

CpkG has previously been shown to catalyse transamination reactions between selected aldehydes and amines (Ruby Awodi, University of Warwick). However, factors such as the substrate scope, preference and stereospecificity of the reaction remained to be investigated. Furthermore, CpkG had yet to be incubated with close structural analogues of **28** and **29** to establish its role in the proposed pathway.

2.1.2 Overproduction, purification and characterisation of CpkG

In order to probe the transaminase activity of CpkG *in vitro*, *E. coli* One Shot® BL21 Star™ cells were first transformed with pET151::cpkG; a plasmid designed to overexpress cpkG from the IPTG-inducible T7 promoter as an N-terminal His₆-fusion protein (engineered by Ruby Awodi). Following overnight induction of cpkG with IPTG, the cells were harvested and lysed. The recombinant protein was then purified from cell-free extracts by immobilised metal-affinity chromatography (IMAC), using Ni-nitriloacetic acid (Ni-NTA) resin.

SDS-PAGE analysis demonstrated that His₆-CpkG was pure and of the expected molecular weight (figure 2.1a). The exact molecular weight of the protein was verified by high-resolution mass spectrometry (HRMS) (figure 2.1b; calculated m/z = 60646 Da, observed m/z = 60648 Da). Despite obtaining some insoluble protein, 10 mg/mL of soluble protein was recovered.

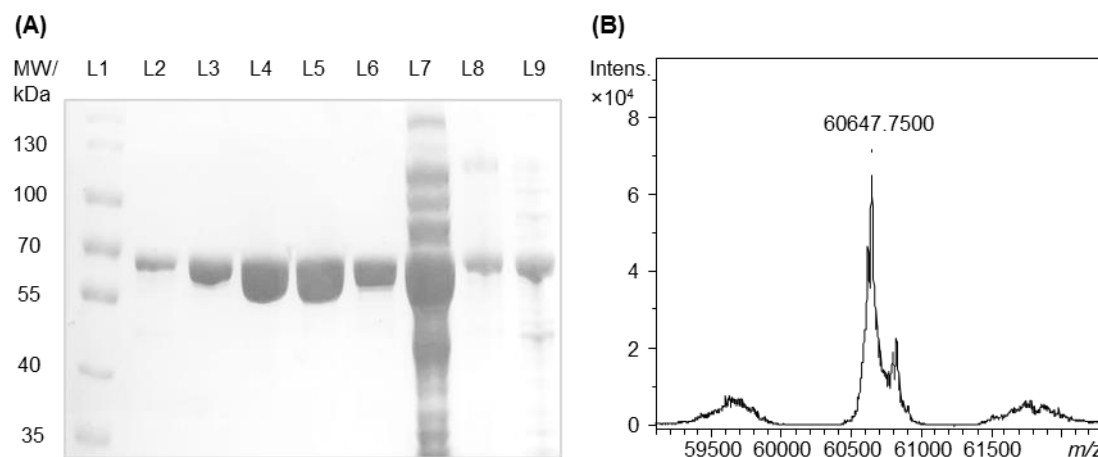


Figure 2.1: (A) Nickel affinity purification of His₆-CpkG. L1: Molecular weight marker; L2: elution with 20 mM imidazole; L3: elution with 50 mM imidazole; L4: elution with 100 mM imidazole; L5: elution with 200 mM imidazole; L6: elution with 300 mM imidazole; L7: flow through from column; L8: soluble protein; L9: insoluble protein. (B) Deconvoluted mass spectrum of His₆-CpkG. The measured and calculated molecular weights of His₆-CpkG are 60648 Da and 60646 Da, respectively.

To investigate the ability of CpkG to covalently bind PLP, a UV/Vis absorption spectrum of the recombinant enzyme purified in the presence of the cofactor was compared to that of PLP alone. At pH 8.0, a characteristic λ_{max} shift from 410 nm to 418 nm was observed, indicating that PLP forms an internal aldimine bond with CpkG (figure 2.2).¹²⁹ Assuming that free PLP and the aldimine adduct have similar molar extinction coefficients, the concentration of bound PLP can be determined. The near-identical absorbance intensities at λ_{max} are thus consistent with a 1:1 molar binding ratio.

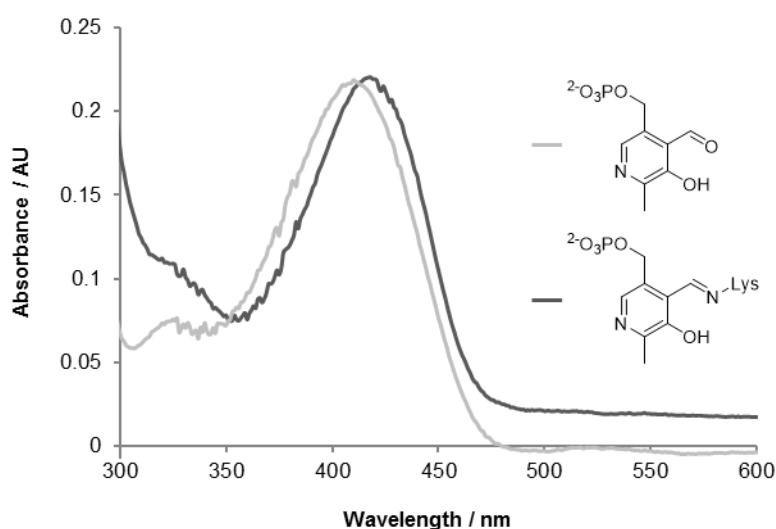
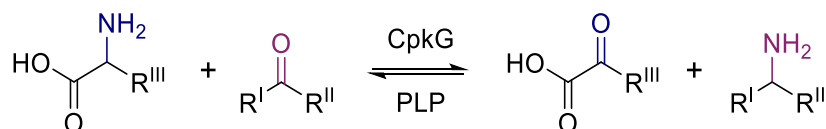


Figure 2.2: UV/Vis spectra and structures of PLP (light grey) and the aldimine adduct of PLP with CpkG (dark grey). The λ_{max} values of the two peaks are 410 nm and 418 nm, respectively.

2.1.3 *In vitro* assays

The activity of CpkG was then investigated *in vitro*, by employing three different assays to detect the four components of the transamination reaction: the amino acid, aldehyde/ketone, α -keto acid and amine (scheme 2.2).⁵⁷

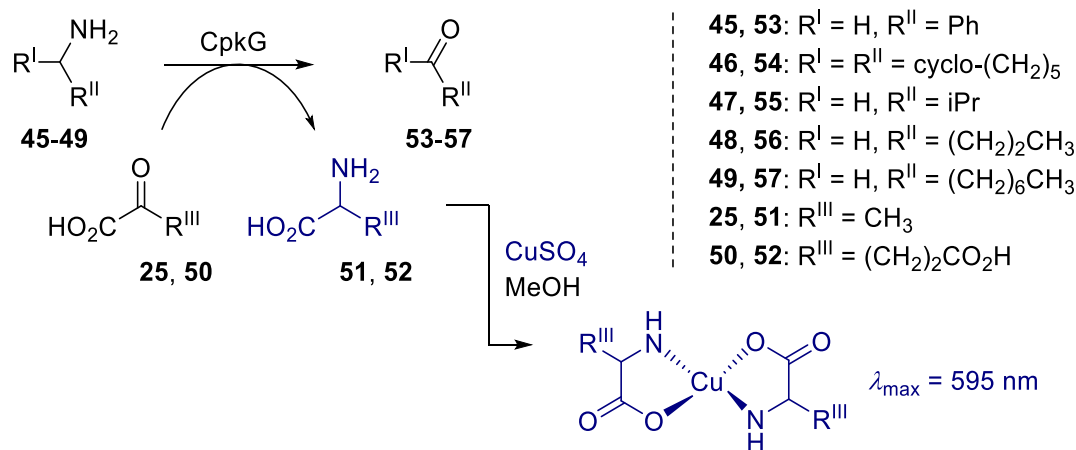


Scheme 2.2: The general transamination reaction catalysed by CpkG.

2.1.3.1 Amino acid and aldehyde detection

In collaboration with Dr. Joleen Masschelein, University of Warwick

The activity and substrate specificity of CpkG were initially explored using a spectrophotometric assay designed to detect the production of amino acids *via* formation of a blue copper (II) complex (scheme 2.3).¹³⁰



Scheme 2.3: Assay employed to detect amino acids produced during the transamination reaction.¹³⁰ Amines (10 mM, **45-49**) and α -keto acids (10 mM, **25** and **50**) tested as substrates for CpkG (20 μ M) are shown.⁵⁷

His₆-CpkG, supplemented with PLP during purification (herein referred to as holo-His₆-CpkG), was incubated with several amines and either pyruvate or α -ketoglutarate, to examine conversion to the corresponding aldehydes/ketones, and either alanine or glutamic acid, respectively. The absorbance at 595 nm was then measured relative to blank and control samples.

All of the amines tested were found to stimulate conversion of the two α -keto acids to the corresponding amino acids, demonstrating that CpkG has a broad substrate tolerance (figure 2.3). No significant differences in relative specificities between benzylamine **45**, cyclohexylamine **46**, isobutylamine **47** and butylamine **48** were observed when using α -ketoglutarate **50** as a cosubstrate, suggesting that linear, branched, cyclic and aromatic amines can be similarly tolerated. For butylamine and benzylamine, considerably lower levels of activity were detected when using pyruvate **25** as a cosubstrate, although it is unclear why. The highest level of activity was observed for octylamine **49**, together with pyruvate, indicating that CpkG prefers medium-chain alkylamines. This is consistent with its proposed role in coelimycin biosynthesis.

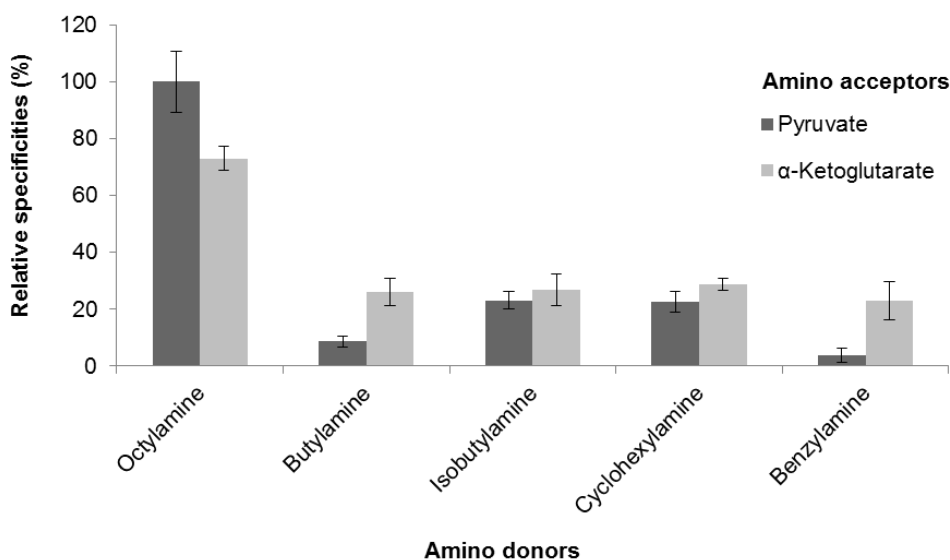
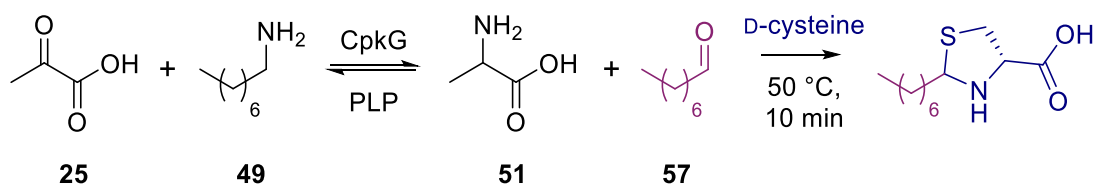


Figure 2.3: Amino donor and acceptor specificities of CpkG as determined by the $\text{CuSO}_4/\text{MeOH}$ staining method.¹³⁰ The specificities for the different substrate combinations are shown relative to the specificity of CpkG for octylamine and pyruvate as amino donor and acceptor. Data are represented as means \pm 1 standard deviation from triplicate experiments.

Utilising the preferred substrates, octylamine and pyruvate, the formation of octanal **57** was examined by LC-MS. Because aldehydes typically have a poor ionisation efficiency, both the product of the reaction and an authentic standard of octanal were derivatised with D-cysteine prior to analysis (scheme 2.4).¹³¹ The retention time and molecular weight of the derivatised reaction product were consistent with those of the authentic standard (figure 2.4), thereby proving that octanal was formed during the reaction.



Scheme 2.4: Assay employed for the detection and derivatisation of octanal **57** with D-cysteine (400 μM),¹³¹ following reaction of pyruvate **25** (200 μM) and octylamine **49** (200 μM) with CpkG (90 μM).

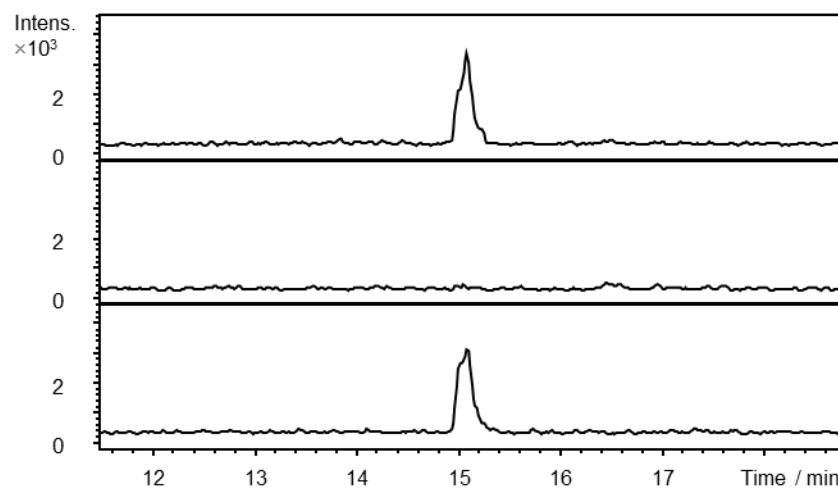
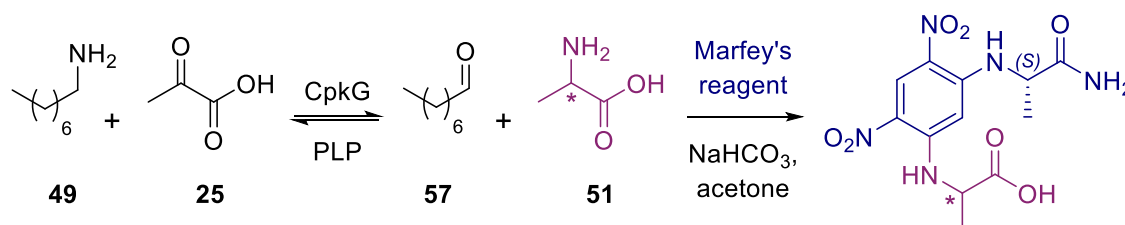


Figure 2.4: Extracted ion chromatograms at $m/z = 232.1366 \pm 0.005$ (corresponding to the $[\text{M}+\text{H}]^+$ ion of the D-cysteine derivative of octanal) from LC-MS analyses of the CpkG-catalysed reaction of octylamine with pyruvate. From top to bottom: enzymatic reaction; control from which the enzyme was omitted; and authentic standard.

2.1.3.1.1 Determination of amino acid stereochemistry

To determine the stereochemical outcome of the CpkG-catalysed reaction of octylamine with pyruvate, the alanine **51** produced was derivatised with enantiomerically-pure Marfey's reagent (scheme 2.5).¹³²



Scheme 2.5: Assay employed to establish the stereochemistry of alanine **51** produced,¹³² following reaction of octylamine **49** (10 mM) and pyruvate **25** (10 mM) with CpkG (20 μM).

The Marfey's-derivatised product was then compared to authentic standards of derivatised L- and D-alanine by LC-MS. As these compounds are diastereomers, they display different retention times under optimised elution conditions. Analysis of the

chromatograms revealed that only L-alanine is formed during the transamination reaction (figure 2.5). Because the reaction is reversible, this suggests that CpkG has preference for L-amino acid substrates, and may function as an (*S*)-selective ω -TA.

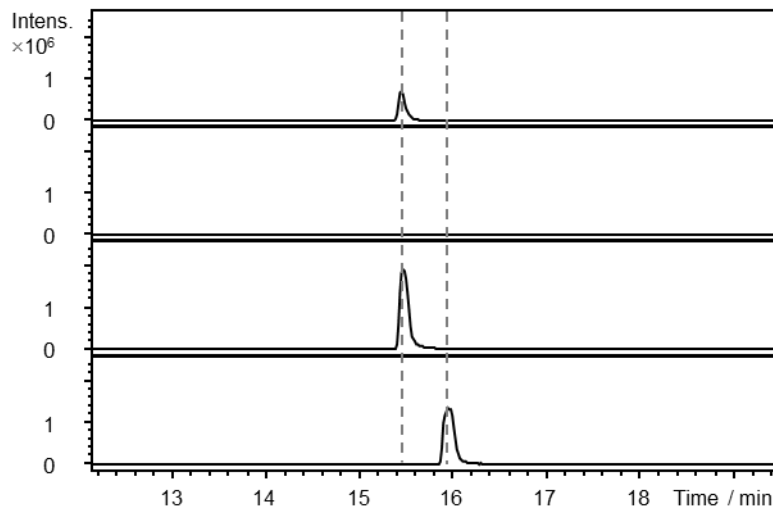
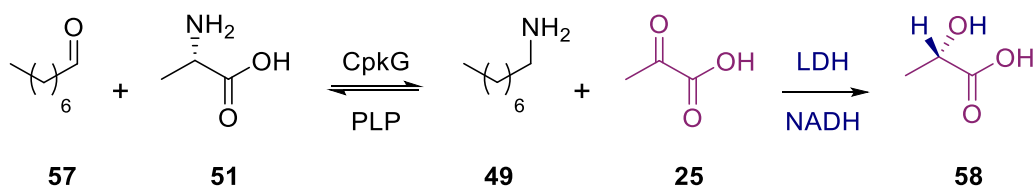


Figure 2.5: Extracted ion chromatograms at $m/z = 342.1044 \pm 0.005$ (corresponding to the $[M+H]^+$ ion of Marfey's-derivatised alanine) from LC-MS analyses of the CpkG-catalysed reaction of octylamine with pyruvate. From top to bottom: Marfey's derivatisation of the product of the reaction; Marfey's derivatisation of the control reaction from which enzyme was omitted; Marfey's derivatisation of an L-alanine standard; Marfey's derivatisation of a D-alanine standard.

2.1.3.2 α -Keto acid and amine detection

Having established that octylamine and pyruvate are good substrates for CpkG in the reverse of the proposed biosynthetic reaction, reductive amination of octanal using alanine as an amino donor was subsequently examined.

Pyruvate formation was monitored using a coupled assay with lactate dehydrogenase (LDH), that exploits the absorption of NADH, but not NAD^+ , at 340 nm (scheme 2.6).¹³³



Scheme 2.6: Assay employed to detect octylamine **49** and pyruvate **25** production, following reaction of octanal **57** (200 μM) and alanine **51** (200 μM) with CpkG (90 μM). Octylamine was observed by LC-MS comparison to an authentic standard, while pyruvate was detected by monitoring the decrease in absorbance at 340 nm after addition of LDH (1 μM) and NADH (200 μM).¹³³

A decrease in absorbance at 340 nm was observed, indicating that NADH was being oxidised during the conversion of pyruvate to lactic acid **58** (figure 2.6a). The octylamine formed during the same reaction was detected by LC-MS comparisons with an authentic standard (figure 2.6b).⁵⁷

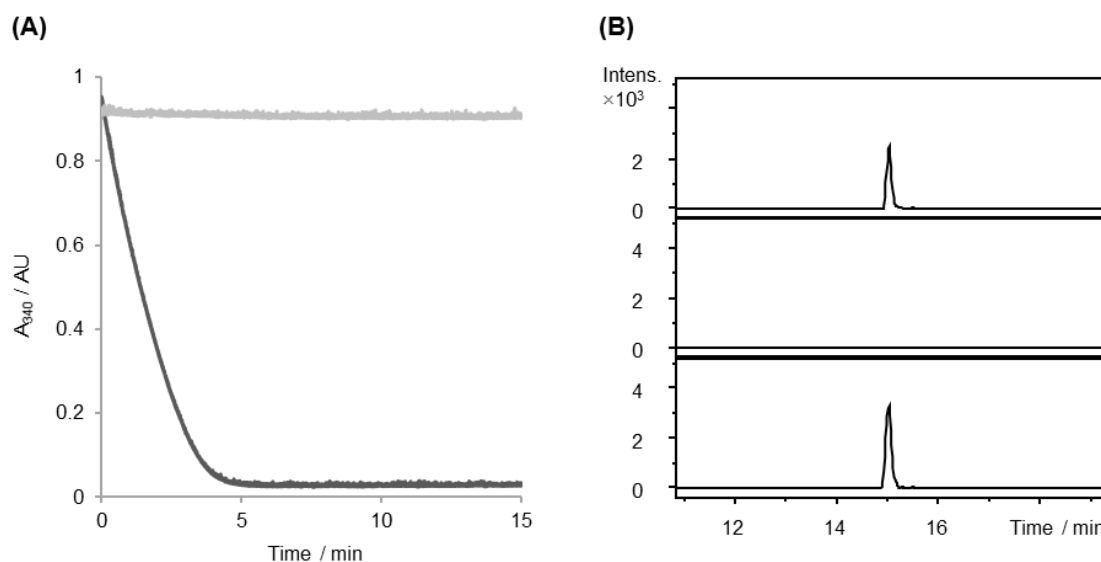
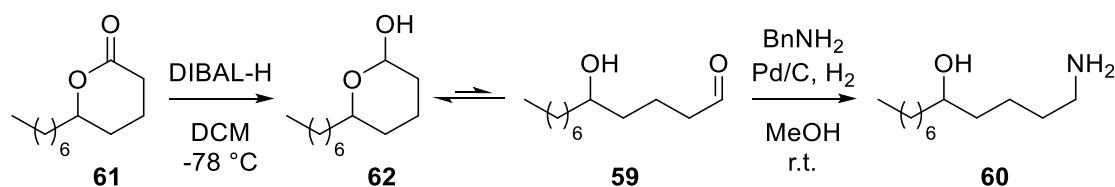


Figure 2.6: (A) Detection of pyruvate formation from L-alanine using a coupled assay with LDH and NADH. The consumption of NADH was determined by monitoring the decrease in absorbance at 340 nm after CpkG addition (dark grey) and without CpkG addition (light grey). (B) Extracted ion chromatograms at $m/z = 130.1590 \pm 0.005$ (corresponding to the $[M+H]^+$ ion for octylamine) from LC-MS analyses of the CpkG-catalysed reaction of octanal with L-alanine. From top to bottom: enzymatic reaction; control reaction from which the enzyme was omitted; and authentic standard.

2.1.3.3 Utilisation of synthetic substrate mimics

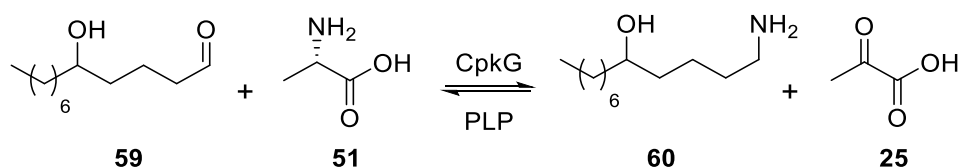
Once the transaminase activity of CpkG had been confirmed, two substrate analogues (**59** and **60**), containing the same 5'-hydroxyl functional group and carbon chain length as the proposed true intermediates of the coelimycin biosynthetic pathway (**28** and **29**), were synthesised (scheme 2.7).

In the first step, commercially-available δ -dodecalactone **61** was reduced with DIBAL-H, following a procedure described by Zhou and coworkers.¹³⁴ 5-Hydroxydodecanal **59**, existing predominantly as lactol **62**, was afforded in moderate yield. Subsequently, benzylamine, Pd/C and H_2 were added to effect a mild, one-pot conversion of the aldehyde to the corresponding primary amine **60** *via* reductive amination. Several milligrams of product were purified by HPLC for use in *in vitro* assays.



Scheme 2.7: Left: Synthesis of 5-hydroxydodecanal **59** by reduction of δ -dodecalactone **61** (1.0 eq.) with DIBAL-H (1.0 eq.). Right: Synthesis of 1-aminododecan-5-ol **60** by reaction of **59** (1.0 eq.) with benzylamine (1.3 eq.), Pd/C (catalytic) and H₂ (1 atm).

Compounds **59** and **60** were then incubated with holo-His₆-CpkG and either alanine **51** or pyruvate **25**, respectively (scheme 2.8). The products of the reaction were then analysed by LC-MS, after derivatisation, if appropriate, as described in sections 2.1.3.1.



Scheme 2.8: General assay for the detection of 1-aminododecan-5-ol **60** (forward reaction) or 5-hydroxydodecanal **59** (reverse reaction), from incubation of the appropriate substrates (200 μ M) with CpkG (90 μ M). For the reverse reaction, **59** was derivatised with D-cysteine (400 μ M) prior to LC-MS analysis.

Amine **60** and aldehyde **59** were readily detected in the forward and reverse reactions, respectively (figure 2.7). This implicates CpkG in catalysing the analogous conversion of **28** to **29** in coelimycin biosynthesis.

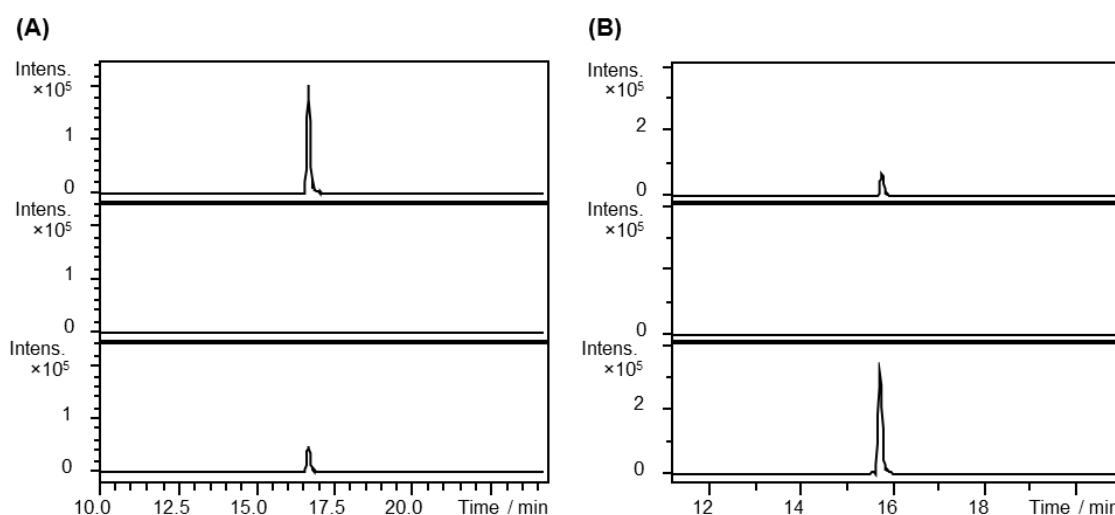


Figure 2.7: (A) Extracted ion chromatograms at $m/z = 202.2165 \pm 0.005$ (corresponding to the $[M+H]^+$ ion for 1-aminododecan-5-ol) from LC-MS analyses of the CpkG-catalysed reaction of 5-hydroxydodecanal with L-alanine. From top to bottom: enzymatic reaction; control reaction from which the enzyme was omitted; and authentic standard. (B) Extracted ion chromatograms at $m/z = 304.1941 \pm 0.005$ (corresponding to the $[M+H]^+$ ion for D-cysteine-derivatised 5-hydroxydodecanal) from LC-MS analyses of the CpkG-catalysed reaction of 1-aminododecan-5-ol with pyruvate. From top to bottom: enzymatic reaction; control reaction from which the enzyme was omitted; and authentic standard.

2.2 Crystal structures of CpkG

2.2.1 Crystallisation screens

In collaboration with Dr. Lona Alkhalaf, University of Warwick

To gain further insight into the transamination mechanism, substrate tolerance and stereoselectivity of CpkG, the recombinant holo-protein was purified by nickel affinity chromatography for crystallisation trials. Initially, trials were performed in the absence of substrate, using three commercially available screens (JCSG-plus, Structure and PACT premier) and a Honeybee robot.

A crystal was obtained from a PACT premier screen (figure 2.8) that diffracted to 2.85 Å. A full data set was acquired, enabling the structure of CpkG to be solved by molecular replacement and refinement (Dr. Lona Alkhalaf, see appendices 2 and 3). CrmG, an ω-TA sharing 50 % sequence identity, was used as the template.¹³⁵

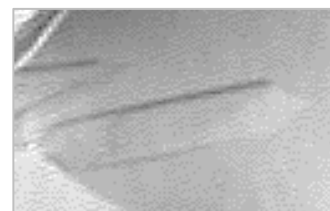


Figure 2.8: A crystal obtained from a PACT premier screen

Unfortunately, the electron density around the cofactor and active site residues was not well defined; possibly because PLP was present in both covalently-bound and unbound forms. Therefore, crystallisation trials were repeated under optimised conditions (0.2 M sodium malonate dibasic monohydrate, 0.15 M bis-tris propane, 20 % w/v PEG, pH 6.5) and in the presence of L-alanine **51**, in order to fully convert PLP to its PMP equivalent. A crystal structure (CpkG_ala) was determined at 2.60 Å resolution.

2.2.2 Overall structure of CpkG

The crystal structure of CpkG_ala confirms that the protein is indeed homodimeric, as it is composed of two identical monomer units (figure 2.9a). Each monomer contains three domains; a large domain, a small domain, and an additional domain (figure 2.9b).

The large domain contains a central seven-stranded mixed β-sheet, while the small domain contains two three-stranded antiparallel β-sheets near to both the *N*- and *C*-termini. These features are characteristic of a fold-type I conformation, adopted by

(*S*)-selective ω -TAs.^{136,137,138} Thus, the structure is consistent with the observation that CpkG catalyses the stereospecific conversion of pyruvate to L-alanine (see section 2.1.3.1.1).

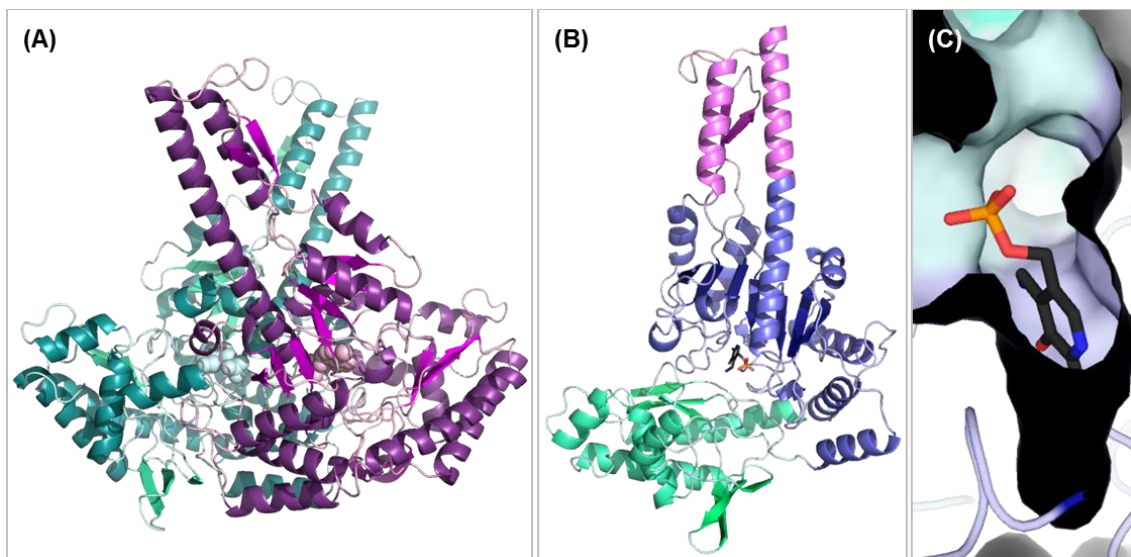
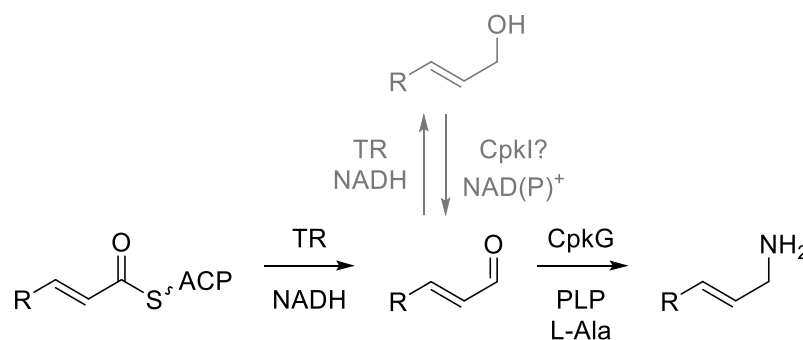


Figure 2.9: (A) A cartoon display of the CpkG homodimer. Monomer units are highlighted in violet and teal, while PMP ligands are depicted as spheres. (B) A cartoon display of the domain architecture of a CpkG monomer. The large domain, small domain and additional domain are highlighted in blue, green and magenta, respectively, while the PMP ligand is depicted as sticks in CPK colouring. (C) The predicted substrate binding tunnel positioned behind the primary amino group of PMP, displayed in surface view.

Both domains also comprise several α -helices and unstructured loops. At the interface of the two domains, a molecule of PMP is bound. The substrate L-alanine or product pyruvate appear absent from the structure, presumably due to diffusion following the first half of the transamination reaction. However, it is possible to discern the putative substrate binding tunnel (figure 2.9c).

The additional domain consists of a two-stranded antiparallel β -sheet and a helical bundle formed from two α -helices (figure 2.9b, magenta). To date, the only other PDB entry reported to contain a structurally similar domain is that of CrmG.¹³⁵ Zhu and coworkers attempted to establish the function of the domain through structure-guided mutagenesis. Deletion of a small helix in the additional domain yielded a mutant with near identical activity compared to wild-type CrmG. Other deletions afforded insoluble variants, suggesting that the domain is important for maintaining the correct protein conformation.¹³⁵

An alternative hypothesis could be that the domain plays a role in protein-protein interactions. In coelimycin P1 biosynthesis, the ACP-bound thioester **22** is reductively cleaved by the TR domain to yield aldehyde **23**, which is further reduced to alcohol **24** (see section 1.3.1.4). However, interaction of the TR domain with CpkG could promote direct conversion of **23** to the corresponding amine before further reduction can occur (scheme 2.9).



Scheme 2.9: Possible biosynthetic pathways leading to the formation of the amine intermediate. Grey: original proposed pathway; black: alternative pathway in which CpkC-TR and CpkG interact with one another, resulting in the direct conversion of the aldehyde to the amine. Note: the postulated isomerisation catalysed by CpkE has been omitted for clarity.

Indeed, a similar phenomenon has previously been reported for the NRPS-derived myxochelins. Reduction of the peptidyl carrier protein (PCP)-tethered thioester yields an aldehyde intermediate, which is then further reduced to afford myxochelin A, or transaminated to produce myxochelin B.¹³⁹ Moreover, as free aldehydes can be harmful to microbial cells,¹⁴⁰ having protein-associated aldehyde could be an important detoxification mechanism in *Streptomyces*. This theory prompted investigations into binding interactions between CpkC-TR and CpkG (see section 2.3).

2.2.3 Binding of PMP

A closer look at the active site reveals that PMP is held in place by several polar interactions (figure 2.10a). A conserved aspartate (Asp317) and glutamine (Gln320) form hydrogen bonds with the pyridine nitrogen and 3'-hydroxyl group of PMP, respectively, while residues Gly129, Ala130 and Thr376 interact with the phosphate moiety. Two additional hydrogen bonds are present between the primary amino group of PMP, and the side-chain carboxylate and amino groups of Glu289 and Lys346, respectively. This lysine residue is predicted to covalently bind PLP through internal and external aldimine linkages during transamination.

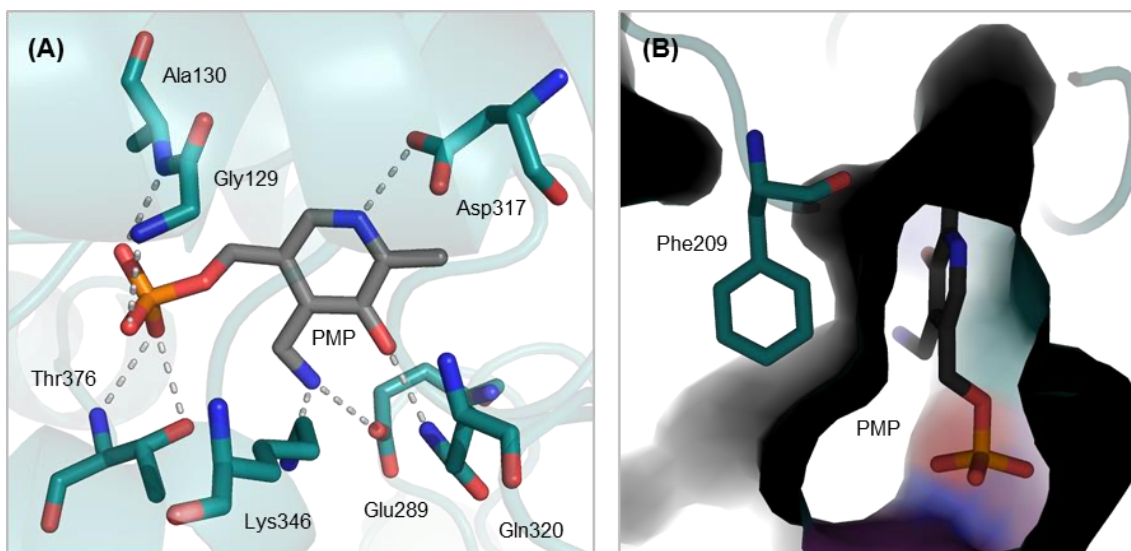


Figure 2.10: (A) Polar interactions of PMP with surrounding residues. (B) Face-to-edge π - π stacking interaction of PMP with Phe209. The carbon chains of PMP and residues are depicted in grey and teal, respectively, while other atoms are shown in CPK colouring.

The aromatic ring of PMP partakes in a non-polar, face-to-edge π - π stacking interaction with Phe209 (figure 2.10b). This residue likely stabilises the formation of external aldimine and quinonoid intermediates during transamination.^{135,141}

2.2.4 Binding of PLP

In an attempt to determine the mechanism of substrate binding, crystallisation trials were repeated in the presence of both L-alanine **51** and 5-hydroxydodecanal **59**. A crystal structure (CpkG_ala_lactol) was subsequently solved to 2.60 Å resolution (Dr. Lona Alkhalaf, University of Warwick).

The binding of L-alanine, 5-hydroxydodecanal or their corresponding products could not be observed, which may again be attributable to product diffusion. However, the PMP cofactor was instead replaced with a PLP-internal aldimine, through the predicted ϵ -amino linkage with Lys346 (figure 2.11a), suggesting that at least one complete transamination had occurred during crystallisation.

The overall conformation and active-site fold are very similar in both the PMP- and PLP-bound states. An additional hydrogen bonding interaction between Thr376 and the imine group of bound PLP appears to stabilise the cofactor (figure 2.11a). An overlay of the two structures reveals that several residues around the substrate binding pocket have

undergone small conformational changes, namely Tyr64, Glu289, Arg488 (figure 2.11b). A similarly-positioned tyrosine residue (Tyr85) in human ornithine aminotransferase (OAT) has been demonstrated to have considerable flexibility and to be a key determinant in substrate recognition.^{142,143}

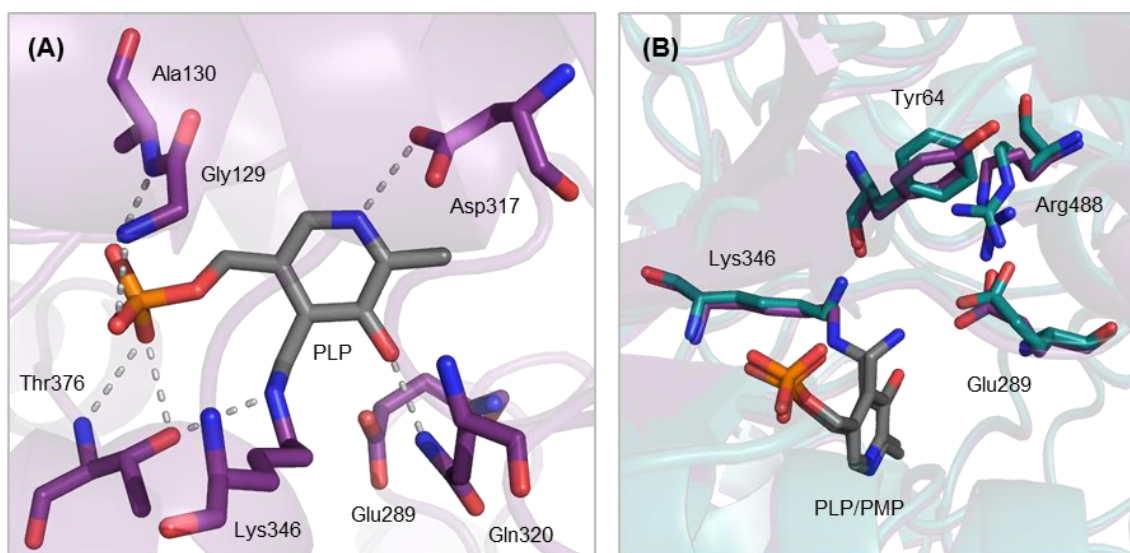


Figure 2.11: (A) Polar interactions of PLP with surrounding residues. (B) An overlay of CpkG_ala (teal) with CpkG_ala_lactol (violet). Residues that have undergone significant conformational alternations are depicted. Nitrogen, oxygen and phosphorous atoms are shown in CPK colours.

Moreover, the corresponding glutamate and arginine residues in OAT have been proposed to operate together as a functional switch, permitting or restricting the binding of substrates at different points in the catalytic cycle.¹⁴⁴ This ‘gateway’ system has since been observed in several other ω -TAs (e.g. acetylornithine aminotransferase from *Thermus thermophilus* HB8¹⁴⁵ and γ -aminobutyrate aminotransferase from *E. coli*¹⁴⁶) and allows the enzyme to confer specificity towards longer-chain amines lacking an α -carboxyl group, while maintaining the ability to bind the α -keto acid acceptor.¹³⁸ Interestingly, these residues have only been reported in α -ketoglutarate-dependent ω -TAs, which contain a second conserved arginine (Arg180 in OAT) to bind the γ -carboxylate group.¹³⁸ This residue is absent in CpkG, which preferentially utilises pyruvate (see section 2.1.3.1). However, the presence of the Glu289-Arg488 ion pair suggests that it may operate in a similar manner.

Future work could explore the soaking of ligands into pre-existing crystals, in order to prevent product diffusion, determine the mode of binding and predict the substrate scope of CpkG.

2.3 Investigations into CpkG and CpkC-TR binding

Following examination of the crystal structure of CpkG, it was proposed that CpkG and CpkC-TR may interact to promote the direct conversion of aldehyde **23** to the corresponding amine in coelimycin biosynthesis (see section 2.2.2). This hypothesis was investigated using two different techniques: native-PAGE analysis and pull-down assays.

2.3.1 Native-PAGE analysis

To investigate potential non-covalent binding interactions between CpkC-TR and CpkG, the *N*-terminal hexahistidine fusion proteins were overproduced in *E. coli* from the appropriate pET151 constructs (engineered by Ruby Awodi, University of Warwick). They were then purified to homogeneity by nickel affinity chromatography and gel filtration (figure 2.12a).

The two proteins were subsequently incubated together, and speculative binding events were probed by monitoring protein migration under non-denaturing conditions. Thus, PAGE gels were prepared in the absence of SDS, and proteins were electrophoresed at 4 °C using a low applied voltage.

If the two proteins interact, a higher band should be apparent in the gel, due to the larger hydrodynamic radius of the complex.¹⁴⁷ However, the only observable bands corresponded to a mixture of CpkC-TR aggregates and the monomer and dimer of CpkG (figure 2.12b). This either suggests that the proteins do not interact, or that binding is transient or too weak for detection. Therefore, a second technique was utilised to distinguish between these possibilities.

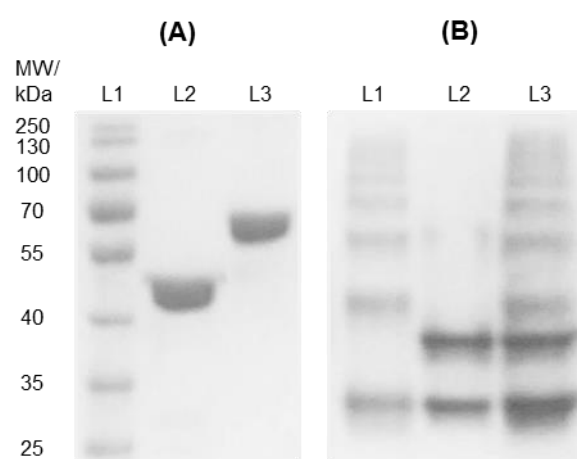


Figure 2.12: (A) SDS-PAGE analysis of His₆-CpkC-TR (L2) and His₆-CpkG (L3), relative to a protein molecular weight marker (L1). The molecular weights of His₆-CpkC-TR and His₆-CpkG are 46.0 and 60.6 kDa, respectively. (B) Native-PAGE analysis of His₆-CpkC-TR (L1), His₆-CpkG (L2) and a 1:1 mixture of His₆-CpkC-TR and His₆-CpkG (L3).

2.3.2 Pull-down assays

Pull-down assays rely on the use of a ‘bait’ protein, which is tagged and immobilised on a specific ligand. Putative interactions with untagged ‘prey’ proteins can be inferred if they are captured and ‘pulled-down’ when the bait protein is eluted.¹⁴⁸

In these assays, His₆-CpkC-TR and Ni-NTA resin were used as the bait and ligand, respectively. The His₆-tag of CpkG was removed using TEV protease, a cysteine protease with a stringent sequence specificity, to obtain speculative prey protein (figure 2.13a). The two proteins were then incubated together and passed through the resin. CpkG could only be observed in the wash fraction and did not co-elute with His₆-CpkC-TR (figure 2.13c), suggesting that the two proteins do not form significant interactions. Similar results were obtained in an inverse assay using His₆-CpkG and CpkC-TR (figure 2.13b, figure 2.13d).

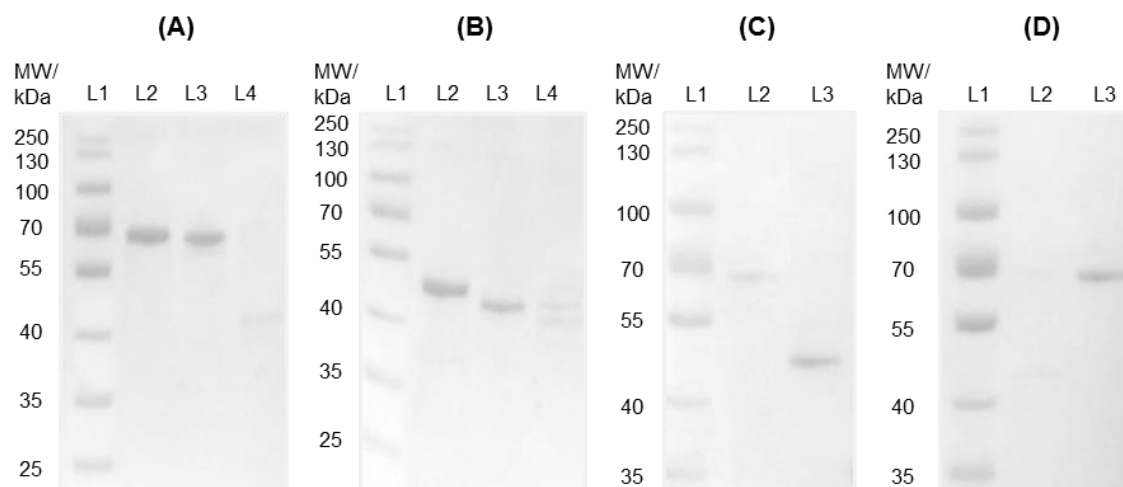


Figure 2.13: Left: SDS-PAGE analysis of (A) CpkG and (B) CpkC-TR, following cleavage of hexahistidine tags with TEV protease. L1: Protein molecular weight marker; L2: His₆-tagged protein standard; L3: Protein collected from wash fraction (20 mM imidazole); L4: Protein collected from elution fraction (300 mM imidazole). Right: SDS-PAGE analysis of fractions collected from pull-down assays, following incubation of (C) His₆-CpkC-TR with CpkG, and (D) His₆-CpkG with CpkC-TR. L1: Protein molecular weight marker; L2: protein collected from wash fraction; L3: protein collected from elution fraction.

Overall, both the native-PAGE and pull-down assays indicate that CpkC-TR and CpkG do not form a heterodimeric complex. This could possibly be explained by analysis of a multiple sequence alignment of CpkG with other putative ω -TAs likely involved in polyketide alkaloid biosynthesis (see section 2.4); part of the additional domain appears

to be missing (figure 2.14). Consequently, the domain may be non-functional, merely existing as an evolutionary relic, or may serve another purpose.

Therefore, it is possible that *Streptomyces coelicolor* utilise an alternative aldehyde detoxification mechanism. Indeed, CpkC-TR is capable of fully reducing ACP-tethered thioester **22** to alcohol **24**, via two NADH-mediated two-electron transfers.

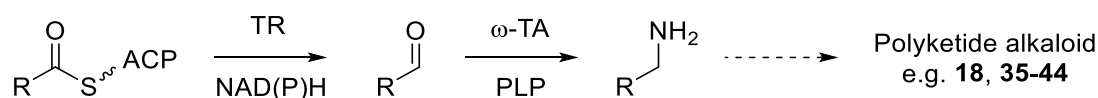


Figure 2.14: A partial display of a multiple sequence alignment of CpkG and homologues encoded within putative polyketide alkaloid biosynthetic gene clusters, produced using T-Coffee and BoxShade software. Darker regions indicate higher levels of homology between proteins. The additional domain of CpkG (residues 142-189) appears to contain fewer amino acids, which may have been necessary for interactions with CpkC-TR.

2.4 Further polyketide alkaloid specialised metabolites

2.4.1 Identification of polyketide alkaloid biosynthetic gene clusters

In addition to coelimycin P1 **18**, there are numerous other polyketide alkaloids, whose gene clusters and biosyntheses remain to be characterised. Examples include latumcidin **35**, nigrifactin **36** and pyrindicin **37** (see section 1.6). Following investigations with CpkC-TR and CpkG, it was proposed that thioester reduction and transamination may constitute a universal mechanism in diverting polyketide intermediates towards alkaloid metabolism (scheme 2.10).⁵⁷



Scheme 2.10: Proposed universal biosynthesis of actinobacterial polyketide alkaloids

To investigate this hypothesis, the GenBank database was examined for gene clusters containing genes encoding a CpkC-TR and CpkG homologue within 35 kb of each other (Dr. Emmanuel de los Santos, University of Warwick). Functional annotation of the sequences obtained using antiSMASH¹⁴⁹ resulted in the identification of 22 actinobacterial gene clusters that are hypothesised to direct the biosynthesis of polyketide alkaloids (figure 2.15).

2.4.2 Prediction of polyketide intermediates

To gain insight into the metabolic products of these gene clusters, the module and domain organisation of the encoded PKSs were analysed in an attempt to predict the structures of the polyketide intermediates. Two online databases, SBSPKS and the NCBI conserved domain search, were used for *in silico* analysis.^{44,150} Domain activities and specificities were then predicted manually by examination of key sequence motifs.

The activity of the acyl carrier protein (ACP) and ketosynthase (KS) domains within each PKS was determined by the presence of a conserved serine residue or a CHH catalytic triad, respectively.^{151,38} Loading KSs were identified by a cysteine-glutamine mutation, which suggests that they function as decarboxylases and that carboxylated acyl-CoA species are recruited as substrates.⁴⁷

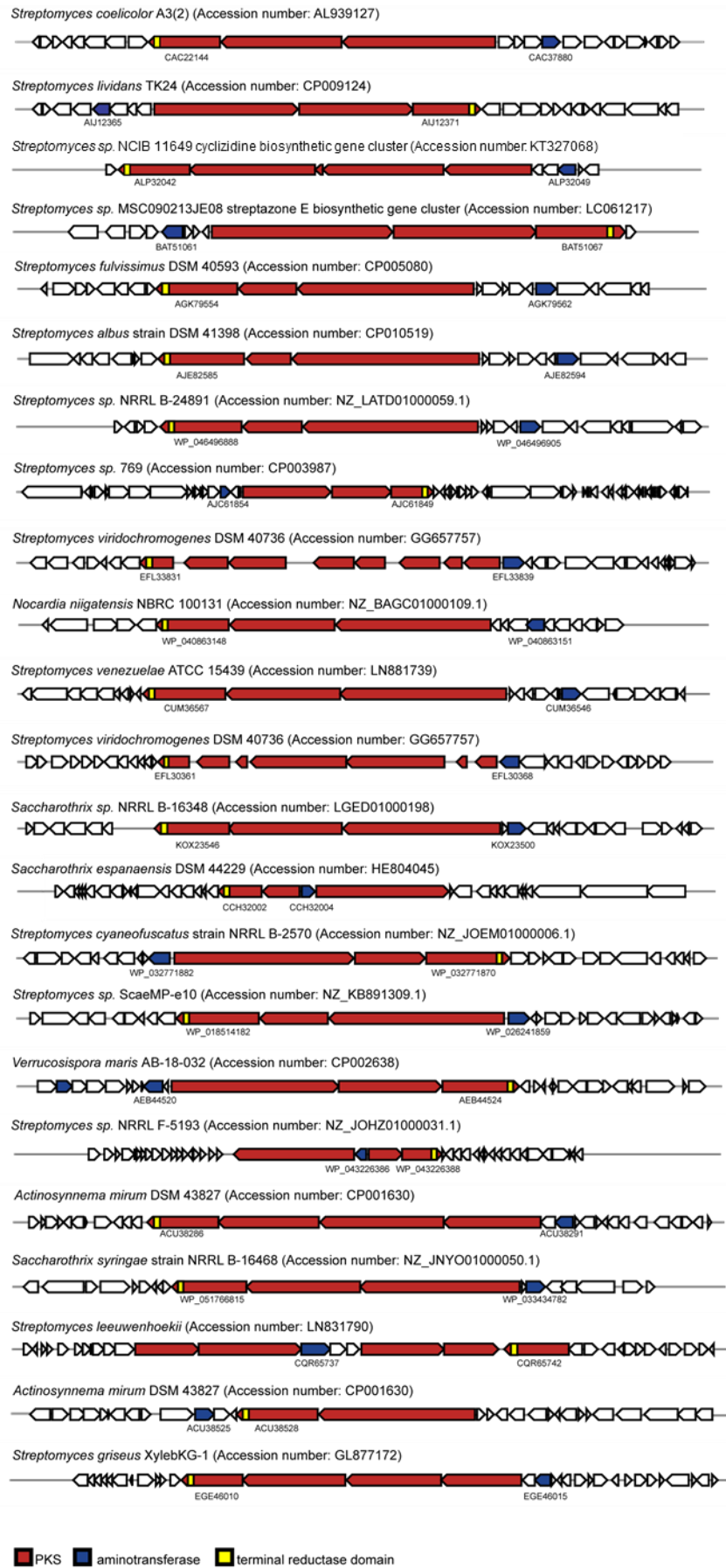


Figure 2.15: Overview of gene clusters harbouring PKS genes and homologues of *cpkC-TR* and *cpkG*⁵⁷

The substrate specificity of the acyltransferase (AT) domains was determined by sequence comparisons with AT domains of known specificity. In particular, the presence of a YASH or HAFH motif, and either a glutamine or branched non-polar residue within the conserved GHSxG motif, indicated the uptake of methylmalonyl- or malonyl-CoA, respectively.^{152,153} In the case of *Saccharothrix espanaensis* DSM 44229, an AT domain was missing from the loading module, and the AT domain present within the fourth module contained unusual RASH and GHSTG motifs, so predictions are tentative.

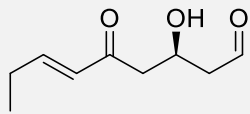
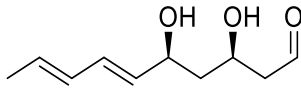
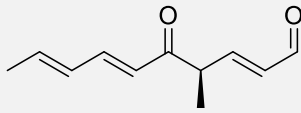
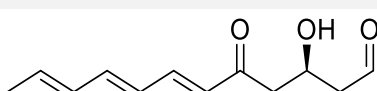
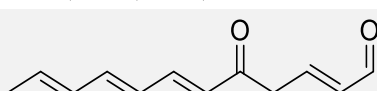
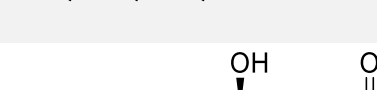
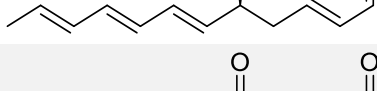
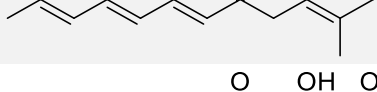
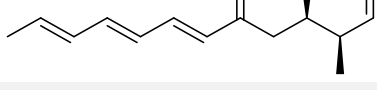
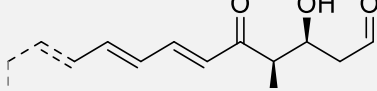
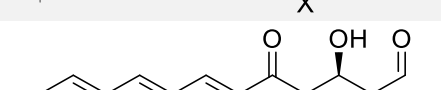

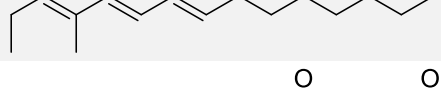
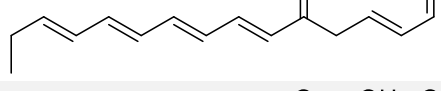
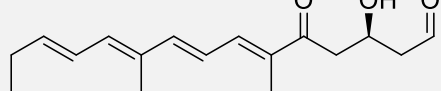
The activity and stereospecificity of the ketoreductase (KR) domains was determined by sequence comparisons with KR domains of known stereospecificity. All of the KR domains analysed contained a conserved LDD (or LDN) motif, indicating that they were B-type KR domains, and thereby suggesting that an (*R*)-configured β -hydroxyl group would be formed.¹⁵⁴ For branched intermediates (e.g. α -methyl- β -ketothioesters), the presence or absence of a proline residue in the catalytic region indicated that the α -substituent would have (*S*)- or (*R*)-stereochemistry, respectively.¹⁵⁵ In several cases, the KR domain was determined to be inactive due to the absence of a catalytic tyrosine or conserved asparagine residue.^{156,155}

The activity of the dehydratase (DH) domains was evaluated by analysing conserved active site residues and motifs. Specifically, considerable alterations to the four signature motifs (HxxxGxxxxP, GYxYGPxF, LPFxW and Dxxx(Q/H)) suggested that the DH domain would be rendered inactive.¹⁵⁷ For all active DH domains, the substrate double-bonds were predicted to have *trans*-geometry, since all β -hydroxyl groups had previously been assigned to have (*R*)-stereochemistry.³⁸

The stereospecificity of the enoyl reductase (ER) domains was predicted by sequence comparisons to ER domains of known stereospecificity. A conserved tyrosine (at position 52 in *E. coli* quinone oxidoreductase) indicated that the α -substituent would have (*S*)-stereochemistry, while a conserved alanine (or occasionally valine or phenylalanine) denoted (*R*)-stereochemistry.^{158,51}

Utilising this information, and assuming a co-linear relationship between the number of modules and the number of precursors used for chain assembly, the structures of the polyketide intermediates were predicted (table 2.1).

Table 2.1: Actinobacterial strain (left), predicted aldehyde intermediate (middle) and predicted corresponding metabolite (right)

<i>Actinosynnema mirum</i> DSM 43827		Novel
<i>Streptomyces fulvissimus</i> DSM 4053 <i>Streptomyces</i> sp. NRRL B-24891		35/ 38/ 41
<i>Streptomyces cyaneofuscatus</i> NRRL B-2570 <i>Streptomyces</i> sp. ScaeMP-e10 <i>Streptomyces viridochromogenes</i> (2)		Novel
<i>Nocardia niigatensis</i> NBRC 100131 <i>Streptomyces</i> sp. NRRL F-5193		18/ 36/ 42
<i>Saccharothrix syringae</i> NRRL B-16468 <i>Saccharothrix</i> sp. NRRL B-16348 <i>Verrucosipora maris</i> AB-18-032 <i>Streptomyces</i> sp. MSC090213JE08		36/ 42
<i>Streptomyces lividans</i> TK24		18
<i>Streptomyces venezuelae</i> ATCC 15439		37
<i>Streptomyces viridochromogenes</i> (1)		37
<i>Saccharothrix espanaensis</i> DSM 44229		Novel
<i>Streptomyces</i> sp. 769		Novel
<i>Streptomyces griseus</i> XylebKG-1		Novel
<i>Streptomyces albus</i> DSM 41398		Novel
<i>Actinosynnema mirum</i> DSM 43827 (2)		39/ 40
<i>Streptomyces</i> NCIB 11649		39/ 40
<i>Streptomyces leeuwenhoekii</i>		Novel

Notably, in each of these assembly lines, the penultimate module either lacks a ketoreductase or dehydratase domain, or harbours a non-functional equivalent. This would result in fully-assembled polyketides with a 5'-keto or hydroxyl group, suggesting that they are predisposed for intramolecular imine formation following transamination.

2.4.3 Prediction of polyketide alkaloid metabolites

Interestingly, thirteen of the predicted aldehyde intermediates were found to have the same carbon skeletons as several known polyketide alkaloids. Of these, two (from *Nocardia niigatensis* NBRC 100131 and *Streptomyces lividans* TK24) are believed to be converted to coelimycin P1 **18** or a derivative thereof, based on sequence similarity to the *cpk* cluster.

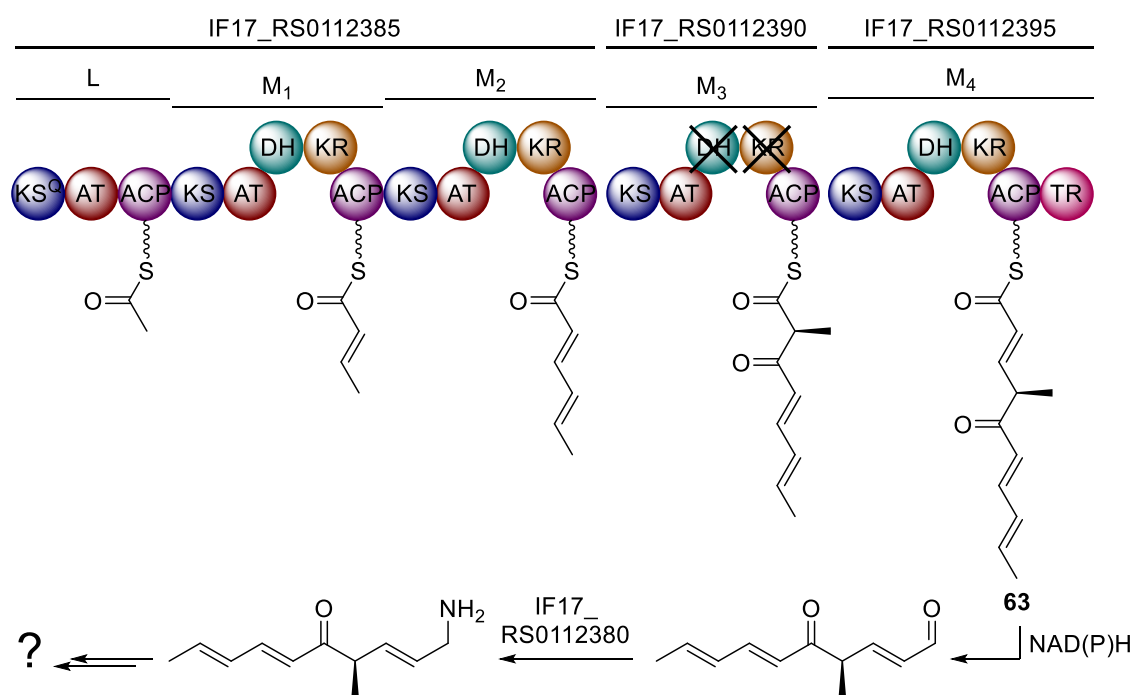
Two other gene clusters produce a polyketide chain that corresponds to the carbon skeleton of cyclizidine **39** or JBIR-102 **40**, while a further five clusters assemble an intermediate corresponding to the scaffold of streptazone E **42** or nigrifactin **36**. Encouragingly, two of these clusters, from *Streptomyces* sp. NCIB11649 and *Streptomyces* sp. MSC090213JE08, have since been shown to direct the production of cyclizidine and streptazone E, respectively,^{159,160} corroborating the robustness of the search algorithm and the reliability of the sequence-based predictions. The polyketide chains assembled by two of the other clusters correspond to the carbon skeleton of pyrindicin **37**, and two further clusters are proposed to direct the biosynthesis of latumcidin **35**, streptazolin **38**, or a closely related metabolite. This is based on the structure of the carbon chains assembled by the PKS and the presence of genes encoding homologues of the cyclases recently reported to be involved in streptazone E biosynthesis.¹⁶⁰ Routes to the formation of known polyketide alkaloids **35-39** and **42** from the corresponding predicted aldehyde intermediates have been tentatively proposed (see appendix 4).

The polyketide chains produced by the PKSs encoded by the remaining nine clusters do not correspond to the carbon skeletons of any known metabolites. Thus, these clusters are likely to direct the biosynthesis of novel polyketide alkaloids.

It is interesting to note that besides the CpkC-TR and CpkG homologues, the other putative tailoring enzymes encoded within the recently characterised cyclizidine and streptazone E biosynthetic gene clusters show little sequence similarity to those associated with the *cpk* cluster. This observation is consistent with the hypothesis that reductive chain release and transamination are a conserved set of enzymatic reactions in polyketide alkaloid biosynthesis, whereas downstream-acting tailoring enzymes are more divergent.

2.4.4 Attempted detection of putative polyketide alkaloid from *Streptomyces cyaneofuscatus* NRRL B-2570

In attempt to detect and characterise one of the predicted novel polyketide alkaloid metabolites, *Streptomyces cyaneofuscatus* NRRL B-2570 was acquired from the ARS culture collection. Analysis of the PKS module and domain organisation revealed that it is likely to assemble thioester **63**, which is then expected to undergo reductive cleavage followed by transamination, catalysed by IF17_RS0112395-TR and IF17_RS0112380, respectively (scheme 2.11). Six other putative post-PKS tailoring enzymes are homologous to oxidoreductases, monooxygenases and cytochrome P450 enzymes (see appendix 5), suggesting that several oxidative steps are likely to follow.



Scheme 2.11: Proposed biosynthesis of polyketide alkaloid intermediates in *Streptomyces cyaneofuscatus* NRRL B-2570

S. cyaneofuscatus NRRL B-2570 was first cultured on growth media (SFM agar) followed by production media (ISP2, R3 and R5 agar). Metabolites were then extracted and analysed by LC-MS (figure 2.16). However, many peaks were apparent and it was difficult to identify the putative polyketide alkaloid(s). Therefore, alternative metabolomics approaches were pursued.

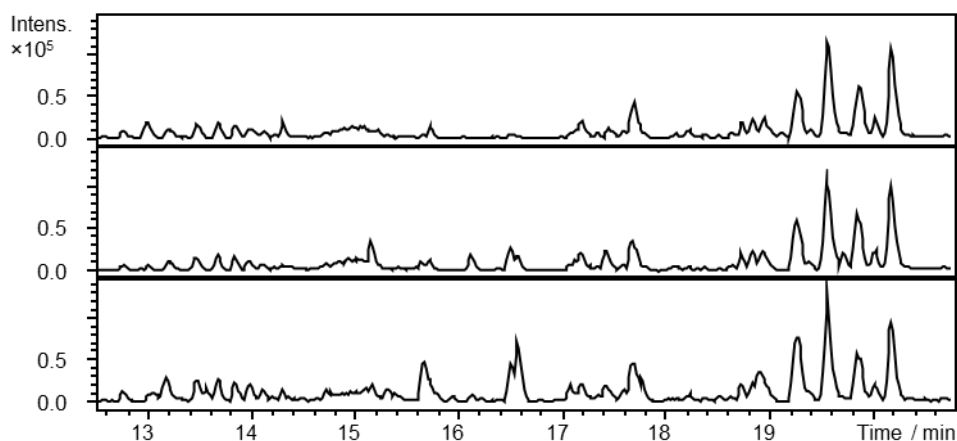


Figure 2.16: Base peak chromatograms from LC-MS analyses of *S. cyaneofuscatus* metabolites extracted from ISP2 (top), R3 (middle) and R5 (bottom) agar

2.4.4.1 Overexpression of a putative SARP

It was noted that the putative polyketide alkaloid cluster within *S. cyaneofuscatus* contained a gene (*IF17_RS0112435*) encoding a *Streptomyces* Antibiotic Regulatory Protein (SARP); a transcriptional activator with a high specificity for specialised metabolite production in actinomycetes.¹⁰⁸ SARPs comprise of a characteristic OmpR-type winged helix-turn-helix DNA binding domain and a bacterial transcriptional activator domain (BTAD), close to the *N*-terminus.¹⁶¹

There are many examples in which SARP overexpression leads to activation or overproduction of the corresponding specialised metabolite.^{162,163,164,165} Therefore, *IF17_RS0112435* was overexpressed in attempt to trigger the (over)production of the putative polyketide alkaloid metabolite(s).

2.4.4.1.1 Cloning of *IF17_RS0112435* into pOSV556t

Primers were designed to allow amplification of *IF17_RS0112435* by PCR and subsequent cloning into pOSV556t; an integrative plasmid under the control of the constitutive *ermE** promoter (figure 2.17). The plasmid contains antibiotic resistant markers for selection in *E. coli* (ampicillin, *amp^r*) and *Streptomyces* (hygromycin, *hyg^r*), in addition to an origin of transfer, *oriT*, which permits interspecies conjugation. A gene encoding an integrase, *int_pSAM2*, allows the plasmid to integrate into the *Streptomyces* chromosome through site-specific recombination between the *attP* plasmid site and *attB* chromosomal site.¹⁶⁶

The forward primer contained a *Hind*III restriction site and an artificial conserved ribosomal binding site, while the reverse primer contained an *Xho*I restriction site. *IF17_RS0112435* was amplified by PCR (figure 2.18a), and the DNA band of the expected size was excised from the gel and purified.

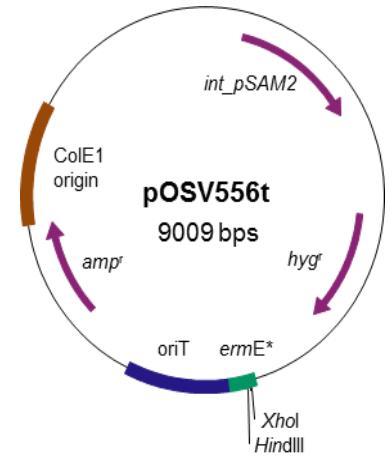


Figure 2.17: Plasmid map of pOSV556t. An integrase gene (*int_pSAM2*), origins of replication (ColE1) and transfer (*oriT*), ampicillin and hygromycin resistance genes (*amp^r* and *hyg^r*) and restriction sites (*Xho*I and *Hind*III) used for cloning are highlighted.

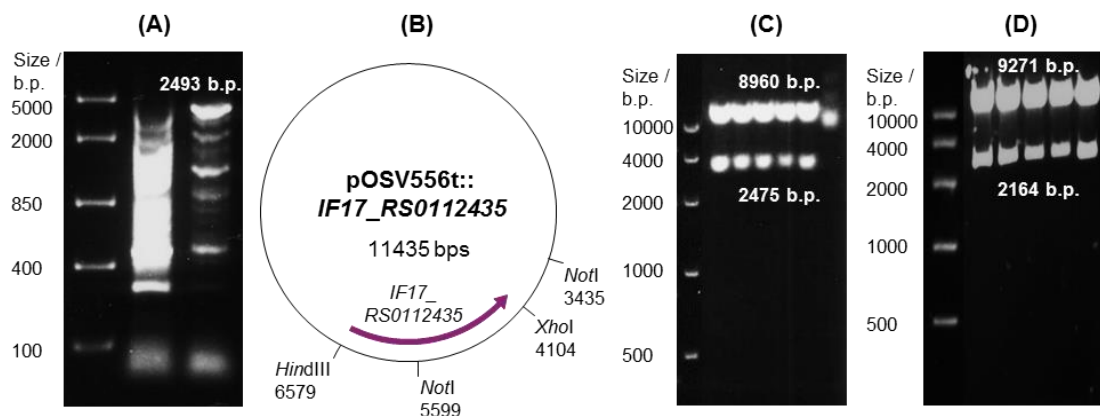


Figure 2.18: (A) Gel from agarose electrophoresis following PCR amplification of *IF17_RS0112435*; (B) plasmid map with restriction sites used for analysis displayed; (C) gel from agarose electrophoresis following restriction digest with *Hind*III and *Xho*I; and (D) gel from agarose electrophoresis following restriction digest with *Not*I.

The PCR product and plasmid were then digested similarly with *Hind*III and *Xho*I, purified and ligated together. The ligation mixture was subsequently used to transform *E. coli* TOP10 cells. Colonies containing the recombinant plasmid were selected for with ampicillin, and several were cultured overnight for plasmid DNA purification. Correct insertion of *IF17_RS0112435* was examined by restriction digestion (*Hind*III, *Xho*I and *Not*I, figure 2.18b-d). Positive clones were confirmed by sequencing.

2.4.4.1.2 Conjugation of pOSV556t::*IF17_RS0112435* into *S. cyaneofuscatus*

The construct (pOSV556t::*IF17_RS0112435*) was used to transform methylation-deficient *E. coli* ET12567/pUZ8002 cells. Transformants were then mixed with *S. cyaneofuscatus* NRRL B-2570 spores to promote conjugation, and cultured on SFM agar supplemented with MgCl₂. The following day the plates were overlaid with nalidixic acid and hygromycin, to selectively kill *E. coli* ET12567/pUZ8002 and wild-type *S. cyaneofuscatus* (not harbouring the plasmid), respectively. The plates were then incubated for 1 week and transconjugants were selected and cultured on SFM agar supplemented with hygromycin, to obtain spore stock solutions of the recombinant strain, denoted *S. cyaneofuscatus*::*SARP*.

2.4.4.2 Inactivation of the PKS

In addition to *SARP* overexpression, a gene encoding part of the PKS (*IF17_RS0112385*) was disrupted, to aid in the deduction of the metabolite(s) associated with the putative polyketide alkaloid cluster.

2.4.4.2.1 Cloning of *IF17_RS0112385* into pKC1132

Primers were designed to target a region encoding the ACP of module 1 and the KS of module 2 in the PKS. Primers contained a *Hind*III and *Xba*I restriction site to facilitate cloning into pKC1132; a 3.5 kb conjugative suicide vector bearing an apramycin resistance gene (*am^r*) and *oriT* from RK2 (figure 2.19).¹⁶⁷

Following amplification of the sequence (figure 2.20a), the PCR product was purified and digested alongside the plasmid. The digested PCR product and plasmid were then ligated and the mixture was used to transform *E. coli* TOP10 cells. Colonies containing the recombinant plasmid were selected for with apramycin, and several were cultured overnight for plasmid DNA purification. Correct insertion of *IF17_RS0112385* was examined by restriction digestion (*Hind*III, *Xba*I and *Sma*I, figure 2.20b-d). Positive clones were confirmed by sequencing.

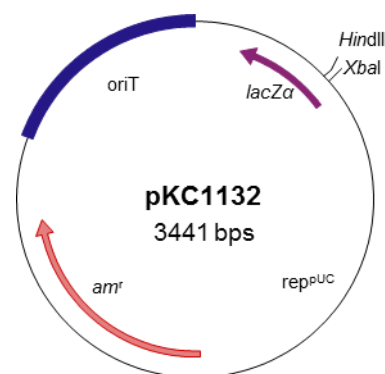


Figure 2.19: Plasmid map of pKC1132. The *lacZα* gene, an apramycin resistance gene (*am^R*), origins of replication (*rep^{pUC}*) and transfer (*oriT*) and restriction sites used for cloning (*Hind*III and *Xba*I) are highlighted.

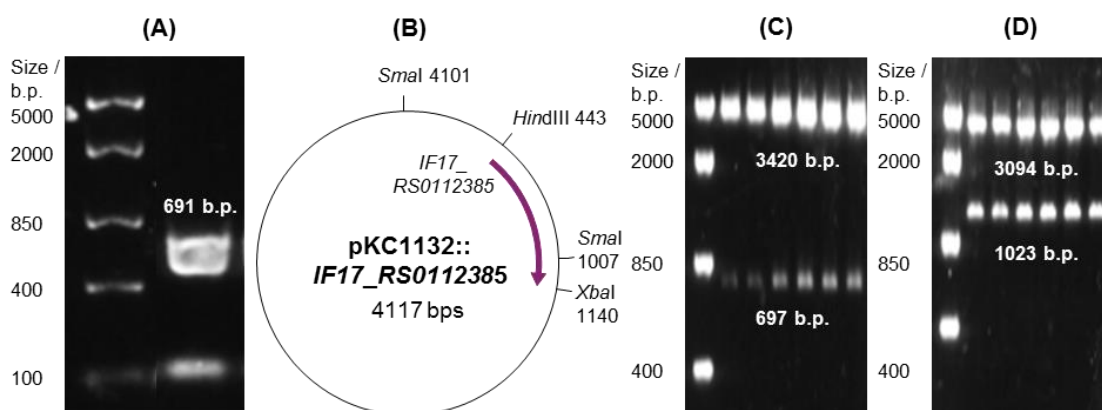


Figure 2.20: (A) Gel from agarose electrophoresis following PCR amplification of *IF17_RS0112385*; (B) plasmid map with restriction sites used for analysis displayed; (C) gel from agarose electrophoresis following restriction digest with *Hind*III and *Xba*I; and (D) gel from agarose electrophoresis following restriction digest with *Sma*I.

2.4.4.2.2 Conjugation of pKC1132::*IF17_RS0112385* into *S. cyaneofuscat*

Transformations of ET12567/pUZ8002 cells with the plasmid clone (pKC1132::*IF17_RS0112385*) and subsequent conjugations were performed as described in section 2.4.4.1.2. However, plates were overlaid with apramycin rather than hygromycin, to select for exconjugants. As pKC1132 is not capable of replicating independently in *Streptomyces*, apramycin-resistant exconjugants could only be produced *via* a single crossover event, resulting in disruption of the PKS gene.

Two conjugations were performed; one to disrupt the PKS gene in the wild-type strain, and one to disrupt the PKS gene in *S. cyaneofuscatus*::SARP, to generate the strains *S. cyaneofuscatus*PKS::am and *S. cyaneofuscatus*::SARP-PKS::am, respectively.

2.4.4.3 Comparative metabolic profiling

To investigate whether SARP overexpression or PKS disruption would induce or abolish metabolite production, respectively, *S. cyaneofuscatus* wild-type and recombinant strains (*S. cyaneofuscatus*::SARP, *S. cyaneofuscatus*PKS::am and *S. cyaneofuscatus*::SARP-PKS::am) were cultured on ISP2, R3 and R5 agar. Metabolites were then extracted and analysed by LC-MS.

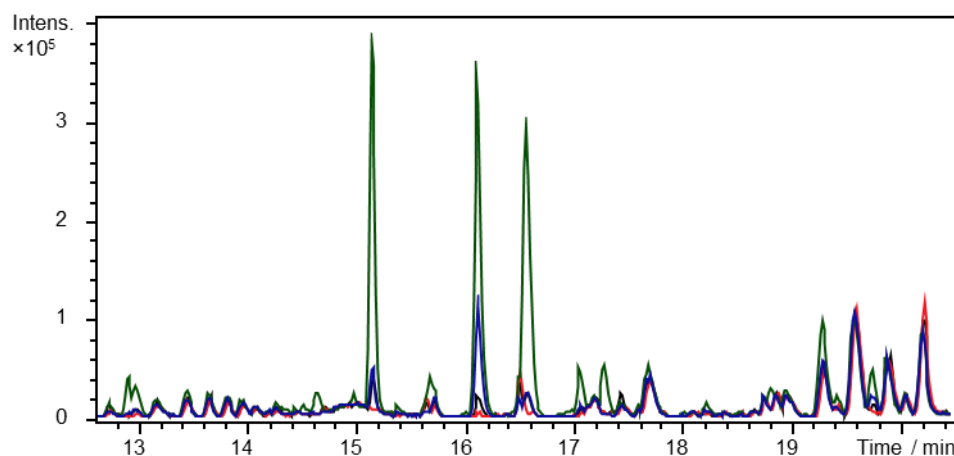


Figure 2.21: Base peak chromatograms from LC-MS analyses of *S. cyaneofuscatus* metabolites extracted from R3 agar. Black trace: wild-type *S. cyaneofuscatus*; red trace: *S. cyaneofuscatus*PKS::am; green trace: *S. cyaneofuscatus*::SARP; blue trace: *S. cyaneofuscatus*::SARP-PKS::am.

In all cases, SARP overexpression appeared to induce the production of three metabolites eluting at 15.2, 16.1 and 16.6 minutes. This was most pronounced using R3 media (figure 2.21). Analysis of the peaks revealed that the metabolites had m/z values of 240.1202, 312.1416 and 411.2103 and calculated molecular formulae of $C_{10}H_{19}NNaO_4$, $C_{13}H_{22}NNaO_6$ and $C_{18}H_{32}N_2NaO_7$, respectively. These formulae seemed promising; all contained one or two nitrogen atoms, predicted to derive from a transamination reaction and possibly an amino acid addition, as in coelimycin P1 biosynthesis. Furthermore, all contained multiple oxygen atoms, which could result from post-PKS tailoring oxidation reactions. For example, the genes *IF17_RS0112400*, *IF17_RS0112425* and

IF17_RS0112430 are hypothesised to encode a putative FAD-dependent monooxygenase, an amine oxidase and a cytochrome P450 enzyme, respectively.

Interestingly, while the corresponding peaks for the PKS disruption mutants were of a much lower intensity, they were still evident. This was assumed to be attributable to homologous recombination, which could result in reformation of the wild-type strain, as mutants were grown in the absence of apramycin.

2.4.4.4 Overproduction, purification and characterisation of putative polyketide alkaloid metabolite(s)

In attempt to overproduce and purify the putative polyketide alkaloid(s), *S. cyaneofuscatus::SARP* spores were cultured on R3 agar. Metabolites were then extracted, dried and separated by HPLC (in collaboration with Dr. Yousef Dashti, University of Warwick). Fractions with m/z values of 240.1202 and 411.2103 were purified sufficiently to obtain NMR data. The compound with an m/z value of 312.1416 was not obtained in adequate yield.

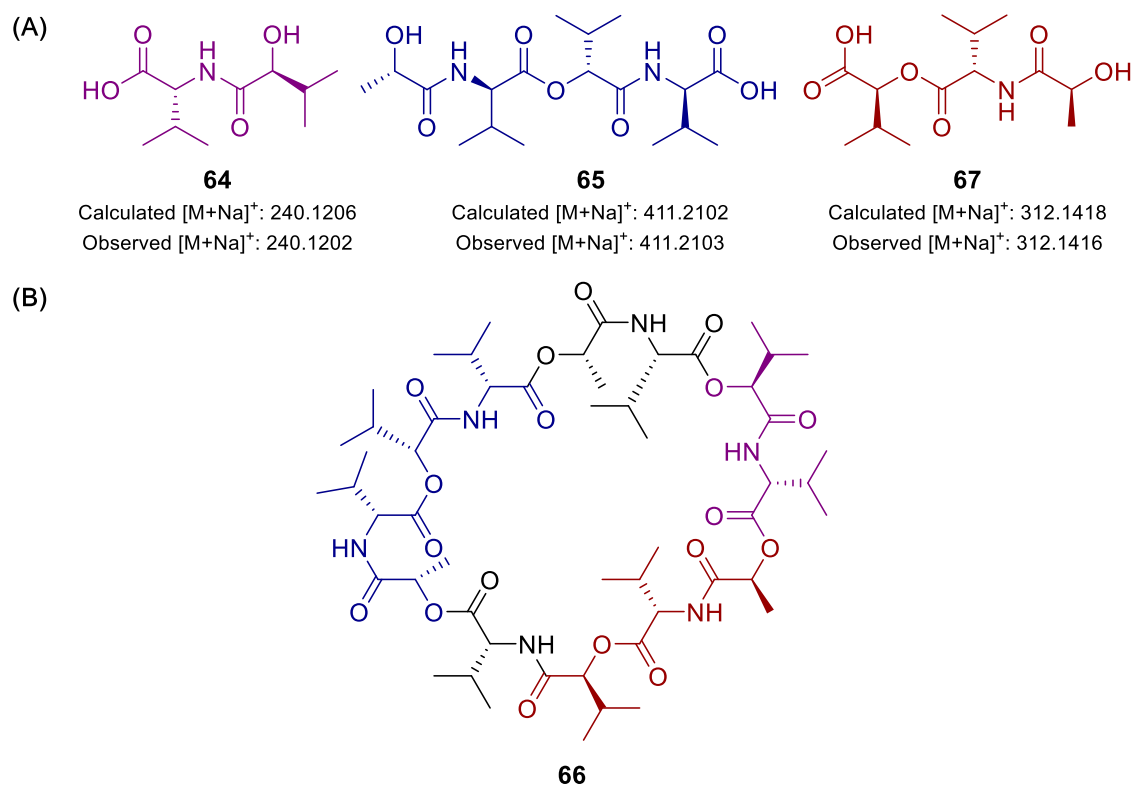


Figure 2.22: (A) Characterised (**64** and **65**) and predicted (**67**) structures of *S. cyaneofuscatus::SARP* metabolite fragments; and (B) structure of valinomycin **66** with isolated and predicted fragments highlighted.

Unfortunately, the structures of the two metabolites were unambiguously determined to be (depsi)peptides **64** and **65** (figure 2.22a). Both compounds appear to be fragments of valinomycin **66** (figure 2.22b), an antibiotic produced by at least eleven *Streptomyces* strains.¹⁶⁸ Indeed, upon searching the genome of *S. cyaneofuscatus* NRRL B-2570, a cluster was determined to be nearly identical to the valinomycin biosynthetic gene cluster in *S. tsusimaensis* ATCC 15141 (figure 2.23).¹⁶⁹ The compound with an observed m/z value of 312.1416 is thus likely to be another valinomycin fragment, **67** (figure 2.22a).

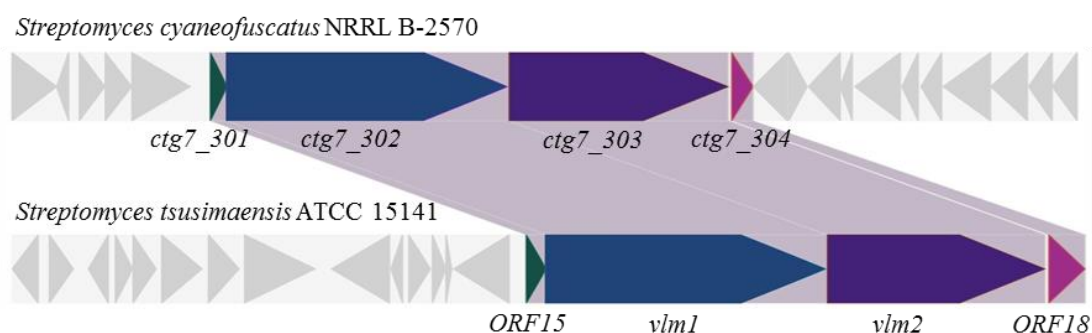


Figure 2.23: Comparison of predicted (top) and known (bottom) valinomycin biosynthetic gene clusters in *Streptomyces cyaneofuscatus* NRRL B-2570 and *Streptomyces tsusimaensis* ATCC 15141¹⁶⁹, respectively (Dr. Emmanuel de los Santos, University of Warwick). Homologous genes are coloured similarly, and are proposed to encode a type II thioesterase (green), NRPS modules 1 and 2 (blue), NRPS modules 3 and 4 (purple) and a transposase (pink).

Hydrolysis of valinomycin likely occurred during the mildly acidic work-up; extractions with ethyl acetate alone did not yield the three metabolites. Indeed, cultivation of *S. cyaneofuscatus*::SARP in AM1 and SMM media, followed by neutral extractions, revealed an additional peak in the LC-MS trace that was absent from the wild-type strain (figure 2.24a). This metabolite had the same molecular formula as valinomycin (figure 2.24b).

It seems interesting that valinomycin production was augmented by overexpressing the SARP from a different gene cluster. However, other examples of SARP-mediated cross-regulation between pathways have been reported, prompting such regulators to be termed ‘cluster-situated’ rather than ‘pathway-specific’. For example, Huang and coworkers showed that induction of *redZ*, a gene encoding a SARP within the undecylprodigiosin (Red) cluster, resulted in activation of the Red, calcium-dependent antibiotic (CDA) and actinorhodin (Act) pathways in *S. coelicolor* A3(2). Deletion of *redZ* abolished expression of *red* genes, but also delayed expression of *cda* and *act* genes.¹⁷⁰

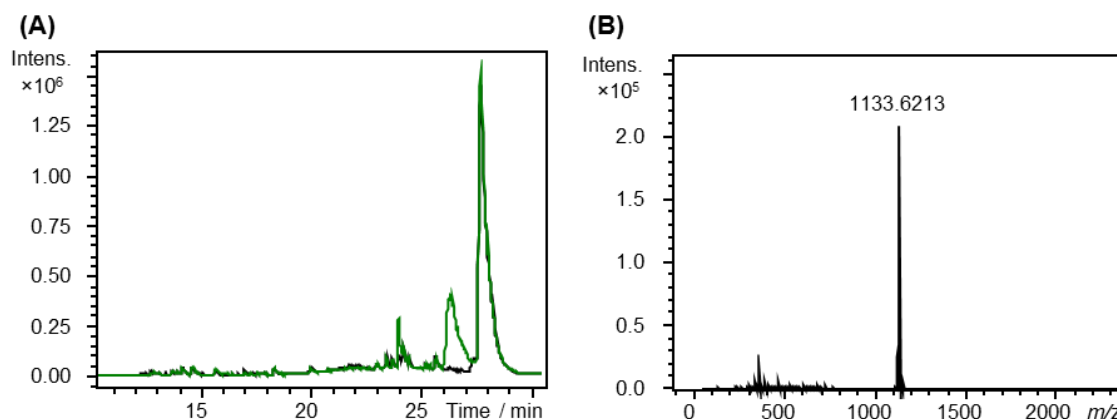


Figure 2.24: (A) Base peak chromatograms from LC-MS analyses of *S. cyaneofuscatus* metabolites extracted from AM1 liquid media. Black trace: wild-type *S. cyaneofuscatus*; green trace: *S. cyaneofuscatus*::*SARP*. (B) Observed $[M+Na]^+$ for metabolite eluting at 26.5 min, likely corresponding to valinomycin (calculated $[M+Na]^+ = 1133.6204$).

Thus far, further attempts to detect the putative polyketide alkaloid metabolite using various media and extraction procedures have been unsuccessful. This could be due to poor expression of the gene cluster under laboratory conditions; certain activating signals may be necessary to effect transcription or translation. Therefore, other approaches could be trialled, such as co-cultivation, deletion of putative repressor genes or promoter exchange.¹⁷¹ Current endeavours within the Challis group are exploring transformation-associated-recombination (TAR) cloning, in attempt to capture the gene cluster for expression in a heterologous host.

2.4.5 Whole genome sequencing of *Streptomyces abikoensis*, a latumcidin producer

In collaboration with Dr. Emmanuel de los Santos, University of Warwick

While bioinformatic searches were underway, *Streptomyces abikoensis* DSM 40831, a known producer of the polyketide alkaloid latumcidin **35**, was acquired from the DSMZ culture collection. It was thought that sequencing the genome could provide a complementary approach to aid in establishing the biosynthetic origin of such metabolites.

Therefore, *S. abikoensis* was cultured and its genomic DNA extracted and purified. The DNA was then sequenced by The Genome Analysis Centre (TGAC) using Pacific Biosciences (PacBio) technologies. 1-15 kb reads were then assembled into a single

contig to yield a draft genome consisting of 7.5 Mbp, with a GC content of 72.2 %. Annotation and mining of the genome revealed the presence of 74 putative specialised metabolite gene clusters, predicted to direct the biosynthesis of a range of PKs, NRPs, RiPPs (ribosomally synthesised and post-translationally modified peptides), terpenes, fatty acids and siderophores (Dr. Emmanuel de los Santos, see appendices 6 and 7).

As expected, a gene cluster encoding a CpkC-TR and CpkG homologue, in addition to a type I modular PKS, was identified (figure 2.25). Analysis of the PKS module and domain organisation (see section 2.4.2) led to the prediction of aldehyde intermediate **68**, which corresponds to the carbon skeleton of latumcidin (figure 2.26). Production of latumcidin by the strain was thereafter reconfirmed by high-resolution LC-MS analysis (Dr. Yousef Dashti and Dr. Daniel Zabala, University of Warwick, see appendix 8). Thus, it seems likely that this gene cluster directs the biosynthesis of latumcidin, although this would need to be corroborated through targeted gene deletions and metabolic profiling.

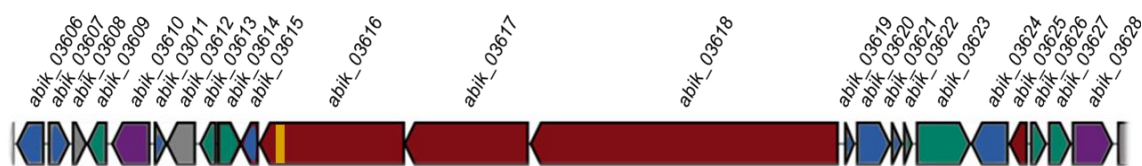


Figure 2.25: The organisation of a gene cluster within *Streptomyces abikoensis* DSM 40831, predicted to direct the biosynthesis of latumcidin **35**. The proposed functions of the proteins encoded by the genes are as follows. Purple: precursor supply; red: PKS component (with TR domain highlighted in yellow); blue: post-PKS tailoring; green: regulatory; grey: hypothetical.

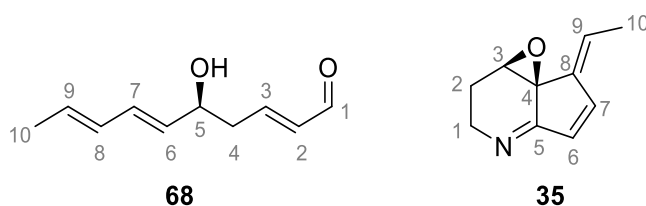
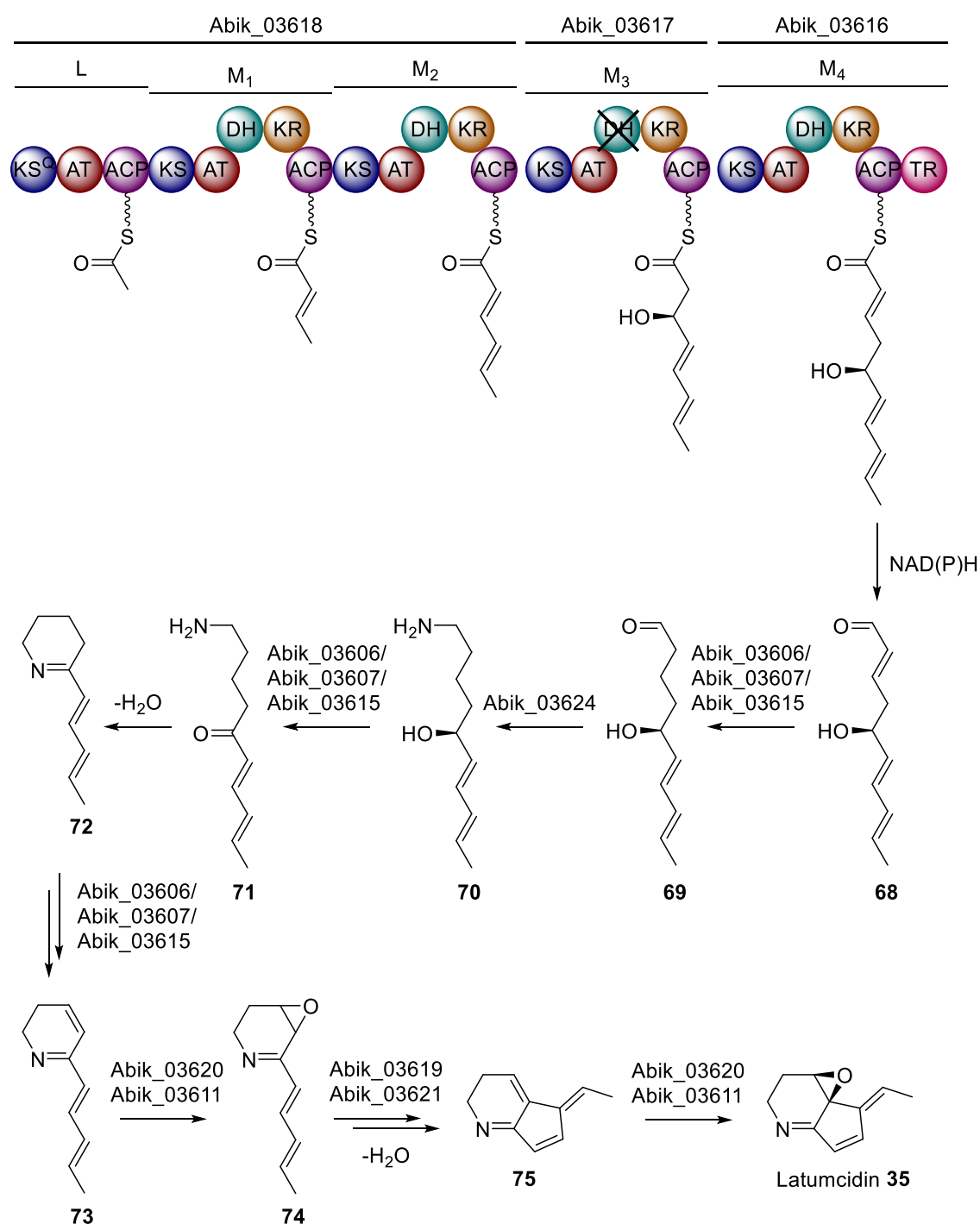


Figure 2.26: Structures of the predicted polyketide aldehyde **68** and latumcidin **35**, with carbon atoms numbered accordingly

Many of the biosynthetic genes within the cluster encode homologues of streptazone E enzymes (see appendix 7). For instance, *abik_03611* and *abik_03620* encode homologues of the putative flavin-dependent monooxygenase and flavin-reductase, respectively, while *abik_03619* and *abik_03621* encode homologues of the two putative cyclases. This led to the hypothesis that the bicyclic scaffold is formed through a similar mechanism to

that proposed by Ohno and coworkers.¹⁶⁰ The remaining three biosynthetic post-PKS tailoring genes (*abik_03606*, *abik_03607* and *abik_03615*) are predicted to encode short-chain dehydrogenases. Utilising this information, a biosynthetic route to latumcidin formation was proposed (scheme 2.12).



Scheme 2.12: A proposed biosynthesis of latumcidin **35**, based on the predicted functions of genes within cluster 43 of *Streptomyces abikoensis* DSM 40831 and the pathways proposed for streptazone E¹⁶⁰ **42** and coelimycin P1³⁰ **18** biosynthesis

Initially, the type I modular PKS likely assembles a pentaketide chain, which is then reductively cleaved by Abik_03616-TR. The *trans*-2,3-double bond of the resulting aldehyde **68** is then likely reduced, possibly by putative dehydrogenases Abik_03606, Abik_03607 or Abik_03615. Interestingly, the gene cluster does not encode a homologue of CpkE, the putative isomerase in coelimycin biosynthesis, suggesting that the double bond is not isomerised. Intermediate **69** then likely undergoes transamination, catalysed by Abik_03624, followed by oxidation, catalysed by one of the aforementioned dehydrogenases. Spontaneous intramolecular cyclisation and a further dehydrogenation would yield intermediate **73**. An epoxidation, catalysed by flavin redox partners Abik_03620 and Abik_03611, and cyclisation, catalysed by cyclases Abik_03619 and Abik_03621, could then ensue. Dehydration, followed by a final epoxidation, would then afford latumcidin **35**. Because only one putative monooxygenase and flavin reductase are encoded within the cluster, it is possible that Abik_03620 and Abik_03611 also catalyse the second epoxidation. This is plausible, as the epoxidation is occurring across the same double bond.

Many of the regulatory genes within the putative latumcidin cluster are homologous to those of the coelimycin cluster (see appendix 7), suggesting a similar mode of transcriptional control. For example, *abik_03609* and *abik_03623* encode homologues of the SARPs CpkN and CpkO, while *abik_03613* and *abik_03626* encode homologues of the TetR-like repressors ScbR and ScbR2, respectively. In addition, *abik_03614* encodes a butenoloid synthase, while *abik_03627* encodes a two-component system histidine-kinase, homologous to CpkM. The function of the putative regulatory gene *abik_03622* is unclear.

Overall, bioinformatics predictions strongly support that this gene cluster directs the biosynthesis of latumcidin; future work could confirm this through *in vivo* gene-knockouts. Furthermore, the hidden biosynthetic potential encoded within the remaining 73 gene clusters could be explored.

2.5 Summary

In summary, CpkG was demonstrated to have transaminase activity and a broad substrate tolerance, readily catalysing the interconversion of a range of primary amines with

aldehydes. Assays with synthetic substrate mimics of proposed intermediates in the *cpk* biosynthetic pathway demonstrated that CpkG is responsible for nitrogen incorporation into coelimycin. Crystal structures of CpkG allowed for the identification of the PLP binding site, key active site residues, a rare additional domain, and hinted at the transamination mechanism. The characteristic fold-type I conformation, coupled with Marfey's analysis, established CpkG as an (*S*)-selective ω -TA. Future work could investigate the biocatalytic capability of CpkG to generate chiral amines from ketones; a key research priority identified by the pharmaceutical industry.¹⁷²

Bioinformatic searches revealed 22 actinobacterial biosynthetic gene clusters encoding homologues of CpkC-TR and CpkG, that are predicted to direct the biosynthesis of both known and novel polyketide alkaloids. Attempts to detect one such metabolite from *S. cyaneofuscatu*s NRRL B-2570 were unsuccessful, instead resulting in the inadvertent overproduction of valinomycin. Other strategies, such as TAR cloning, may prove more effective. Moreover, the characterisation of TR and ω -TA enzymes within these systems could lead to promising alternative methods for the efficient and biocatalytic synthesis of polyketide alkaloids in the near future.

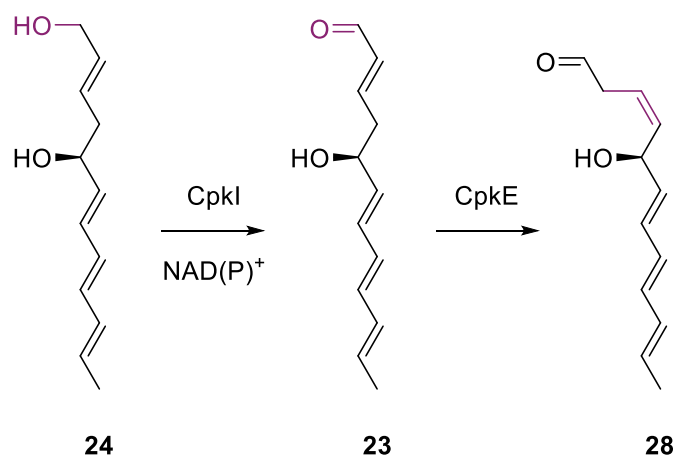
Sequencing the genome of the known latumcidin producer, *S. abikoensis* DSM 40831, enabled the discovery of a gene cluster containing *cpkC-TR* and *cpkG* homologues, as predicted. Together with the recent characterisation of the streptazone E and cyclizidine clusters, these findings strongly support the hypothesis that thioester reduction and transamination are a universal sequence of reactions in polyketide alkaloid biosynthesis. Future work could involve the generation of gene-knockouts or the purification of recombinant enzymes to aid in the elucidation of latumcidin biosynthesis. In addition, genome mining could accommodate the prediction and identification of further interesting specialised metabolites from *S. abikoensis*.

3 RESULTS AND DISCUSSION II: Characterisation of post-PKS tailoring enzymes in coelimycin biosynthesis

3.1 CpkI and CpkE

3.1.1 Introduction

CpkI, which shows sequence similarity to known short-chain dehydrogenases (SDHs), could catalyse the re-oxidation of alcohol **24** to aldehyde **23** in coelimycin biosynthesis. CpkE, belonging to the α,β -hydrolase fold superfamily of enzymes, has been tentatively suggested to catalyse a subsequent isomerisation of the *trans*-configured 2,3-double bond to a *cis*-configured 3,4-double bond, to afford intermediate **28** (scheme 3.1).³⁰ To investigate these possibilities, it was necessary to clone and overexpress the biosynthetic genes and purify the corresponding proteins.



Scheme 3.1: Proposed reactions catalysed by CpkI and CpkE in coelimycin biosynthesis³⁰

3.1.2 Cloning, overproduction and purification of proteins

Primers were designed to allow for directional cloning of *cpkI* and *cpkE* into the pET151/D-TOPO expression vector (figure 3.1). A four-base overhang (5'-CACC) was introduced in the forward primer to enable the PCR product to anneal to the overhang (5'-GGTG) on the complementary strand of the vector in the correct orientation. Both the forward and reverse primers contained 16-25 nucleotides of the coding sequences of *cpkI* or *cpkE*, including the start and stop codons, as appropriate.

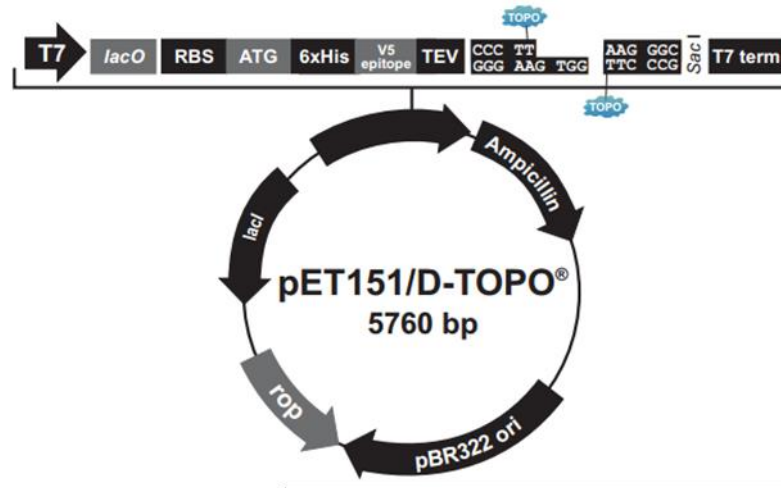


Figure 3.1: Map of the pET151/D-TOPO vector used for the assembly of protein expression constructs.¹⁷³ Key features (clockwise): T7 promoter; *lac* operator (*lacO*); ribosome binding site (RBS); start codon (ATG); *N*-terminal His₆-tag; V5 epitope; TEV recognition site; TOPO® cloning site; *SacI* restriction site; T7 transcription termination region; ampicillin resistance gene (β -lactamase); pBR322 origin of replication (*ori*); ROP gene; and *lacI* gene.

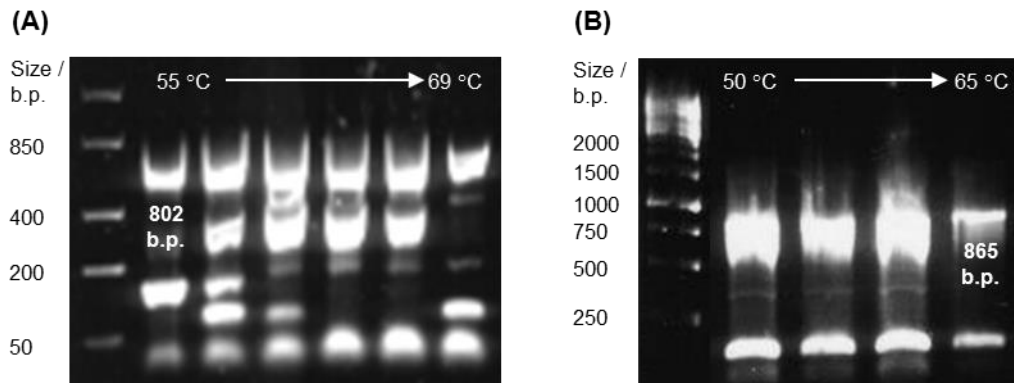


Figure 3.2: Gels from agarose electrophoresis following PCR amplifications of (A) *cpkI*; and (B) *cpkE*, at selected annealing temperatures. The expected sizes of the PCR products are labelled.

Using these primers, the genes were amplified by PCR (figure 3.2) and ligated with pET151/D-TOPO. The ligation mixtures were then used to transform *E. coli* One Shot® TOP10 cells. Colonies harbouring the plasmid were selected for with ampicillin and several were cultured overnight for plasmid DNA isolation. Correct insertion of the genes was then determined by restriction digestion using selected endonucleases (figure 3.3a). 8/12 *cpkI* and 7/12 *cpkE* clones appeared to be positive (figure 3.3b) and several were sequenced, to ensure that mutations had not been introduced during PCR.

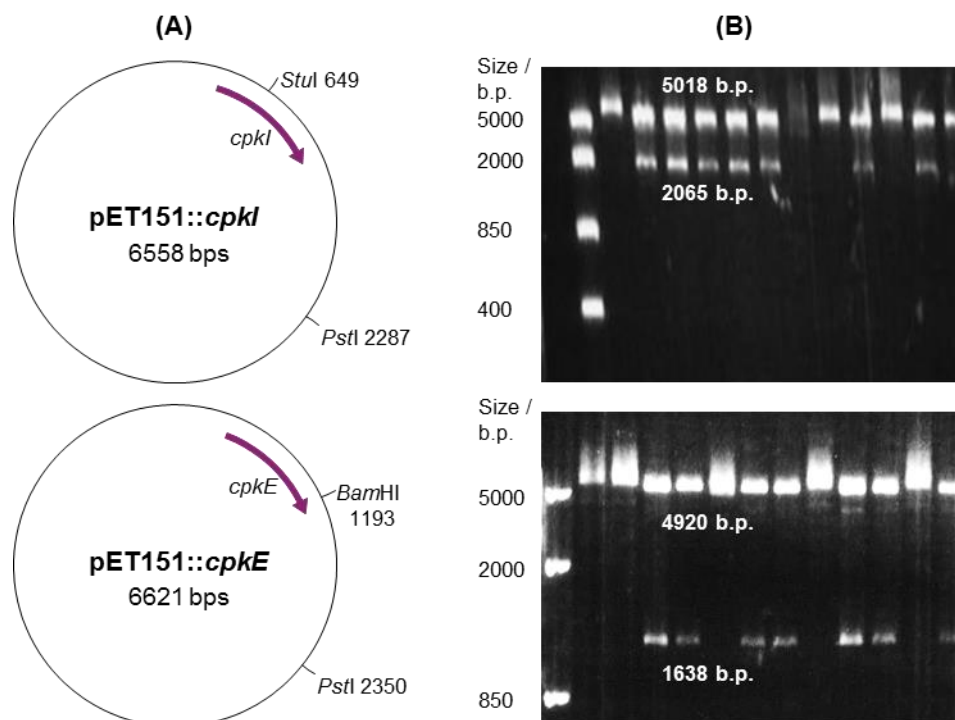


Figure 3.3: (A) Plasmid maps with restriction sites used for analysis; and (B) gels from agarose electrophoresis following restriction digests, of pET151::cpkI (top) and pET151::cpkE (bottom). The expected sizes of the DNA fragments are labelled.

Following confirmation of the sequences of the inserts, the constructs were transferred to *E. coli* One Shot® BL21 Star™ (DE3); a strain suitable for high-level recombinant protein expression. Subsequently, *cpkI* and *cpkE* were overexpressed and the corresponding *N*-terminal His₆-fusion proteins purified by nickel affinity chromatography, as described in section 2.1.2.

SDS-PAGE analysis indicated that the proteins were pure and of the expected molecular weights (figure 3.4). The exact molecular weights of the proteins were later verified by HRMS using maximum entropy deconvolution algorithms (figure 3.5; calculated molecular weights of His₆-CpkI and His₆-CpkE are 31214 and 34813 Da, and respectively).



Figure 3.4: SDS-PAGE analysis of purified His₆-CpkI (L2) and His₆-CpkE (L3) alongside a protein molecular weight marker (L1)

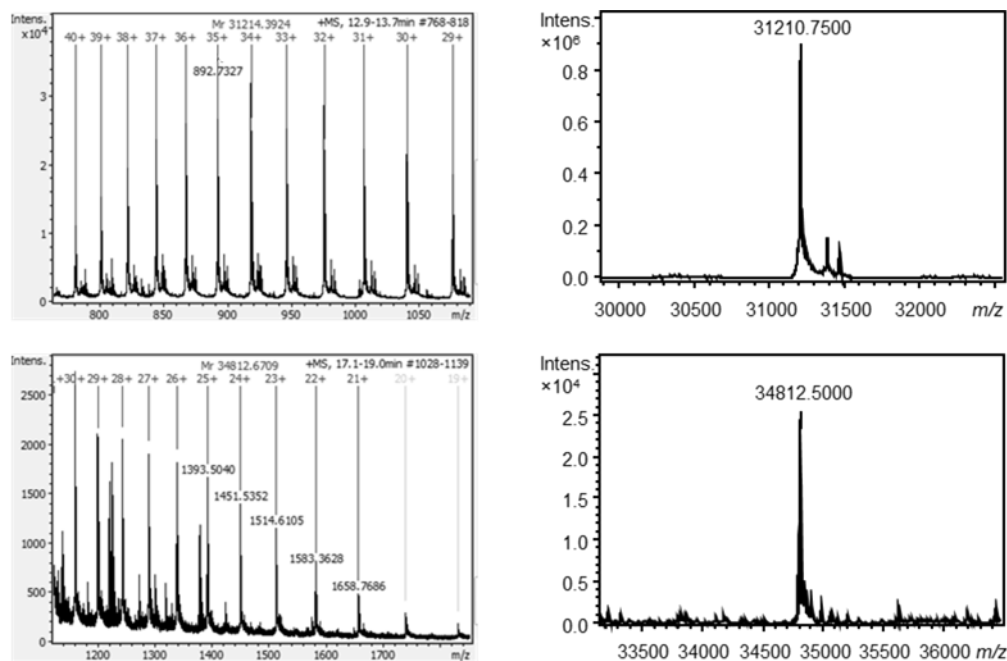
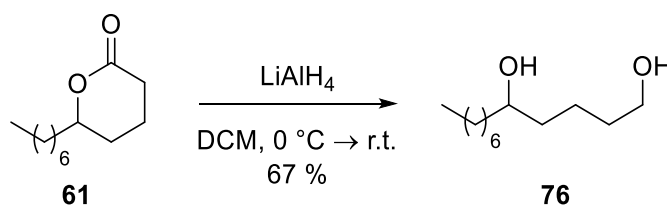


Figure 3.5: Measured (left) and deconvoluted (right) mass spectra of His₆-CpkI (top) and His₆-CpkE (bottom). The calculated molecular weights of the proteins are 31214 and 34813 Da, respectively.

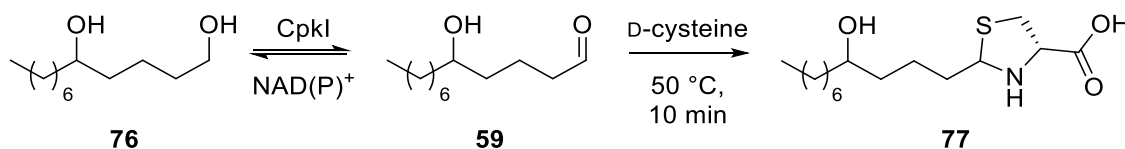
3.1.3 Preliminary *in vitro* assays

To investigate the hypothesised oxidoreductase activity of CpkI, a substrate analogue (dodecane-1,5-diol, **76**), containing the same 1,5-diol functionality and carbon chain length as the proposed true intermediate of the coelimycin biosynthetic pathway (**24**), was synthesised. This involved reducing commercially-available δ -dodecalactone **61** with two equivalents of LiAlH₄ (scheme 3.2).



Scheme 3.2: Synthesis of dodecane-1,5-diol **76** by reduction of δ -dodecalactone **61** (1.0 eq.) with LiAlH₄ (1.5 eq.)

Subsequently, dodecane-1,5-diol **76** was incubated with His₆-CpkI and either NAD⁺ or NADP⁺ as a cofactor. Any 5-hydroxydodecanal **59** produced during the reaction was then derivatised with D-cysteine (scheme 3.3) and analysed by LC-MS alongside a similarly derivatised authentic standard. Unfortunately, a peak corresponding to the molecular weight of (4S)-2-(4-hydroxyhexyl)thiazolidine-4-carboxylic acid **77** could not be detected (figure 3.6a).



Scheme 3.3: Assay employed to probe the possible oxidation of dodecane-1,5-diol **76** (200 μM) by CpkI (90 μM) and NAD(P)⁺ (200 μM), followed by derivatisation with D-cysteine (400 μM)

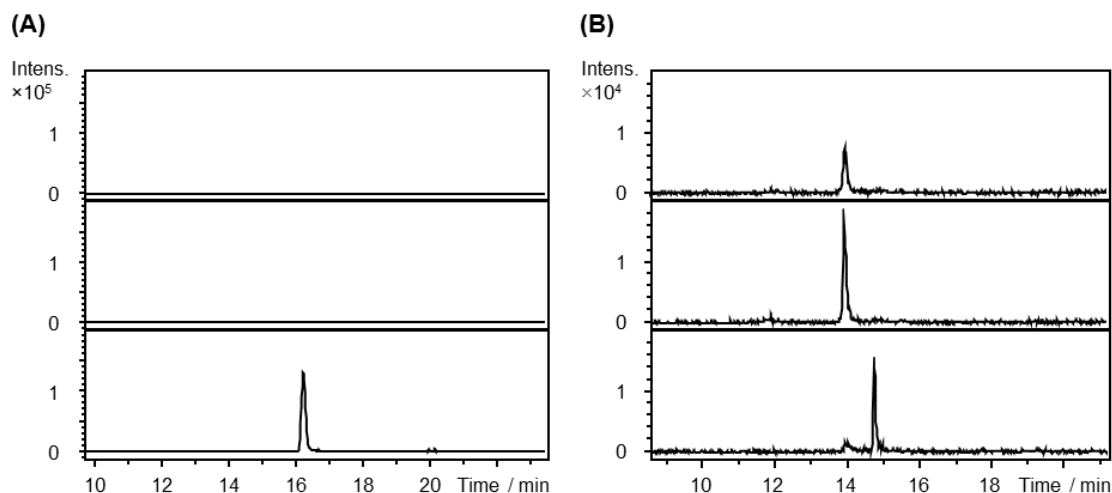
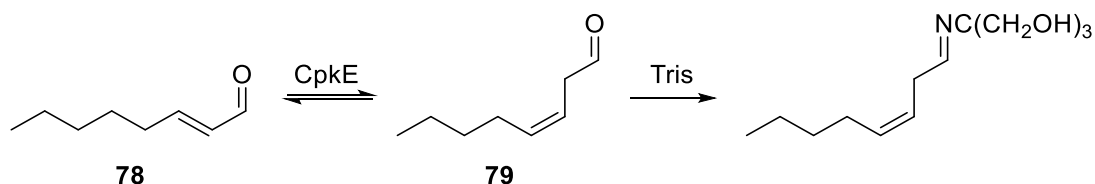


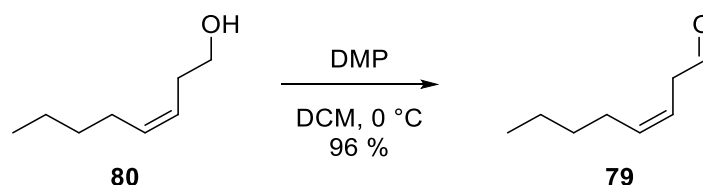
Figure 3.6: (A) Extracted ion chromatograms at $m/z = 304.1941 \pm 0.05$ (corresponding to the $[\text{M}+\text{H}]^+$ ion for (4*S*)-2-(4-hydroxyhexyl)thiazolidine-4-carboxylic acid). From top to bottom: reaction of His₆-CpkI with dodecane-1,5-diol and NAD⁺; reaction of His₆-CpkI with dodecane-1,5-diol and NADP⁺; and D-cysteine derivatised authentic standard. (B) Extracted ion chromatograms at $m/z = 230.1751 \pm 0.05$ (corresponding to the $[\text{M}+\text{H}]^+$ ion for tris-derivatised octenal). From top to bottom: reaction of *trans*-octen-2-al with His₆-CpkE and tris; control reaction from which the enzyme has been omitted; and authentic standard of tris-derivatised *cis*-octen-3-al. Experiments were performed in duplicate.

To investigate the hypothesised isomerase activity of CpkE, the recombinant protein was incubated with *trans*-octen-2-al **78**; a substrate mimic of the proposed true intermediate of the coelimycin biosynthetic pathway (**23**), containing the same α,β -unsaturated aldehyde functionality. Tris(hydroxymethyl)amino methane (tris) was added to act as a simultaneous buffer and derivatising agent (scheme 3.4).



Scheme 3.4: Assay employed to probe the possible isomerisation of *trans*-octen-2-al **78** (200 μM) by CpkE (100 μM), and derivatisation with tris (20 mM)

The mixture was then analysed by LC-MS alongside an authentic standard of derivatised *cis*-octen-3-al **79**, which was synthesised by oxidising *cis*-octen-3-ol **80** with DMP (scheme 3.5) and used immediately due to rapid *cis* to *trans* isomerisation.



Scheme 3.5: Synthesis of *cis*-octen-3-al **79** by oxidation of *cis*-octen-3-ol **80** (1.0 eq.) with DMP (1.3 eq.)

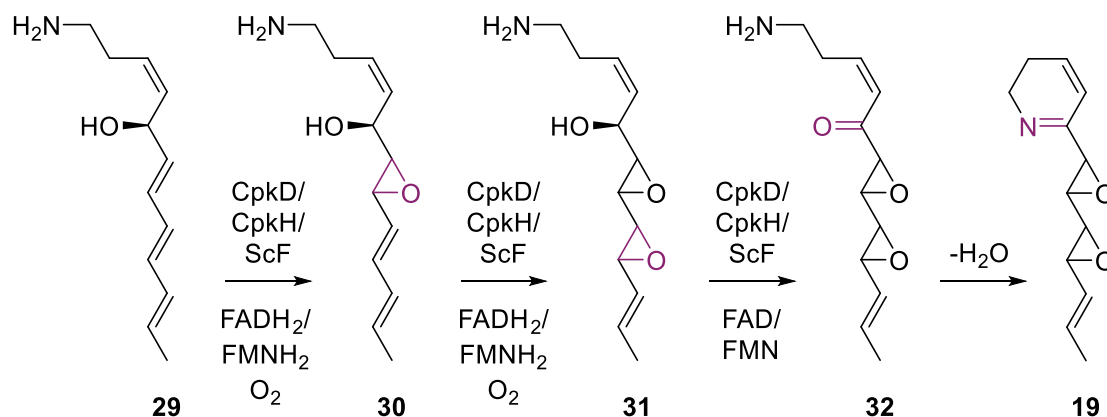
Examination of the retention times of the tris-derivatised isomers revealed that no *trans* to *cis* isomerisation had occurred during the reaction (figure 3.6b).

Therefore, it seems likely that CpkI and CpkE do not catalyse these reactions in coelimycin biosynthesis; instead they may catalyse other steps or have different enzymatic activities. For example, CpkE, as a member of the α,β -hydrolase fold superfamily, could function as an esterase, lipase, haloalkane dehalogenase, peroxidase or epoxide hydrolase.¹⁷⁴ Alternatively, it may be necessary to optimise reaction conditions or synthesise closer substrate mimics to effect catalysis (see section 3.2.4.4 for further investigations).

3.2 CpkD, ScF and CpkH

3.2.1 Introduction

CpkD, ScF and CpkH show sequence similarity to known flavin-dependent dehydrogenases and epoxidases, and have been proposed to catalyse the oxidative conversion of intermediate **29** to coelimycin A **19** (scheme 3.6).³⁰



Scheme 3.6: Proposed reactions catalysed by CpkD, CpkH and ScF in coelimycin biosynthesis³⁰

Previous attempts to decipher the roles of the oxidative enzymes and the order of pathway intermediates *in vivo* yielded inconclusive results (see section 1.4). Therefore, the genes were cloned and overexpressed, and the corresponding proteins purified, for *in vitro* investigations.

3.2.2 Cloning, overproduction and purification of proteins

cpkD, *scF* and *cpkH* were amplified by PCR and cloned into pET151 as described in section 3.1.2 (figure 3.7). However, because ScF and CpkH are known substrates of the TAT secretion system,⁹⁵ the corresponding genes were amplified as truncated sequences, without the *N*-terminal TAT signal peptide (predicted using PRED-TAT¹⁷⁵), to prevent secretion of the proteins during expression.

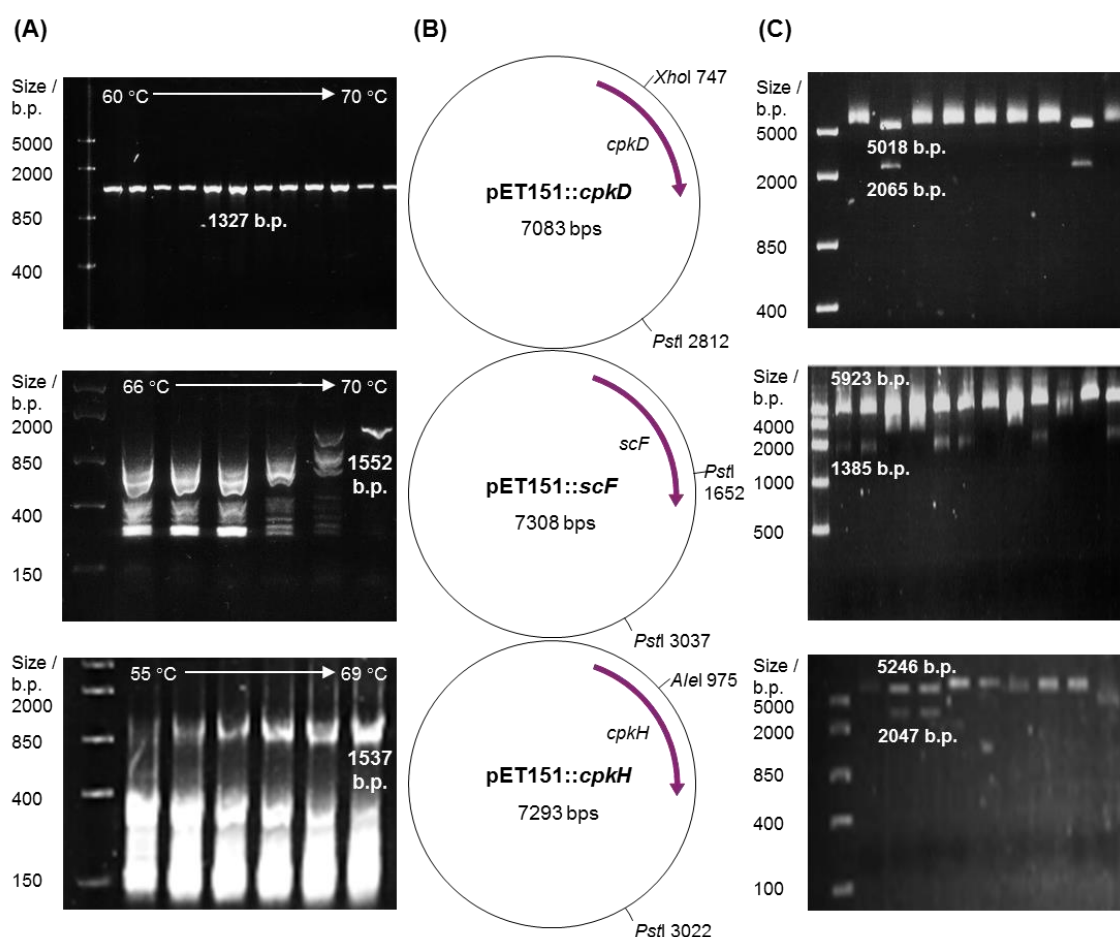


Figure 3.7: (A) Gels from agarose electrophoresis following PCR amplifications at selected annealing temperatures; (B) plasmid maps with restriction sites used for analysis; and (C) gels from agarose electrophoresis following restriction digests of pET151 constructs harbouring *cpkD* (top), *scF* (middle) and *cpkH* (bottom). The expected sizes of the DNA fragments are labelled.

Following verification of the inserts by sequencing, the proteins were overproduced in *E. coli* and purified by nickel affinity chromatography. All appeared to be pure and of the correct molecular weight by SDS-PAGE analysis (figure 3.8; calculated molecular weights of His₆-CpkD, His₆-ScF and His₆-CpkH are 51807, 60363 and 58827 Da, and respectively). Interestingly, the proteins were bright yellow in colour, indicative of flavin binding.

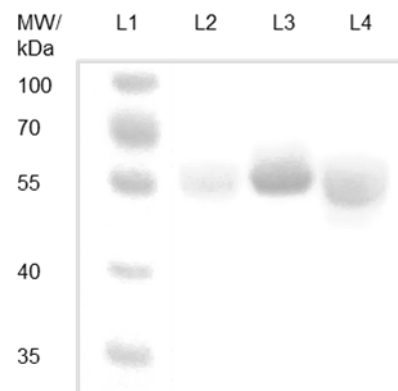


Figure 3.8: SDS-PAGE analysis of purified His₆-CpkD (L2), His₆-ScF (L3) and His₆-CpkH (L4) alongside a protein molecular weight marker (L1)

3.2.3 Examination of flavin binding and NAD(P)H utilisation

3.2.3.1 UV/Vis spectroscopic analysis to identify the flavin cofactor

To determine if the flavin cofactor was present, and if so, the form of flavin bound (flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD)), absorbance spectra for each protein were recorded from 300-600 nm (figure 3.9). The wavelength at maximum absorbance, λ_{max} , was then compared to literature values for FAD (450 nm) and FMN (446 nm).^{176,177} ScF and CpkH were both found to bind FAD, while for CpkD the result was ambiguous, suggesting that the enzyme could bind either or both cofactors.

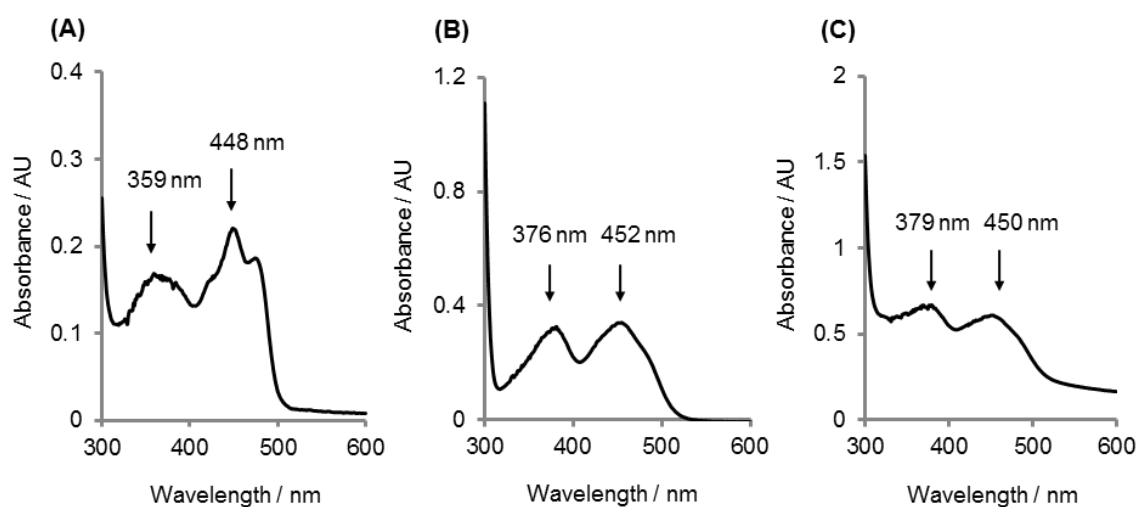


Figure 3.9: UV/Vis spectra of flavoproteins: (A) His₆-CpkD; (B) His₆-ScF; and (C) His₆-CpkH

3.2.3.2 Mass spectrometric analysis to determine the mode of binding

To determine whether the flavin cofactor was bound covalently or non-covalently, the intact proteins were analysed by HRMS (figure 3.10).

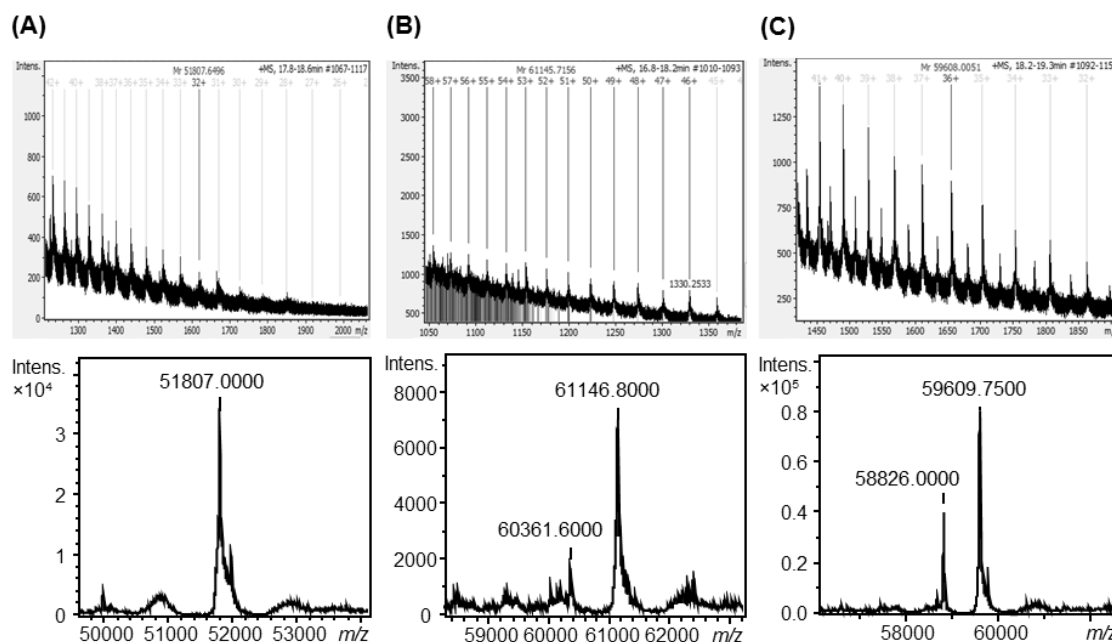


Figure 3.10: Measured (top) and deconvoluted (bottom) mass spectra of (A) His₆-CpkD; (B) His₆-ScF; and (C) His₆-CpkH. The calculated protein molecular weights are 51807 Da, 60363 Da (+ FAD = 61148 Da) and 58827 Da (+ FAD = 59611 Da), respectively.

For CpkD, there was a single peak corresponding to the molecular weight of the apo-protein, suggesting non-covalent binding of flavin. For ScF and CpkH, additional peaks were apparent at approximately +784 Da, indicating covalent binding of FAD. This observation is consistent with the measured UV/Vis absorption spectra (figure 3.9) and with ScF and CpkH being extracellular enzymes. Presumably, stronger binding of flavin would be favourable, as once excreted, the proteins would have limited access to free flavin.¹⁷⁸ Furthermore, covalent flavin attachment could aid with stabilising the protein tertiary structure during transport across the lipid bilayer.¹⁷⁹

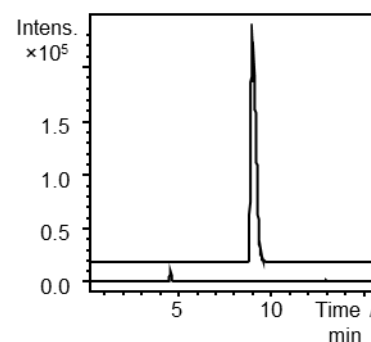


Figure 3.11: Extracted ion chromatograms at $m/z = 784.1499 \pm 0.05$ (top trace, corresponding to the $[M-H]^-$ ion of FAD) and $m/z = 455.0973 \pm 0.05$ (bottom trace, corresponding to the $[M-H]^-$ ion of FMN)

To deduce the nature of the cofactor bound non-covalently to CpkD, the enzyme was boiled and the supernatant analysed by LC-MS. Peaks corresponding to FMN (calculated $m/z = 455.0963$ Da, observed = 455.0970 Da) and FAD (calculated $m/z = 784.1499$ Da, observed = 784.1495 Da) were apparent (figure 3.11), suggesting that either both cofactors are present, or that solely FAD is bound, but fragments within the mass spectrometer.

3.2.3.3 Prediction of sites of covalent FAD attachment

In an attempt to predict the amino acid residue(s) covalently bound to FAD, the sequences of ScF and CpkH were compared to those of characterised flavoproteins using the NCBI BLAST database. ScF and CpkH were found to share a respective 56 % and 49 % sequence identity with TamL, an FAD-dependent dehydrogenase involved in tirandamycin biosynthesis.¹⁸⁰ The crystal structure of TamL revealed that two residues, Cys122 and His62, are covalently attached to FAD *via* the C-6 atom of the isoalloxazine ring and the 8 α -methyl group (figure 3.12).

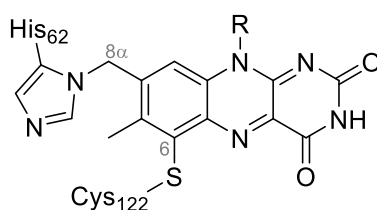


Figure 3.12: Amino acid residues involved in the covalent binding of FAD to TamL. Note: Only the isoalloxazine moiety of FAD is shown.

A multiple sequence alignment showed that these residues are conserved in ScF and CpkH (figure 3.13), suggesting a similar bicovalent mode of binding. Attempts to experimentally validate this by tryptic digestion and MS/MS analysis were unsuccessful, presumably due to photodegradation of FAD adducts.

TamL	7	VAPGDI RYEDLRRGENLRVGDPEETHLVGSAAEIEQVLSRAVRS GKRVAVRSGGHCYED
CpkH	52	VTPGDPRYATLVAGANGRWKGTDPDYLVASDADQVVQAVRET LAKGLRFVAVKSGGHCYED
ScF	60	VTRSDPRYQGLISGSNQRWVGNPDYIRVSSAGQVVRVAVQEA VDKGLKIAVRSGGHCDED
TamL	67	FVANS DVRVVMDSRLSAVGFDEERGAFAVEAGATLGAVYKTLFRVWGVTLPGGACPDVG
CpkH	112	FTTNSGVRVLIDLSAMAAIEFDASRRFAFAGPGAQLGSVYQKLYDGWGVTLPGGTCP SVA
ScF	120	FVANRDVRVVIDMAGMDSVTYDRERRAFVAVGPGARLGI VYRTLYKRWGVVLPGGTCPIVG

Figure 3.13: A partial view of a multiple sequence alignment of TamL, ScF and CpkH, produced using T-Coffee and BoxShade software. Darker regions indicate higher levels of homology. Residues implicated in covalent FAD binding are highlighted with a purple box.

3.2.3.4 UV/Vis spectroscopic analysis of flavin reduction by NAD(P)H

To examine whether CpkD, ScF and CpkH could bind and utilise NAD(P)H to reduce their flavin cofactors, the enzymes were incubated with either NADH or NADPH and the absorbance monitored at 340 nm. In each case, a decrease in absorbance could be observed upon NADH addition (figure 3.14), implying that NADH was being oxidised during the concomitant reduction of flavin. Little or no change in absorbance could be observed upon NADPH addition. This suggests that all three enzymes selectively bind NADH, and are capable of catalysing redox chemistry without a partner protein. This is consistent with the absence of a flavin reductase-encoding gene within the *cpk* biosynthetic gene cluster.

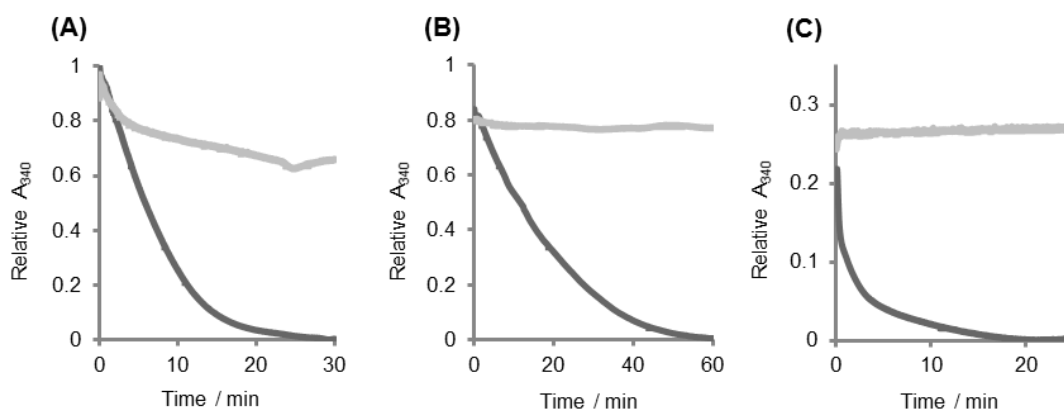


Figure 3.14: UV/Vis monitoring at A₃₄₀ to detect the oxidation of NADH (dark grey) or NADPH (light grey), by (A) His₆-CpkD; (B) His₆-ScF; and (C) His₆-CpkH

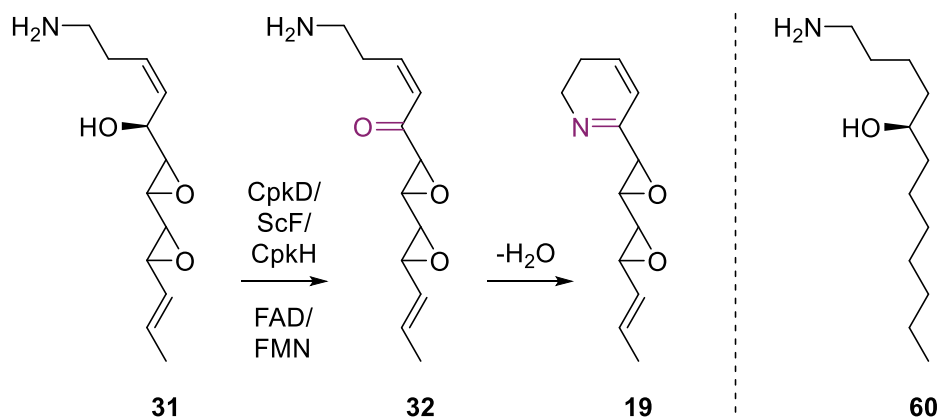
3.2.4 Probing flavin-dependent dehydrogenase activity

3.2.4.1 Incubation of 1-aminododecane-5-ol with CpkD, ScF and CpkH

One of the three flavoproteins, CpkD, ScF or CpkH, has been proposed to catalyse the oxidation of the C-5 hydroxyl group of **31** during coelimycin biosynthesis. This is then predicted to be followed by spontaneous intramolecular imine formation (scheme 3.7).

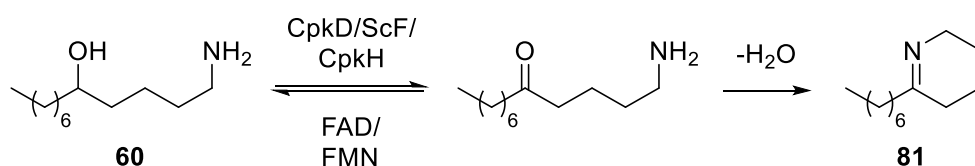
To investigate which of the enzymes exhibit dehydrogenase activity, 1-aminododecane-5-ol **60** (figure 3.15), previously synthesised for assays with CpkG (see section 2.1.3.3), was incubated separately with His₆-CpkD, His₆-ScF and His₆-CpkH (scheme 3.8). The reaction mixtures were then analysed by LC-MS. In the CpkH-catalysed reaction, a peak with the same molecular weight and predicted molecular formula as 6-heptyl-2,3,4,5-

tetrahydropyridine **81** was apparent (figure 3.16). The peak was not evident in the analogous CpkD or ScF reaction mixtures.



Scheme 3.7: Proposed oxidation reaction catalysed by CpkD, ScF or CpkH

Figure 3.15: Proposed substrate mimic of **31**



Scheme 3.8: Assay employed to probe flavin-dependent dehydrogenase activity, by incubation of 1-aminododecan-5-ol (200 μM) with CpkD, ScF or CpkH (20 μM)

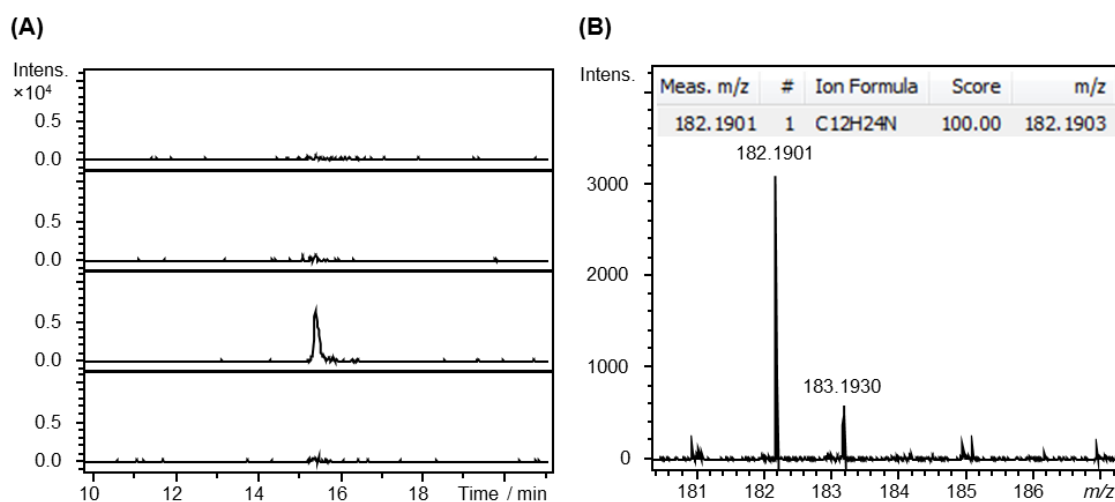
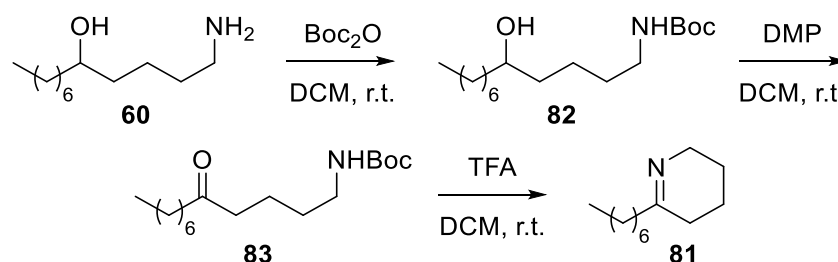


Figure 3.16: (A) Extracted ion chromatograms at $m/z = 182.1903 \pm 0.05$ (corresponding to the $[M+H]^+$ ion for 6-heptyl-2,3,4,5-tetrahydropyridine). From top to bottom: reaction of 1-aminododecan-5-ol with CpkD, ScF, CpkH and without enzyme. (B) Measured m/z and predicted molecular formula for the compound eluting at 15.4 min.

In order to confirm the product of the CpkH-catalysed reaction, an authentic standard of 6-heptyl-2,3,4,5-tetrahydropyridine **81** was synthesised from 1-aminododecan-5-ol **60** in three steps (scheme 3.9). This involved a Boc-protection of the amine, followed by a

DMP-mediated oxidation of the hydroxyl group. Deprotection of the amine with TFA and a subsequent base work-up afforded the desired product in moderate yield.



Scheme 3.9: Synthesis of 6-heptyl-2,3,4,5-tetrahydropyridine **81** from 1-aminododecan-5-ol **60**; 18 % yield over 5 steps from δ -dodecalactone **61**

The assay with CpkH was then repeated. LC-MS analysis indicated that the product of the enzymatic reaction and the authentic standard displayed identical retention times (figure 3.17). Thus, CpkH is a flavin-dependent dehydrogenase and likely catalyses the oxidation of **31** to **32** in coelimycin biosynthesis. This supports the data obtained from comparative metabolic profiling with a *cpkH* deletion mutant (see section 1.4).

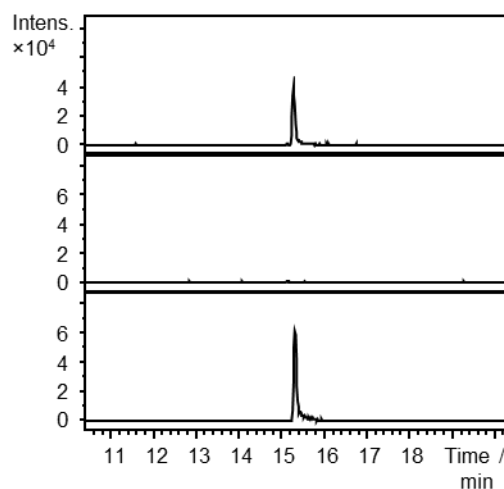


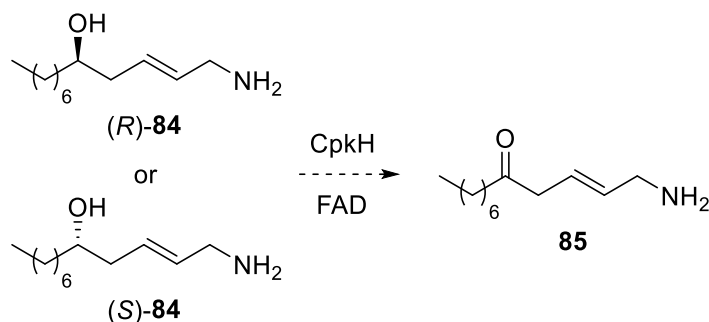
Figure 3.17: Extracted ion chromatograms at $m/z = 182.1903 \pm 0.05$ (corresponding to the $[M+H]^+$ ion for 6-heptyl-2,3,4,5-tetrahydropyridine). From top to bottom: reaction of 1-aminododecan-5-ol with CpkH; control reaction from which the enzyme was omitted; and authentic standard.

Notably, NAD^+ cannot be used for the regeneration of FAD, due to the unfavourable redox potential ($E_{\text{cell}}^{\circ} = -0.1 \text{ V}$)¹⁸¹; *in vivo*, the oxidation of $FADH_2$ is performed by electron carriers such as cytochromes.⁷⁷ Therefore, in these assays, molecular oxygen, phenazine methosulfate (PMS) and ferrocenium hexafluorophosphate were trialled as external oxidants, however low levels of conversion of **60** to **81** were consistently observed. Thus, reaction conditions may require optimisation, or alternatively, CpkH may only catalyse one turnover, possibly due to product inhibition.

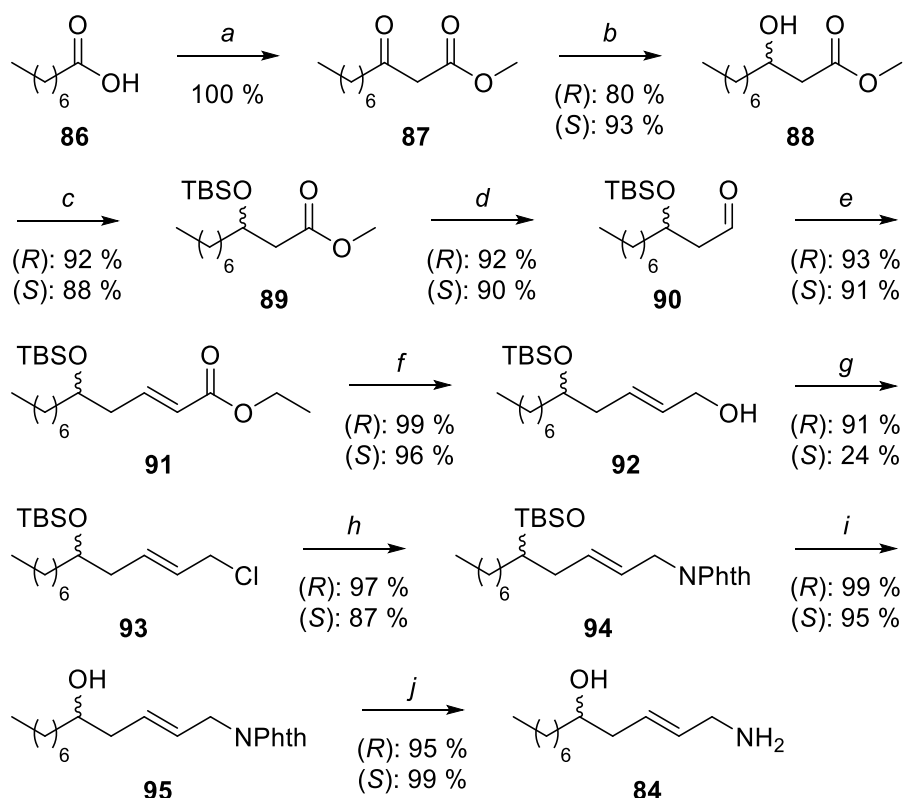
3.2.4.2 Synthesis of (*R,E*)- and (*S,E*)-1-aminododec-2-en-5-ol

In order to investigate the enantio- and configurational selectivity of CpkH, the synthesis of two enantiomers of **60** bearing *trans*-double bonds ((*R,E*)- and (*S,E*)-1-aminododec-2-

en-5-ol **84**) was devised. According to KR stereochemical predictions and the proposed coelimycin biosynthetic pathway, only the (*R*)-enantiomer is expected to be oxidised. In the absence of additional tailoring enzymes, intramolecular cyclisation should not occur, due to restrictions imposed by the *trans*-double bond, meaning that the presence of the ketone intermediate could be conclusively confirmed (scheme 3.10). Furthermore, the potential of CpkI and CpkE to reduce or isomerise the *trans*-double bond of **84** could be examined.

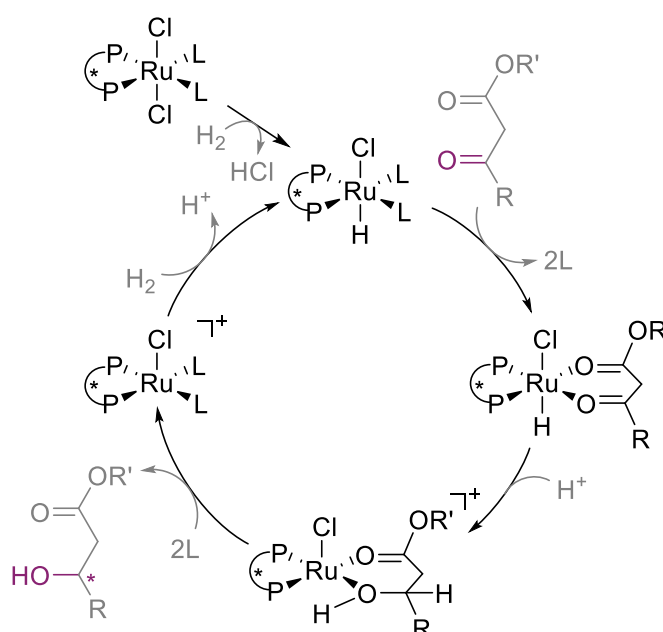


Scheme 3.10: Proposed CpkH-catalysed oxidation of (*R*)-**84** or (*S*)-**84**



Scheme 3.11: Synthetic routes to (*R*)-**84** and (*S*)-**84**. **a**) CDI, monomethyl malonate potassium salt, MgCl_2 , THF, r.t., o/n; **b**) $[\text{Ru}(\text{OAc})_2((R)\text{-BINAP})]$ or $[\text{Ru}(\text{OAc})_2((S)\text{-BINAP})]$, MeOH, HCl, H_2 (49 atm), 50°C , o/n; **c**) *tert*-butyl-dimethylsilyl chloride, imidazole, DMF, $0^\circ\text{C} \rightarrow \text{r.t.}$, o/n; **d**) DIBAL-H, DCM, -78°C , 1 h; **e**) ethyl(triphenylphosphoranylidene)acetate, THF, r.t., o/n; **f**) DIBAL-H, DCM, -78°C , 3 h; **g**) (*R*)-enantiomer: triphenylphosphine, *N*-chlorosuccinamide, THF, r.t., 3.5 h; (*S*)-enantiomer: triethylamine, tosyl chloride, DCM, $0^\circ\text{C} \rightarrow \text{r.t.}$, o/n; **h**) potassium phthalimide, DMF, r.t., 72 h; **i**) (1*S*)-(+)-10-camphorsulfonic acid, MeOH, DCM, r.t., 1.5 h; **j**) 40% aq. MeNH_2 , EtOH, r.t., 2 h.

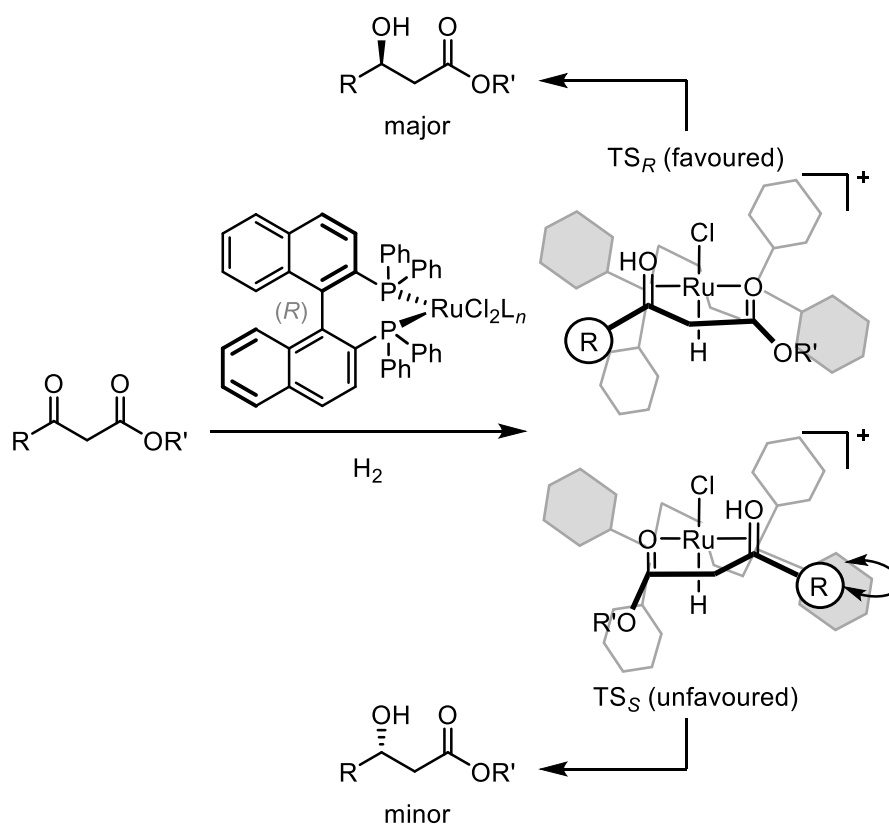
A synthetic route was developed, enabling the assembly of both enantiomers in ten steps (scheme 3.11). In the first reaction, octanoic acid **86** was derivatised with 1'-carbonyldiimidazole (CDI), and the resulting activated imidazolide mixed with monomethyl malonate potassium salt and magnesium chloride to form methyl-3-oxodecanoate **87** in quantitative yield. The β -keto ester was then asymmetrically reduced with Noyori's (*R*)- or (*S*)-BINAP-Ru dichloride catalyst¹⁸² under a high pressure of H₂ (scheme 3.12), to afford (*R*)- or (*S*)-methyl-3-hydroxydecanoate **88**, respectively. Due to the bias provided by the diastereomeric transition states (scheme 3.13), the hydride transfers proceeded in high enantiomeric purity (see section 3.2.4.3).¹⁸³



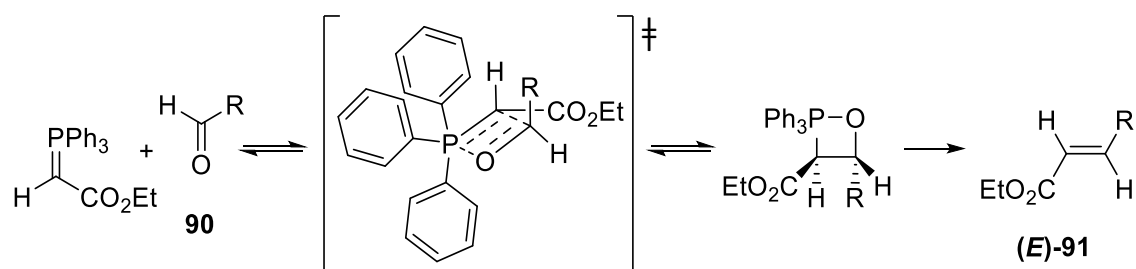
Scheme 3.12: Catalytic cycle for the asymmetric hydrogenation of β -keto esters by $[\text{Ru}(\text{Cl})_2\text{BINAP}]$, generated *in situ* from $[\text{Ru}(\text{OAc})_2\text{BINAP}]$ and methanolic HCl. In the first step, the active Ru^{II} monohydride complex is formed from reaction of the dichloride precatalyst with H₂. The β -keto ester substrate then chelates to the complex, displacing two solvent ligands. The keto oxygen is then protonated, increasing the electrophilicity at the carbonyl carbon and thereby facilitating intramolecular hydride transfer. The chiral β -hydroxy ester product is then displaced by incoming solvent ligands. The ruthenium complex then reacts with H₂ to regenerate the monohydride, thus completing the catalytic cycle. L = solvent; P*P = BINAP.¹⁸³

The two enantiomers were then protected with trimethyl silyl (TMS) chloride, and the resulting β -silyl ether esters **89** reduced with one equivalent of diisobutylaluminium hydride (DIBAL-H) to afford the corresponding aldehydes **90**. Ethyl(triphenylphosphoranylidene)acetate was subsequently added to install the required double bond *via* a Wittig reaction.^{184,185} Because a stabilised ylide was used, thermodynamic conditions prevailed, resulting in the predominant formation of the more

stable *trans*-oxaphosphetane (scheme 3.14).¹⁸⁶ This led to the isolation of *E* : *Z* isomers in >19 : 1 ratio.



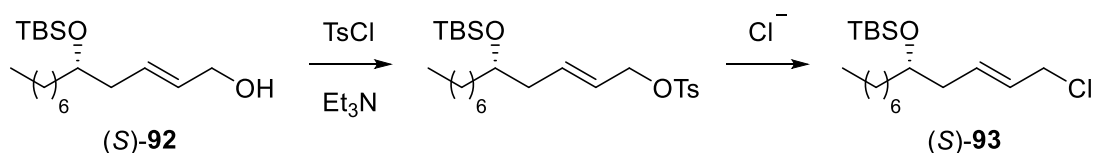
Scheme 3.13: Enantio-determining transition states during reaction of a β -keto ester with $[\text{RuCl}_2(R)\text{-BINAP}]$ and H_2 , adapted from Noyori et al.¹⁸³ The phenyl rings projecting forward and backward are depicted as shaded and colourless hexagons, respectively. BINAP rings have been omitted for clarity. In TS_R , the *R* group occupies an open quadrant within the chiral template. In TS_S , there is a significant nonbonding repulsion between the *R* group and the equatorial phenyl ring, indicated by the double-headed arrow. Thus, TS_R , and therefore the (*R*)- β -hydroxy-ester product, are favoured.



Scheme 3.14: Reaction of aldehyde **90** with ethyl(triphenylphosphoranylidene)acetate to form (*E*)-**91** as the major product. Unfavourable 1,2-steric interactions between the *R* group and the ester moiety are minimised in the *trans*-oxaphosphetane transition state.¹⁸⁶

Esters **91** were then reduced with excess DIBAL-H to afford the corresponding alcohols **92**. Subsequently, (*S*)-**92** was reacted with *p*-toluenesulfonyl chloride and triethylamine

in an attempt to form the tosylate, however a chlorinated side product, (*S*)-**93** (24 %), was isolated instead. This likely resulted from further reaction of the tosylate intermediate with a chloride ion, due to the enhanced electrophilicity provided by the vinyl group (scheme 3.15). Indeed, similar observations have been reported for benzyl alcohols.¹⁸⁷



Scheme 3.15: Reaction of vinyl alcohol (*S*)-**92** with TsCl and Et₃N to form a tosylate intermediate. The leaving group is subsequently displaced by a chloride ion to afford (*S*)-**93**.

Since chlorides are also effective leaving groups, the side product was reacted on with potassium phthalimide to install a primary amino moiety *via* the Gabriel synthesis,¹⁸⁸ affording (*S*)-**94** in 87 % yield. As this reaction worked well, and phthalimide displacements of alkyl tosylates are often reported in poor to moderate yields,^{189,190,191,192,193} rather than attempting to optimise sulfonylation conditions for the other enantiomer, the direct synthesis of chloride (*R*)-**93** was attempted.

Initially, (*R*)-**92** was reacted with thionyl chloride. However, due to the highly acidic conditions, a mixture of compounds formed, including the *cis*-isomer and a possible Michael-addition product. Therefore, in a second attempt, (*R*)-**92** was reacted with triphenylphosphine and *N*-chlorosuccinamide. Complete, clean conversion of alcohol **92** to chloride **93** was observed in 3 hours.

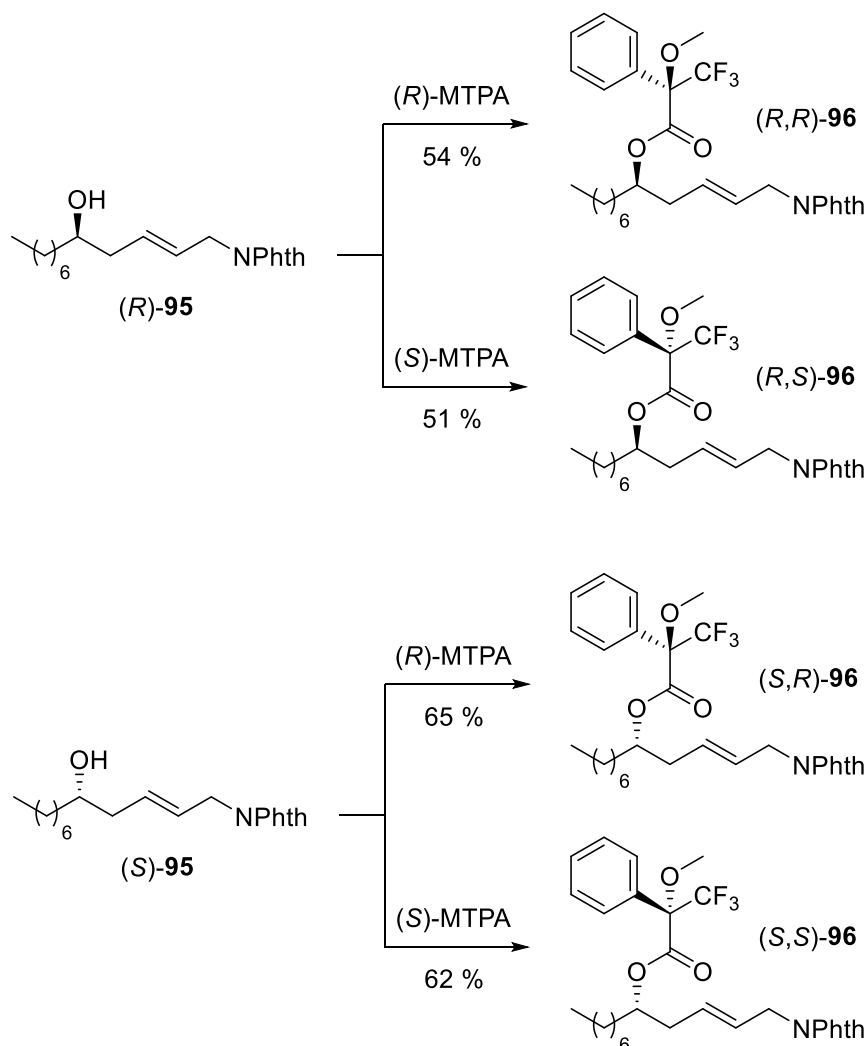
The chloride moiety of (*R*)-**93** was then displaced by potassium phthalimide as described for the (*S*)-enantiomer, to afford (*R*)-**94** in 97 % yield. Tetrabutylammonium fluoride (TBAF) was subsequently added to remove the silyl protecting group, however the reaction proceeded extremely slowly (> 4 days). Therefore, the TBS-intermediates were instead reacted with (1*S*)-(+)-10-camphorsulfonic acid (CSA), resulting in complete deprotection to afford (*R*)- and (*S*)-**95** in 1.5 hours.

In the final step, the phthalimido protecting groups were efficiently removed using aqueous methylamine. This reagent was chosen over the more commonly-used hydrazine, which has been shown to reduce unactivated carbon-carbon double bonds in the presence of oxygen.¹⁹⁴

Overall, the hypothetical CpkH substrates, (*R,E*)- and (*S,E*)-1-aminododec-2-en-5-ol **84**, were synthesised in 52 % and 13 % yield, respectively, over ten steps.

3.2.4.3 Determination of absolute stereochemistry and enantiomeric purity of (*R,E*)- and (*S,E*)-1-aminododec-2-en-5-ol

In order to deduce the enantiomeric purity and confirm the absolute stereochemistry of the two putative substrates, the phthalimido intermediates **95** were derivatised with (*R*)- and (*S*)-Mosher's acid¹⁹⁵ to yield four diastereomers: (*R,R*)-, (*R,S*)-, (*S,R*)- and (*S,S*)-**96** (scheme 3.16). The ¹H NMR shifts of the diastereomers were then compared.



Scheme 3.16: Synthesis of Mosher's esters of (*R*)- and (*S*)-**95**. Reaction conditions: (*R*)- or (*S*)-**95** (1.0 eq.), (*R*)- or (*S*)-Mosher's acid (MTPA: α -methoxy- α -trifluoromethylphenylacetic acid, 1.1 eq.), EDC (1.1 eq.), DMAP (0.1 eq.), DCM, r.t.

As expected, the pairs of enantiomers ((*R,R*) and (*S,S*); (*R,S*) and (*S,R*)) displayed identical $^1\text{H-NMR}$ spectra, while the pairs of diastereomers ((*R,R*) and (*R,S*); (*S,R*) and (*S,S*)) exhibited slightly different chemical shifts. These differences are mainly attributable to the magnetic shielding effect exerted by the phenyl ring; proximal protons (residing above or below the plane of the phenyl ring) experience a higher degree of shielding and thus have a more upfield chemical shift, while distal protons are less affected (figure 3.18).^{196,197}

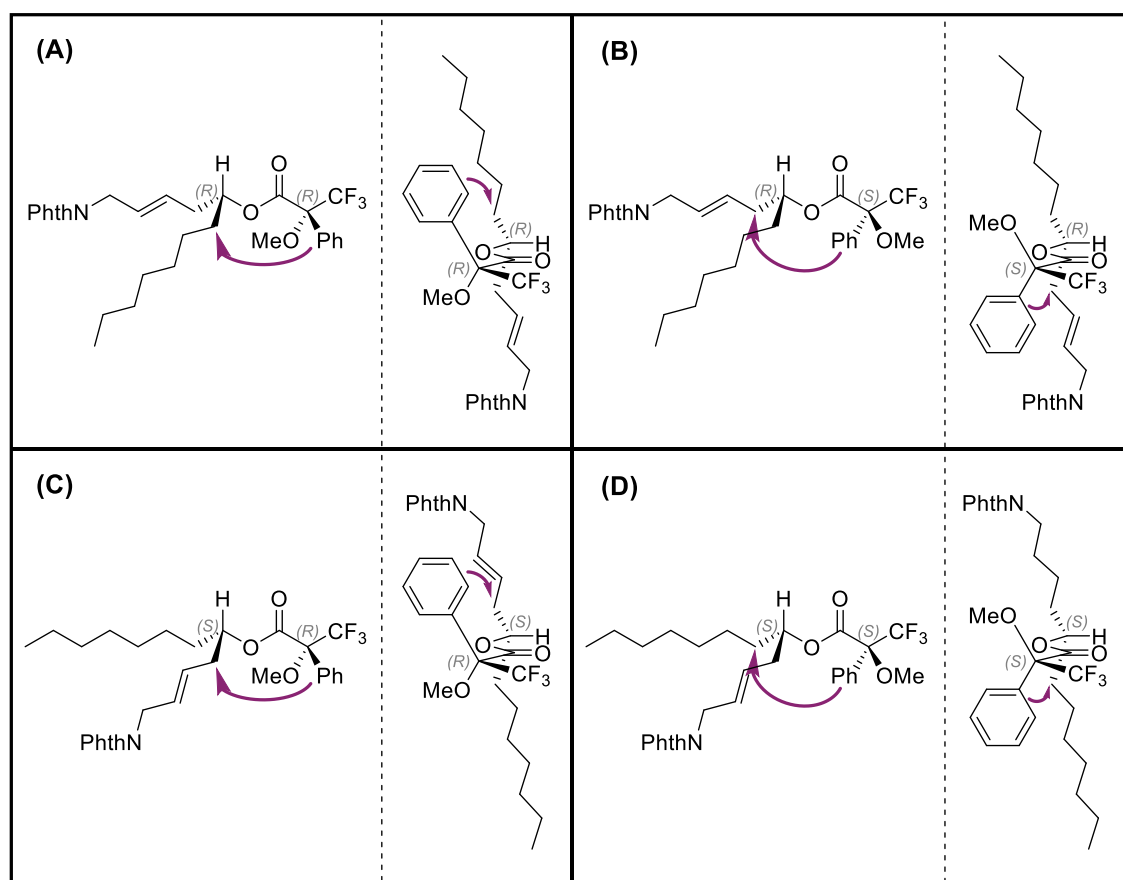


Figure 3.18: Two different representations of (A) the (*R,R*)-diastereomer; (B) the (*R,S*)-diastereomer; (C) the (*S,R*)-diastereomer; and (D) the (*S,S*)-diastereomer of Mosher's-derivatised **95**, adapted from Hoyer and coworkers.¹⁹⁷ The shielding effect exerted by the phenyl group is indicated with a purple arrow.

Therefore, protons 1-7 in the (*R,S*)- and (*S,R*)-diastereomers would be expected to be more shielded relative to those of the (*R,R*)- and (*S,S*)-diastereomers (figure 3.19). This effect can be readily observed and is particularly pronounced for protons 3 ($\Delta\delta^{SR} = -0.07$ ppm), 4 ($\Delta\delta^{SR} = -0.11$ ppm) and 5 ($\Delta\delta^{SR} = -0.10$ ppm) (figure 3.20), thus confirming the predicted absolute stereochemistry of both compounds.

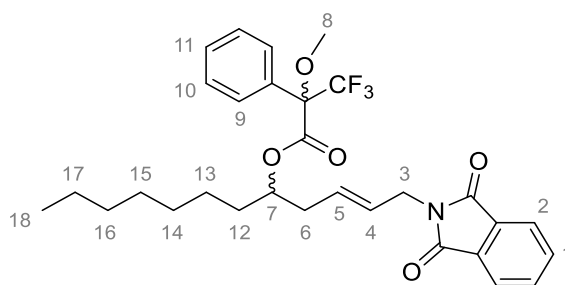


Figure 3.19: Structure of **96** with protons labelled

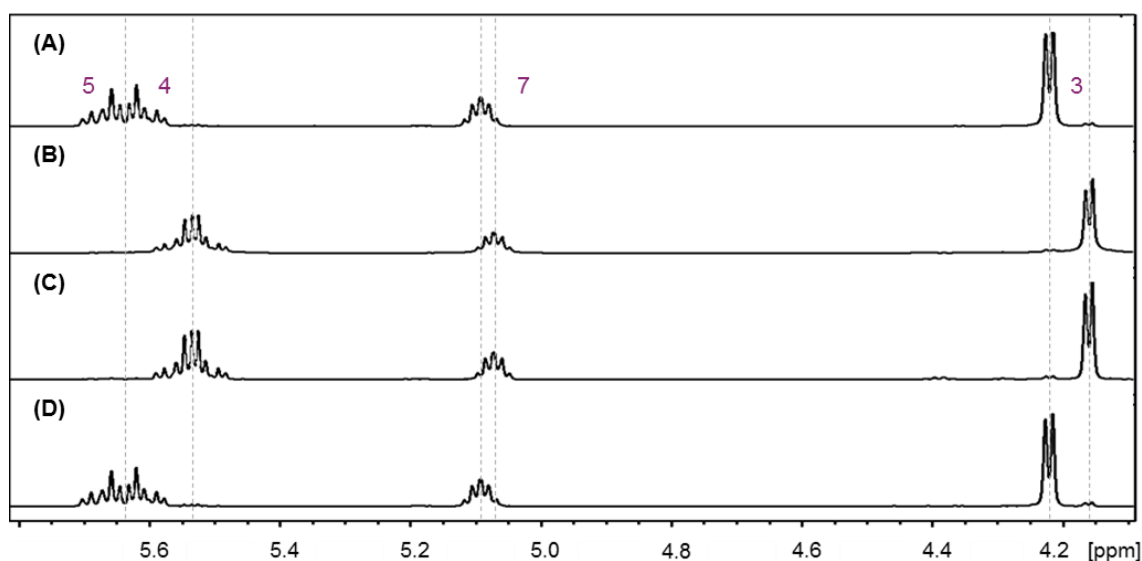


Figure 3.20: A comparison of ^1H NMR spectra from 4.2–5.8 ppm of (A) (*R,R*)-**96**; (B) (*R,S*)-**96**; (C) (*S,R*)-**96**; and (D) (*S,S*)-**96**. Protons 3, 4, 5 and 7 in spectra (B) and (C) are shifted upfield compared to those of (A) and (D).

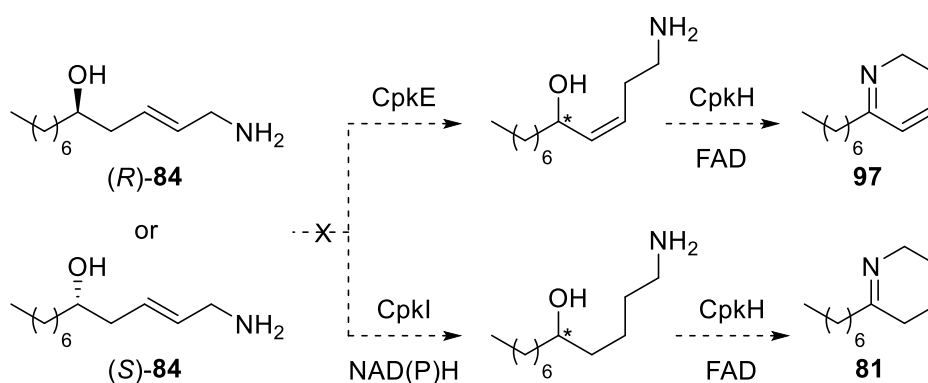
By measuring the intensities of analogous resonances, the enantiomeric ratio (er) and enantiomeric excess (ee) of both (*R*)- and (*S*)-**95** were determined to be 96 : 4 and 92 %, respectively.

3.2.4.4 Incubation of (*R,E*)- and (*S,E*)-1-aminododec-2-en-5-ol with CpkH, CpkE and CpkI

Following the synthesis of (*R,E*)- and (*S,E*)-1-aminododec-2-en-5-ol **84** and the confirmation of their absolute stereochemistry, the putative substrates were incubated separately with His₆-CpkH as described in section 3.2.4.1 (scheme 3.10). The supernatants were then analysed by LC-MS.

Peaks corresponding to the molecular weight of the oxidised compound, (*E*)-1-aminododec-2-en-5-one **85**, could not be detected in the extracted ion chromatograms from either reaction mixture. Peaks corresponding to (*R,E*)- and (*S,E*)-**84** had not decreased in intensity relative to control experiments, suggesting that the starting materials had not been consumed during the assay.

It therefore seems likely that these compounds need to be processed by other PKS-tailoring enzymes prior to oxidation by CpkH. For example, a CpkE-catalysed isomerisation, as tentatively suggested in the originally-proposed biosynthetic pathway, or a CpkI-catalysed reduction of the double bond, could yield suitable *cis*- or saturated substrate analogues, respectively (scheme 3.17).



Scheme 3.17: Assays employed to investigate the hypothetical conversion of (*R,E*)- or (*S,E*)-1-amino-dodec-2-en-5-ol **84** (200 μ M) to 6-heptyl-2,3-dihydropyridine **97** or 6-heptyl-2,3,4,5-tetrahydropyridine **81**, catalysed by CpkH (20 μ M) and either CpkE (20 μ M) or CpkI (20 μ M), respectively.

To investigate these hypotheses, (*R,E*)- and (*S,E*)-**84** were incubated with His₆-CpkH, in conjunction with either His₆-CpkE or His₆-CpkI. For His₆-CpkI, both NADH and NADPH were trialled as cofactors. Unfortunately, peaks corresponding to 6-heptyl-2,3-dihydropyridine **97** or 6-heptyl-2,3,4,5-tetrahydropyridine **81** could not be detected in LC-MS chromatograms.

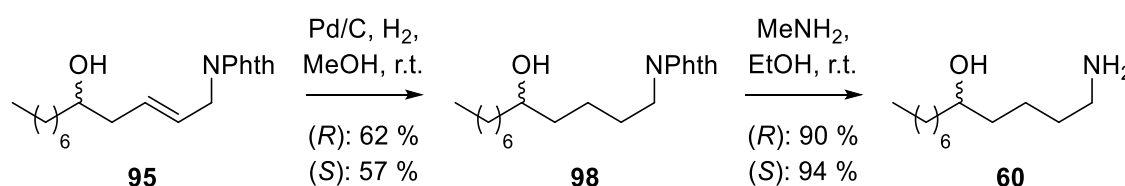
It is possible that these compounds are not ideal substrate mimics for CpkE or CpkI, or that the enzymes catalyse alternative reactions or exhibit different activities. For example, given that the coelimycin biosynthetic pathway is proposed to proceed through several epoxide intermediates, it seems plausible that CpkE could function as an epoxide hydrolase. This could be a detoxification mechanism, to prevent reaction of the epoxides with DNA or other cellular components, or even part of the intended biosynthetic

pathway, which is not observed due to the apparent trapping of coelimycin A **19** by amino acid components within the media. Indeed, a closer inspection of the sequence of CpkE reveals the presence of an aspartic acid residue in place of the usually-conserved catalytic nucleophilic serine residue within the GxSxG motif (GHDWG). This is a characteristic feature among dehalogenases and epoxide hydrolases.¹⁹⁸

Moreover, it is conceivable that CpkI does catalyse the reduction of the *trans*-double bond, but during polyketide chain assembly. This could be investigated by loading suitable substrates onto the ACP domain within CpkC, prior to incubation with CpkI and NAD(P)H. Alternatively, the final DH domain within the PKS could be inactive, despite sequence predictions. This would result in the presence of a 3'-hydroxyl residue rather than the proposed *trans*-double bond, meaning that ring formation could still occur. Future studies could focus on the synthesis of *cis*-, 3'-hydroxyl- and epoxide-containing substrate analogues for *in vitro* assays with CpkH, CpkC-DH and CpkE. Furthermore, $\Delta cpkE$ and $\Delta cpkI$ deletion mutants could be generated, to aid in determining the function of the corresponding enzymes *in vivo*.

3.2.4.5 Synthesis of (*R*)- and (*S*)-1-aminododecane-5-ol

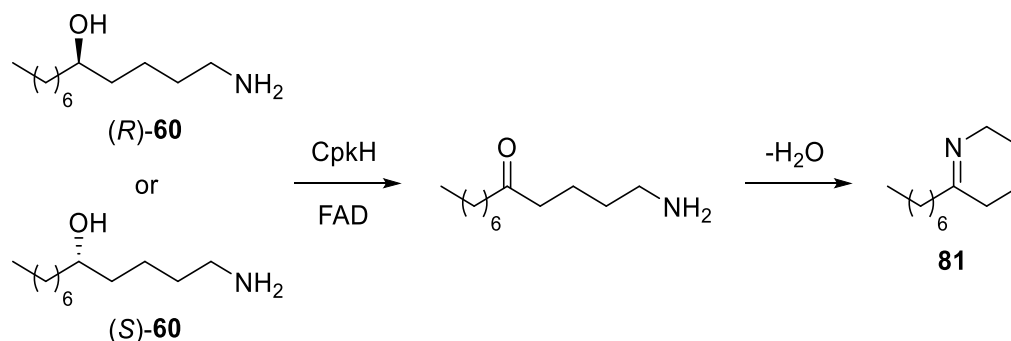
In order to gain insight into the stereospecificity of CpkH, enantiomers of the known saturated substrate, 1-aminododecan-5-ol **60**, were synthesised. This involved a Pd/C-catalysed reduction of the carbon-carbon double bond of intermediates (*R,E*)- and (*S,E*)-2-(5-hydroxydodec-2-en-1-yl)isoindoline-1,3-dione **95** with H₂, followed by a methylamine-mediated deprotection of the primary amine group (scheme 3.18). The first step proceeded in moderate yield, due to the unforeseen instability of the phthalimido group under hydrogenation conditions, while the final step afforded (*R*)- and (*S*)-**60** in >90 % yield.



Scheme 3.18: Left: Reduction of (*R,E*)- and (*S,E*)-**95** (1.0 eq.) with Pd/C (catalytic) and H₂ (1 atm), to form (*R*)- and (*S*)-**98**. Right: Deprotection of (*R*)- and (*S*)-**98** (1.0 eq.) with MeNH₂ (excess), to form (*R*)- and (*S*)-**60**.

3.2.4.6 Incubation of (*R*)- and (*S*)-1-aminododecane-5-ol with CpkH

Following the synthesis of (*R*)- and (*S*)-**60**, the enantiomeric substrates were incubated with His₆-CpkH in independent assays (scheme 3.19). The supernatants were then analysed by LC-MS alongside a negative control and an authentic standard of 6-heptyl-2,3,4,5-tetrahydropyridine **81**.



Scheme 3.19: Assay employed to investigate the oxidation of (*R*)- or (*S*)-**60** (200 μM), catalysed by CpkH (20 μM)

From the extracted ion chromatograms, it is clear that CpkH has a strong preference for the (*R*)-enantiomer of **60** (figure 3.21). This is consistent with KR sequence predictions, and thus supports the proposed stereochemistry of the 5'-hydroxyl-containing intermediates within the coelimycin biosynthetic pathway.

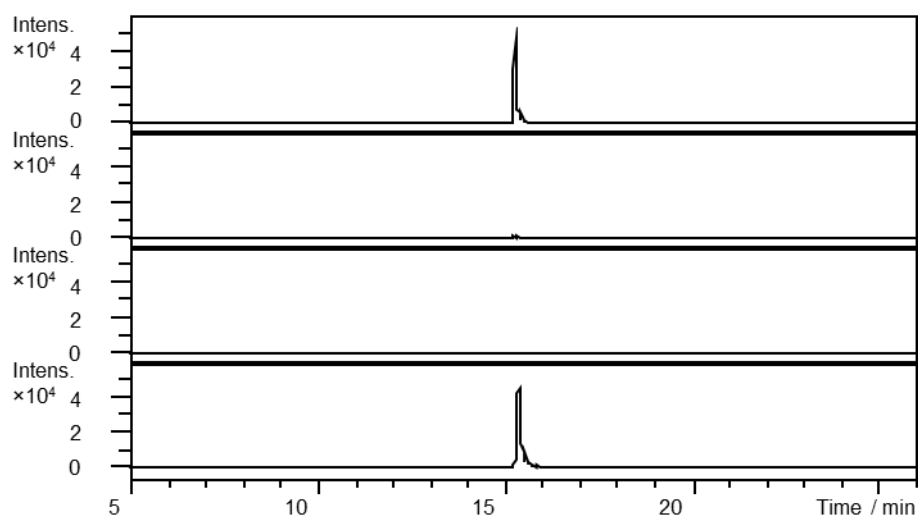


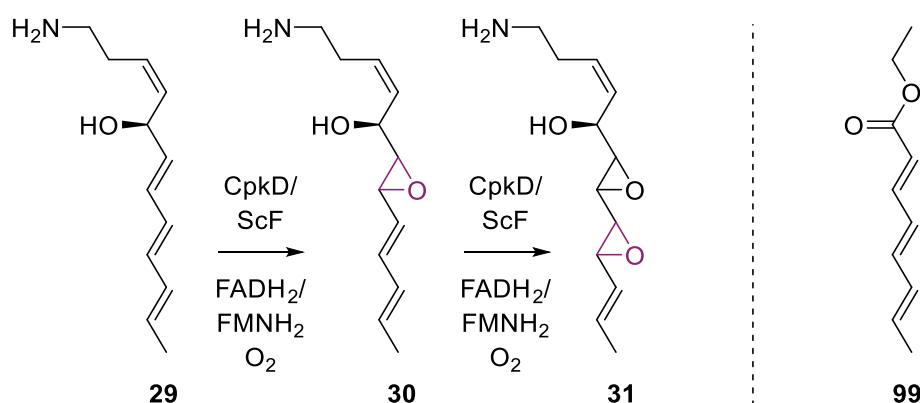
Figure 3.21: Extracted ion chromatograms at $m/z = 182.1903 \pm 0.05$ (corresponding to the $[\text{M}+\text{H}]^+$ ion for 6-heptyl-2,3,4,5-tetrahydropyridine). From top to bottom: reaction of (*R*)-1-aminododecan-5-ol with His₆-CpkH (ion count = 45640 a.u.); reaction of (*S*)-1-aminododecan-5-ol with His₆-CpkH (ion count = 1864 a.u.); control reaction from which the enzyme was omitted; and authentic standard. Independent experiments were performed in triplicate.

In order to quantify the preference of CpkH for (*R*)-**60**, the intensities of the product peaks were recorded. The ratio of the ion counts (45640 : 1864 = 96 : 4) corresponded to the enantiomeric purity of the synthetic substrates, suggesting that the product peak in the assay with (*S*)-**60** is attributable to the presence of (*R*)-enantiomeric impurities. Therefore, this implicates CpkH as being a stereospecific, rather than stereoselective, FAD-dependent dehydrogenase.

3.2.5 Probing flavin-dependent monooxygenase activity

3.2.5.1 Synthesis of ethyl (*2E,4E,6E*)-octa-2,4,6-trienoate

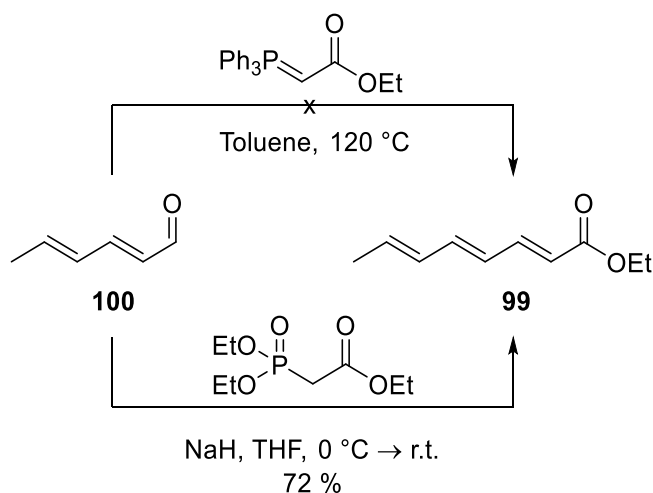
Following the characterisation of CpkH as a flavin-dependent dehydrogenase, it was hypothesised that CpkD and/or ScF are likely to act as flavin-dependent monooxygenases in coelimycin biosynthesis (scheme 3.20).



Scheme 3.20: Proposed epoxidation reactions catalysed by CpkD and/or ScF

Figure 3.22: Proposed substrate mimic of **29**

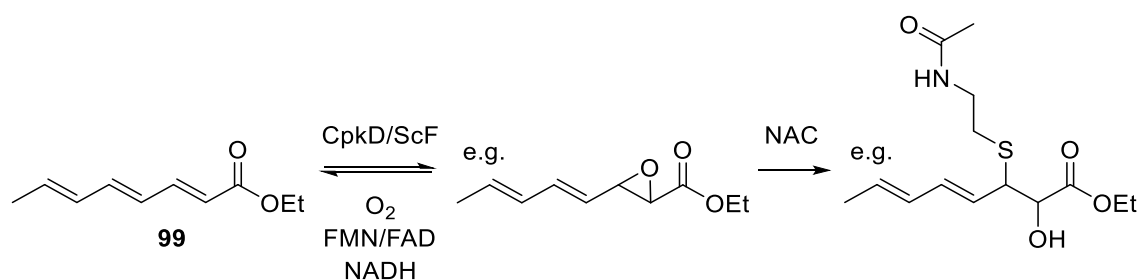
To probe their prospective monooxygenase activities, the synthesis of a simple substrate mimic of **29** containing a triene functionality (**99**, figure 3.22) was planned (scheme 3.21). In a first attempt, *trans,trans*-2,4-hexadienal **100** was reacted with (carbethoxymethylene)triphenylphosphorane under Wittig^{184,185} conditions. However, the high temperature employed (120 °C) caused polymerisation to occur. Therefore, instead, **100** was mixed with triethyl-4-phosphonoacetate and sodium hydride at low temperature in a Horner-Wadsworth-Emmons¹⁹⁹ reaction. Ethyl (*2E,4E,6E*)-octa-2,4,6-trienoate **99** was afforded in good yield, in a 9 : 1 ratio of *2E* : *2Z* isomers.



Scheme 3.21: Attempts to synthesise ethyl (2E,4E,6E)-octa-2,4,6-trienoate **99** from *trans,trans*-2,4-hexadienal **100**

3.2.5.2 Incubation of ethyl (2E,4E,6E)-octa-2,4,6-trienoate with CpkD and ScF

Following the synthesis of triene ester **99**, the putative substrate was incubated with His₆-CpkD and His₆-ScF independently. His₆-CpkD, which binds flavin non-covalently, was supplemented with additional FMN or FAD. NADH and *N*-acetylcysteamine (NAC) were also added to both reaction mixtures, to reduce the flavin cofactors to their required forms (FMNH₂ or FADH₂, see section 3.2.3.4) and to trap any formed epoxides, respectively (scheme 3.22).



Scheme 3.22: Putative epoxidation of **99** (200 μM) by CpkD/ScF (20 μM), FMN/FAD (40 μM) and NADH (40 μM), followed by trapping with NAC (400 μM)

The supernatants were then analysed by LC-MS. Clusters of peaks were apparent in the CpkD-catalysed reaction, with molecular ions corresponding to a NAC ring-opened mono-epoxide ($[\text{M}+\text{Na}]^+ = 324.1240$) and bis-epoxide ($[\text{M}+\text{Na}]^+ = 340.1189$) (figure 3.23). The same peaks were not evident in the analogous ScF assay. Furthermore, in reactions where CpkD had been supplemented with FAD or FMN, the peaks were approximately 6-fold and 40-fold more intense, respectively. This suggests that CpkD has

a preference for FMN, but can utilise both cofactors, and is thus consistent with the UV/Vis and HRMS data previously obtained (see sections 3.2.3.1 and 3.2.3.2). Peaks corresponding to tris-epoxide derivatives could not be detected, in agreement with the structures of the proposed intermediates in coelimycin biosynthesis.

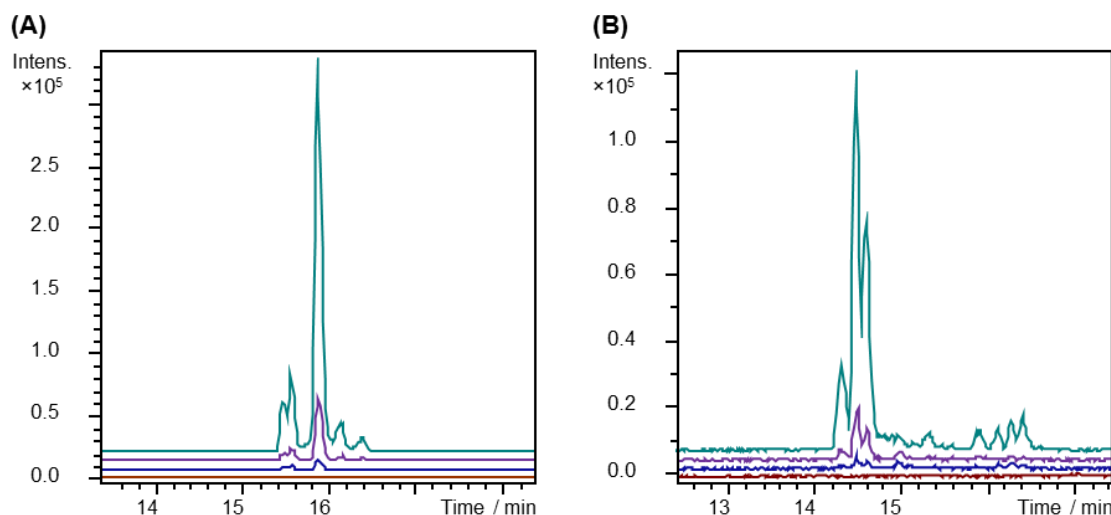


Figure 3.23: Extracted ion chromatograms at (A) $m/z = 324.1240 \pm 0.05$ (corresponding to the $[M+Na]^+$ ion for the NAC-derivatised mono-epoxide) and (B) 340.1189 ± 0.05 (corresponding to the $[M+Na]^+$ ion for the NAC-derivatised bis-epoxide), Brown trace: reaction with His₆-ScF; blue trace: reaction with His₆-CpkD; purple trace: reaction with His₆-CpkD, supplemented with additional FAD; green trace: reaction with His₆-CpkD, supplemented with additional FMN. Independent experiments were performed in triplicate.

It was hypothesised that the clusters of peaks may correspond to a range of regioisomers and diastereomers of NAC-derivatised mono- and bis-epoxides (figure 3.24). However, to first prove that the compounds were products of CpkD catalysis, several control reactions were performed: without substrate; without NAC; without CpkD; with boiled CpkD; and without NADH (figure 3.25).

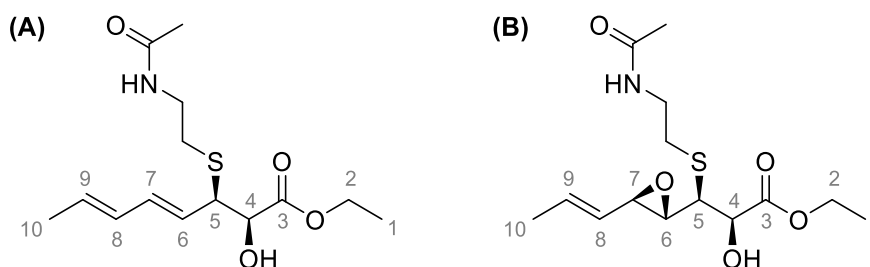


Figure 3.24: Possible structures of a NAC ring-opened mono-epoxide (A) and bis-epoxide (B). Other structures may include regioisomers (e.g. epoxidation at carbons 6-7 and NAC addition at carbon 7), Michael-addition products (e.g. epoxidation at carbons 4-5 and NAC addition at carbon 9) and related diastereomers.

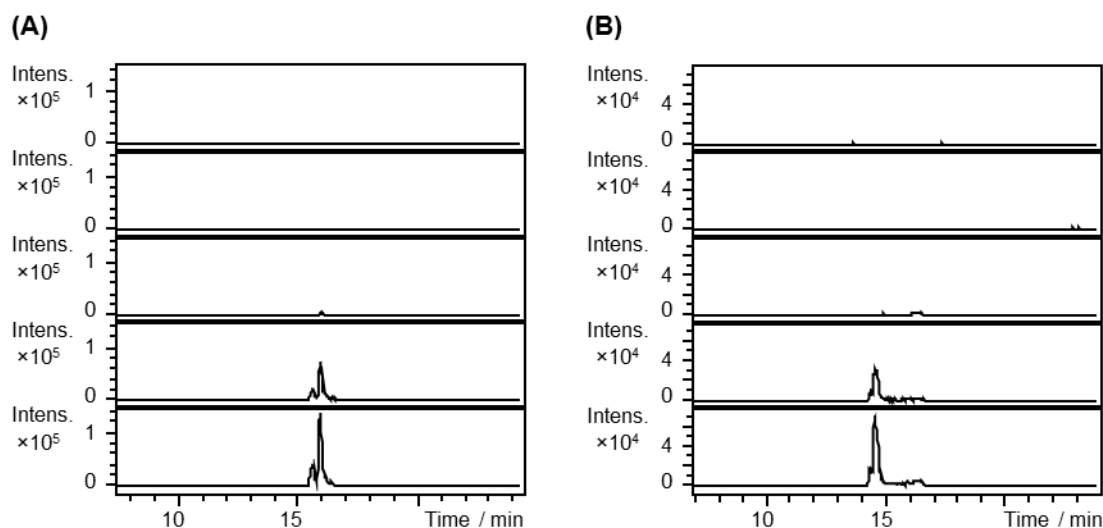


Figure 3.25: Extracted ion chromatograms at (A) $m/z = 324.1240 \pm 0.05$ (corresponding to the $[M+Na]^+$ ion for the NAC-derivatised mono-epoxide) and (B) 340.1189 ± 0.05 (corresponding to the $[M+Na]^+$ ion for the NAC-derivatised bis-epoxide) arising from the incubation of His₆-CpkD with triene ester **99**, FMN, NADH and NAC. From top to bottom: reaction without substrate; reaction without NAC; reaction without His₆-CpkD; reaction with boiled His₆-CpkD; and reaction without NADH. Independent experiments were performed in triplicate.

As expected, product peaks could not be detected in the assays without substrate or NAC. However, a small peak ($\sim 1/1000$ intensity) was apparent in the reaction without CpkD, but to which FMN had been added. It is possible that FMN can perform the epoxidation chemistry alone, but less efficiently. For instance, CpkD may serve to bind FMN and the substrate within close proximity, while favourably altering their pKa. Consistent with this, Schneider and coworkers recently reported that flavin derivatives can act as stand-alone catalysts, which are less effective than their flavoprotein counterparts.²⁰⁰

Surprisingly, a relatively intense cluster of peaks could be detected in the reaction with boiled His₆-CpkD. To investigate this phenomenon, circular dichromism (CD) spectra of the enzyme were obtained, following incubation at 4 °C or 100 °C for 1 hour. Comparison of the spectra revealed that His₆-CpkD remains partially folded upon boiling, and thus is an exceptionally stable protein. This therefore accounts for the apparent catalytic activity observed (figure 3.26).

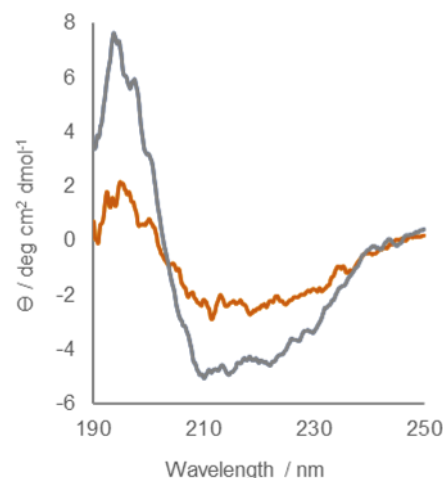
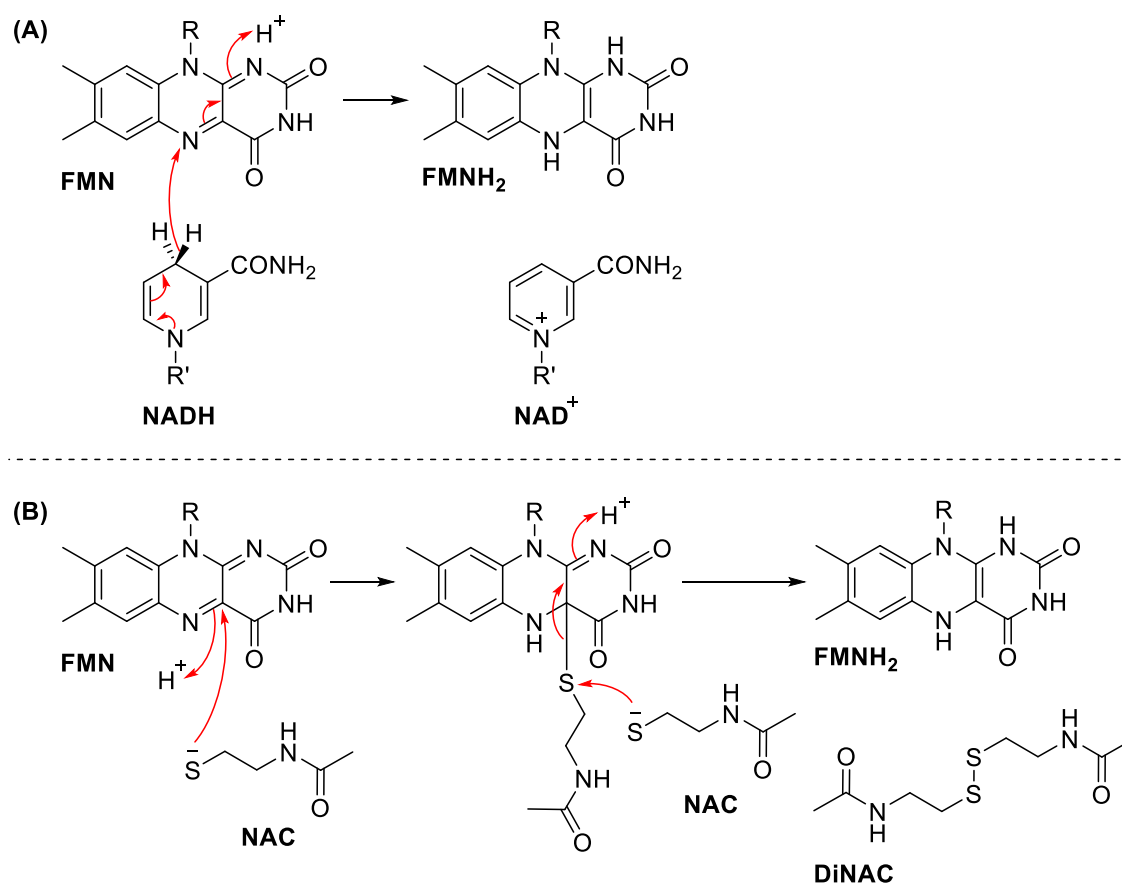


Figure 3.26: CD spectra of non-denatured (grey) and boiled (orange) His₆-CpkD



Scheme 3.23: Mechanism of FMN reduction by NADH (A) and NAC (B)

The same peaks were also evident in the reaction without NADH. This was unexpected, as reduced flavin is necessary for monooxygenase activity. However, thiols are also known to reduce flavin, *via* the formation of a thiol-C(4a) flavin adduct.^{201,202} Therefore, NAC may be the responsible reducing agent (scheme 3.23). Indeed, a molecular ion corresponding to the oxidised NAC dimer (diNAC, $m/z = 259.0545$) could be detected in assays in which NAC had been added (figure 3.27). The presence of diNAC within the assay from which FMN was omitted can be attributed to the auto-oxidation of NAC; the NAC stock was determined to contain 10 % diNAC by NMR.

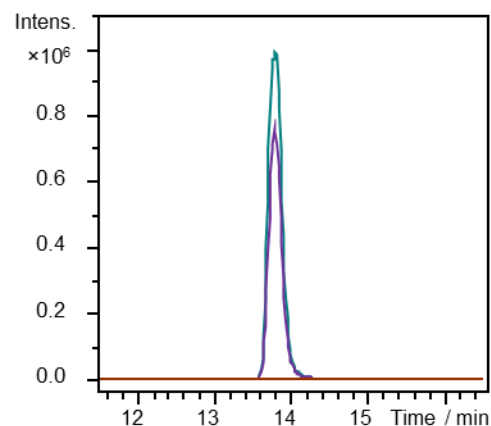


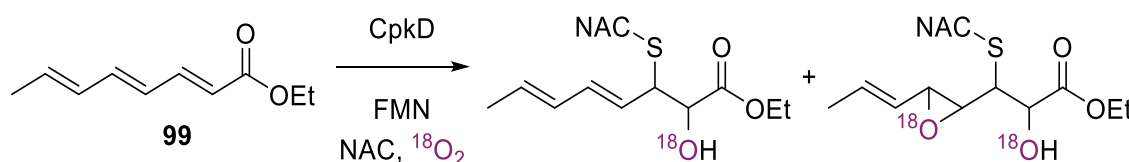
Figure 3.27: Extracted ion chromatograms at $m/z = 259.0545 \pm 0.005$ (corresponding to the $[M+Na]^+$ ion for diNAC). Green trace: incubation of His₆-CpKd with triene ester **99**, FMN, NADH and NAC; purple trace: control reaction from which FMN was omitted; and brown trace: control reaction from which NAC was omitted.

3.2.5.3 Tandem mass spectrometry and $^{18}\text{O}_2$ incorporation experiments

In collaboration with Dr. Lijiang Song, University of Warwick

To gain insight into the structures of the compounds, the molecular ions at $m/z = 324$ and 340 were subject to tandem mass spectrometry (Dr. Lijiang Song, University of Warwick). The data seemed promising, with fragment ions corresponding to plausible molecular formulae (see appendix 9). However, it proved challenging to propose mechanistically-viable structures.

To further probe the origin of the compounds, the assay was repeated under an $^{18}\text{O}_2$ atmosphere (scheme 3.24). Because monooxygenases transfer one atom of molecular oxygen per catalytic reaction, any resulting mono- or bis-epoxide products should contain one or two atoms of ^{18}O , respectively. This would result in an increase in intensity of peaks at +2 or +4 Da by HRMS.



Scheme 3.24: Putative epoxidation of **99** (200 μM) by CpkD (20 μM) and FMN (40 μM), followed by derivatisation with NAC (400 μM), to yield NAC-adducts of a mono- and bis-epoxide (possible structures are shown). Under an $^{18}\text{O}_2$ -atmosphere, one or two atoms of ^{18}O would be expected to be incorporated into the mono and bis-epoxide derivatives, respectively, resulting in an increase in intensity of peaks of 326 and 344 Da.

Therefore, $^{18}\text{O}_2$ was bubbled through triene ester **99**, His₆-CpkD, FMN and NAC, prior to, and during incubation. Analysis of the extracted ion chromatograms for the mono-epoxide derivative revealed a substantial increase in intensity of a peak at +2 Da ($m/z = 326$), relative to the control reaction performed in air (figure 3.28). Similarly, the extracted ion chromatograms for the bis-epoxide derivative indicated an increase in intensity of peaks at +2 and +4 Da ($m/z = 342$ and 344) (figure 3.29). The peaks at $m/z = 324$ and 340 , corresponding to the ^{16}O isotopes, had decreased in intensity, accordingly.

It should be noted that there is not 100 % ^{18}O incorporation, likely due to ^{16}O isotopic impurities within the canister (~3 %) and difficulties associated with complete exclusion of air from the reaction mixtures.

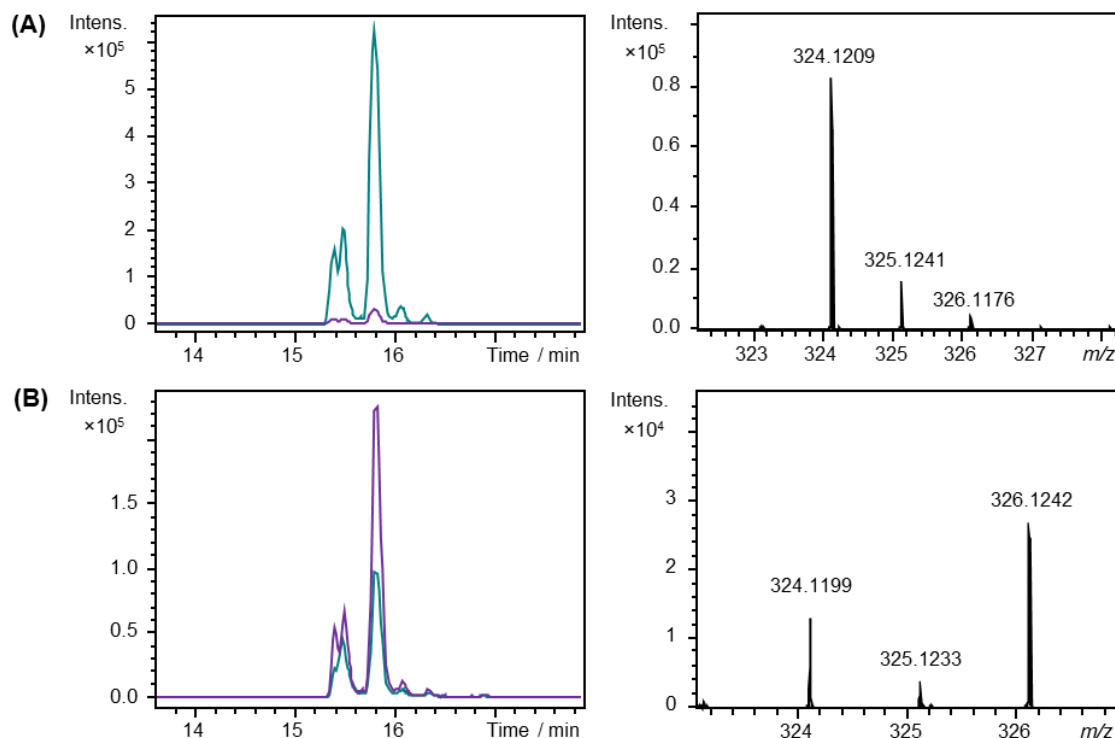


Figure 3.28: Extracted ion chromatograms from LC-MS analysis of reactions of CpkD, FMN, triene ester **99** and NAC, in air (A) and under $^{18}\text{O}_2$ (B). The extracted ion chromatograms at $m/z = 324.1240 \pm 0.05$ (green) and $m/z = 326.1240 \pm 0.05$ (purple) correspond to the ^{16}O and ^{18}O isotopes of the mono-epoxide, respectively. Independent assays were performed in duplicate.

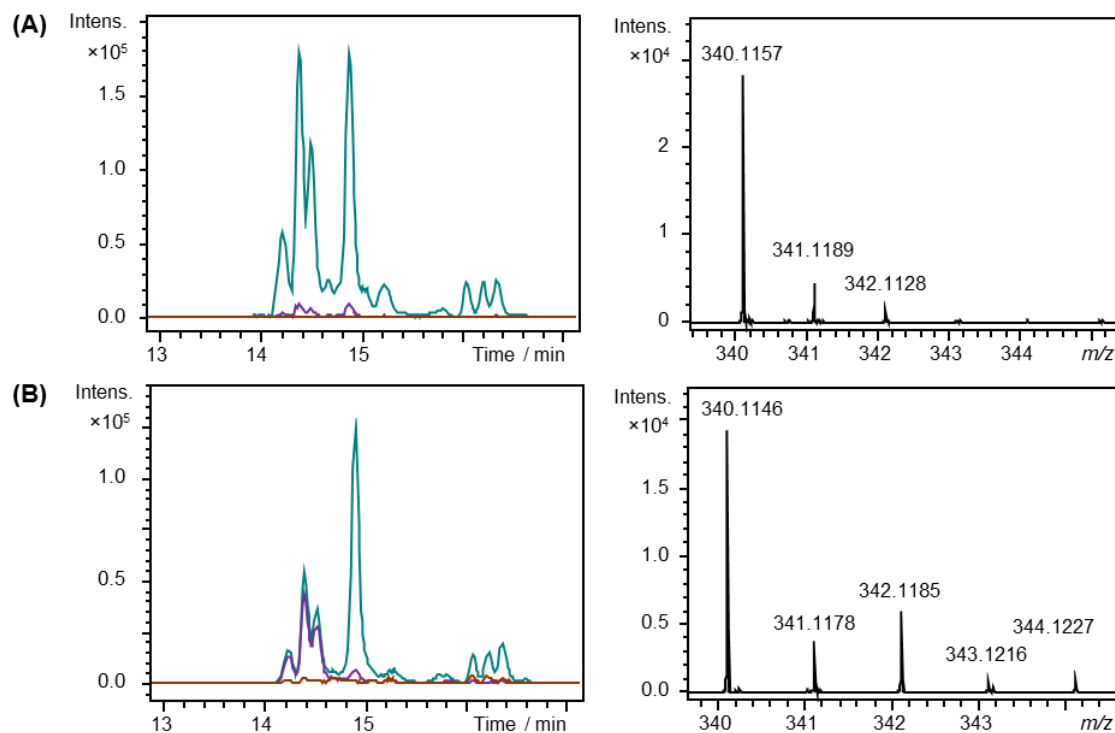


Figure 3.29: Extracted ion chromatograms from LC-MS analysis of reactions of CpkD, FMN, triene ester **99** and NAC, in air (A) and under $^{18}\text{O}_2$ (B). The extracted ion chromatograms at $m/z = 340.1189 \pm 0.05$ (green), $m/z = 342.1189 \pm 0.05$ (purple) and $m/z = 344.1189 \pm 0.05$ (brown) correspond to the ^{16}O , $^{18}\text{O}_1$ and $^{18}\text{O}_2$ isotopes of the bis-epoxide, respectively. Independent assays were performed in duplicate.

To quantify the incorporation of $^{18}\text{O}_2$ into the NAC-derivatised mono-epoxide, the intensities of the peaks at $m/z = 324$ (*A*) and 326 (*A*+2) for the reaction and control mixtures were recorded (table 3.1). The intensity of the *A*+2 peak in the control experiment was then subtracted from the analogous peak in the $^{18}\text{O}_2$ reaction, to account for contributions from naturally occurring isotopes (e.g. $^{13}\text{C}_2$, ^{34}S) (table 3.2, figure 3.30a). The ratio of unlabelled product : singly-labelled product was determined to be 1 : 1.91.

Similarly, to quantify the incorporation of $^{18}\text{O}_2$ into the NAC-derivatised bis-epoxide, the intensities of the peaks at $m/z = 340$ (*B*), 342 (*B*+2) and 344 (*B*+4) for the reaction and control mixtures were recorded (table 3.1). The values were then corrected, as described previously. However, for the doubly-labelled product, isotopic contributions from both the unlabelled and singly-labelled product were considered. Because the latter could not be calculated, the contribution was assumed to be equal to that of the unlabelled product to the singly-labelled product ($100 : 5.16 = 28.4 : 1.47$) (table 3.2, figure 3.30b). The ratio of unlabelled : singly-labelled : doubly-labelled product was determined to be 1 : 0.28 : 0.05. It is unclear why the incorporation of ^{18}O is lower in the bis-epoxide, compared to the mono-epoxide.

Table 3.1: Observed and relative intensities of peaks corresponding to isotopes of mono-epoxide ($m/z = 324$ and 326) and bis-epoxide ($m/z = 340$, 342 and 344) derivatives, from reactions in air and under $^{18}\text{O}_2$

	<i>m/z</i>	Control (in air)		Reaction (under $^{18}\text{O}_2$)	
		Observed intensity / a.u.	Relative intensity / %	Observed intensity / a.u.	Relative intensity / %
Mono-epoxide	324 (<i>A</i>)	29600	100	13400	100
	326 (<i>A</i> +2)	1370	4.63	26100	195
Bis-epoxide	340 (<i>B</i>)	9270	100	8970	100
	342 (<i>B</i> +2)	478	5.16	3010	33.6
	344 (<i>B</i> +4)	67.0	0.720	646	7.20

Overall, these data support the incorporation of ^{18}O into the enzymatic products, and thus imply that CpkD functions as a monooxygenase. It therefore seems likely that CpkD

catalyses both epoxidations during coelimycin biosynthesis, leaving the role of ScF open to speculation. Future work could focus on the optimisation of the isotope incorporation experiments and the synthesis of the proposed true triene intermediate, **29**, for use as a substrate. Furthermore, the enzymatic products could be purified by HPLC and characterised by NMR, or a variety of authentic standards could be synthesised for LC-MS comparisons.

Table 3.2: The corrected relative intensities of peaks corresponding to mono- and bis-epoxide derivatives, following subtraction of isotopic contributions (e.g. $^{13}\text{C}_2$, ^{34}S , $^{13}\text{C}_4$, $^{34}\text{S}_2$, $^{13}\text{C}_2^{34}\text{S}$) from unlabelled and singly-labelled products

	<i>m/z</i>	Relative intensity / %	Isotopic contributions		Corrected relative intensity / %
			From unlabelled product / %	From singly-labelled product / %	
Mono-epoxide	324 (<i>A</i>)	100	-	-	100
	326 (<i>A</i> +2)	195	4.63	-	191
Bis-epoxide	340 (<i>B</i>)	100	-	-	100
	342 (<i>B</i> +2)	33.6	5.16	-	28.4
	344 (<i>B</i> +4)	7.20	0.720	1.47	5.01

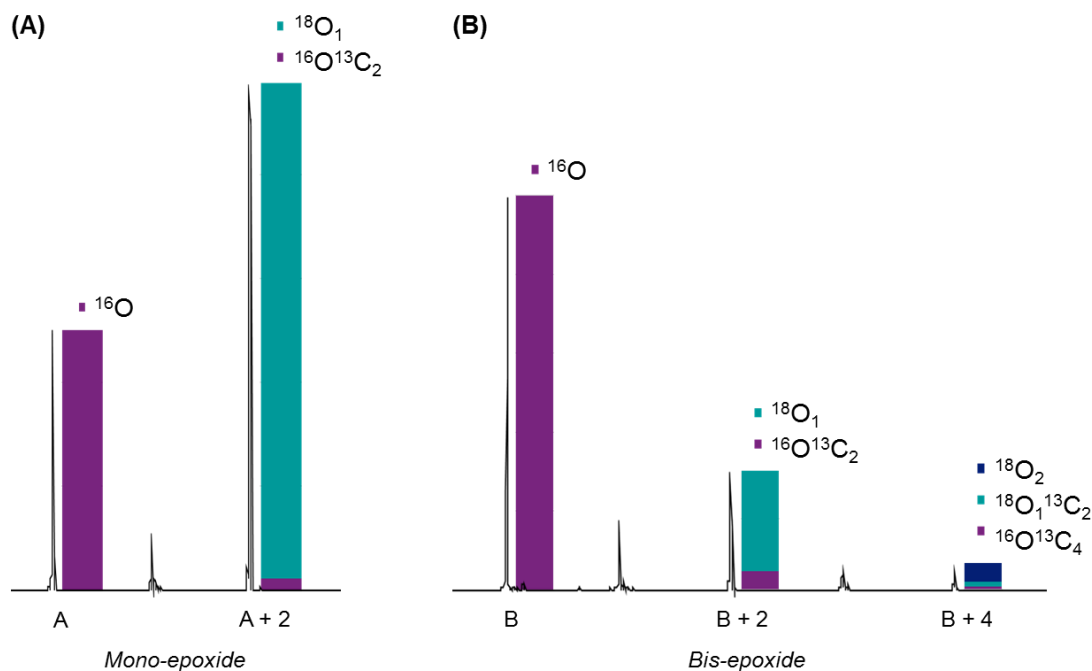


Figure 3.30: Histograms depicting the relative contributions of isotopes to the *A*+2, *B*+2 and *B*+4 peaks, following reaction of triene ester **99** with His₆-CpkD, FMN and NAC under an $^{18}\text{O}_2$ -enriched atmosphere. For simplicity, contributions of naturally occurring isotopes are annotated as $^{13}\text{C}_2$ and $^{13}\text{C}_4$ responses. Calculations and data display are adapted from Johnson and coworkers.²⁰³

Analysis of the sequence of CpkD and *in silico* predictions using Cofactory²⁰⁴ indicate that the enzyme contains a fingerprint ‘DG’ motif and a single Rossmann fold.²⁰⁵ Together with the fact that CpkD is encoded by a single gene, and is capable of binding NADH, this suggests that CpkD belongs to the group A subcategory of monooxygenases.⁸³ These enzymes are normally involved in the hydroxylation of aromatic substrates, although a few examples of epoxidases have also been reported.^{206,207,208} Interestingly, CpkD appears to be the first example of a group A monooxygenase capable of binding FMN, in addition to FAD. Crystallisation studies could aid in the elucidation of the mode of binding of FMN, NADH and the triene substrate, and could provide insight into the epoxidation mechanism.

3.3 Summary

The coelimycin post-PKS tailoring genes (*cpkI*, *cpkE*, *cpkD*, *scF* and *cpkH*) were cloned into pET151 and the corresponding recombinant proteins overproduced in *E. coli* and purified. All proteins were determined to be pure and of the correct molecular weight by SDS-PAGE and HRMS analysis.

His₆-CpkI and His₆-CpkE were incubated with synthetic dodecane-1,5-diol **76** and *trans*-octen-2-al **78**, to investigate their potential oxidoreductase and isomerase activities, respectively. Products could not be detected, suggesting that the assays require optimisation, or that the enzymes catalyse other reactions. A multiple sequence alignment of CpkE with other members of the α , β -hydrolase fold superfamily suggested that it may function as an epoxide hydrolase.

His₆-CpkD, His₆-ScF and His₆-CpkH were all characterised as flavoproteins by UV and HRMS analysis. Specifically, His₆-CpkD was observed to bind both FAD and FMN non-covalently, while His₆-ScF and His₆-CpkH were found to bind FAD covalently. His₆-CpkH was subsequently shown to oxidise racemic 1-aminododecan-5-ol **60** to 6-heptyl-2,3,4,5-tetrahydropyridine **81**, and was thus characterised as an FAD-dependent dehydrogenase. To probe the stereo- and configurational selectivity of CpkH, four further substrates ((*R*)-**60**, (*S*)-**60**, (*R,E*)-**84** and (*S,E*)-**84**) were synthesised in excellent yields over 10-12 steps. The enzyme specifically oxidised (*R*)-**60** rather than (*S*)-**60**, in agreement with KR predictions and the proposed coelimycin biosynthetic pathway.

Analogous substrates containing the *trans*-double bond were not tolerated in assays with CpkH alone, nor coupled assays with CpkE or CpkI. This led to the proposal that the released polyketide product may contain a saturated carbon-carbon bond or a 3'-hydroxyl functional group.

Preliminary evidence from assays with synthetic triene ester **99**, in addition to fragmentation studies and $^{18}\text{O}_2$ incorporation experiments, implicate CpkD as an FMN-dependent monooxygenase. Future work could involve purifying the enzymatic products or synthesising a range of stereo- and regioisomeric standards for LC-MS comparisons. Furthermore, crystallisation studies could assist in understanding of the mode of cofactor and substrate binding, and the mechanism of epoxidation. The role of ScF in coelimycin biosynthesis remains to be investigated.

4 CONCLUSIONS AND PERSPECTIVES

4.1 Biochemical and structural characterisation of CpkG

CpkG was characterised as an ω -transaminase with a broad substrate tolerance. The His₆-tagged enzyme readily catalysed the conversion of several linear, branched, cyclic and aromatic amines to the corresponding aldehydes, utilising two different amino acceptors. The highest level of activity was observed for octylamine; the compound closest in structure to the proposed substrate of CpkG in the coelimycin biosynthetic pathway. Further complementary assays allowed for the detection of other products of the transamination reaction. The conversion of synthetic substrate mimic **59** to **60**, and vice versa, demonstrated that CpkG is responsible for nitrogen incorporation into coelimycin.

Crystal structures revealed that His₆-CpkG is homodimeric, with a rare tri-domain architecture. Comparison of the structures and analysis of the active site allowed for the identification of key residues and provided insight into the transamination mechanism. Soaking of various amines and carbonyl compounds (e.g. 1-aminododecan-5-ol **60** and 5-hydroxydodecanal **59**) into pre-existing crystals could determine the mode of substrate binding within CpkG, while molecular modelling could then unveil the substrate scope. The overall fold-type I conformation, in conjunction with Marfey's analysis, indicated that CpkG is an (*S*)-selective ω -transaminase. The unusual additional domain prompted binding assays with CpkC-TR, however an interaction could not be observed. A multiple sequence alignment of CpkG with other ω -transaminases involved in polyketide alkaloid biosynthesis indicated that the additional domain in CpkG could be non-functional.

Future work could examine the potential of CpkG as a biocatalyst for the generation of chiral amines from ketones. Interestingly, an engineered (*R*)-selective ω -transaminase has been recently employed in the large-scale production of the anti-diabetic sitagliptin.²⁰⁹

4.2 Identification of further polyketide alkaloid biosynthetic gene clusters

Investigations with CpkC-TR and CpkG led to the hypothesis that thioester reduction and transamination may be a conserved sequence of reactions in polyketide alkaloid biosynthesis. Comprehensive computational searches uncovered 22 cryptic biosynthetic

gene clusters from actinobacterial genomes that encode homologues of CpkC-TR and CpkG, in addition to a type I modular PKS. Based on PKS sequence predictions, thirteen gene clusters were proposed to direct the biosynthesis of known polyketide alkaloids, while the remaining nine are likely associated with novel metabolites. Attempts to detect one such metabolite from *Streptomyces cyaneofuscatus* NRRL B-2570 were unsuccessful; SARP overexpression instead resulted in the overproduction of the unrelated antibiotic valinomycin. Other strategies could explore the deletion of repressor genes, promoter exchange or TAR cloning.

Sequencing the genome of the known latumcidin producer, *Streptomyces abikoensis* DSM 40831, led to the identification of a gene cluster containing *cpkC-TR* and *cpkG* homologues, as predicted. Furthermore, two of the gene clusters detected from the bioinformatics search have since been linked to streptazone E and cyclizidine production. These findings strongly support the hypothesis that thioester reduction and transamination are universal steps in polyketide alkaloid biosynthesis. Future work could involve generating gene-knockouts to link latumcidin production to the putative gene cluster in *S. abikoensis*, as well as exploring the vast biosynthetic potential encoded within the additional 73 clusters. Moreover, the genetic modification of polyketide biosynthetic gene clusters to include a *cpkG* homologue, and a TE-TR domain swap, could allow for the diversion of polyketide pathways towards the production of a range of novel and structurally-diverse polyketide alkaloids.

4.3 *In vitro* investigations with CpkE and CpkI

The coelimycin biosynthetic genes *cpkI* and *cpkE* were cloned into pET151, and the corresponding *N*-terminal His₆ fusion proteins overproduced in *E. coli* and purified. The proteins were determined to be pure and of the correct molecular weight by SDS-PAGE and HRMS analysis.

The potential oxidoreductase and isomerase activities of His₆-CpkI and His₆-CpkE were investigated by incubating the enzymes with synthetic dodecane-1,5-diol **76**, *trans*-octen-2-al **78**, and both enantiomers of (*E*)-1-aminododec-2-en-5-ol **84**, as appropriate. Products

could not be observed, suggesting that the enzymes are very substrate-specific, or catalyse alternative biosynthetic steps.

Examination of the sequence of CpkE revealed the presence of an aspartic acid residue in place of the conserved catalytic serine residue, indicating that it may function as an epoxide hydrolase. Additionally, CpkI might act as an external enoyl reductase, to reduce the carbon-carbon double bond of thioester **22** during polyketide chain assembly. Future work could therefore involve the synthesis of epoxide-containing substrates for incubation with CpkE, and CoA-thioesters harbouring *trans*-2,3-double bonds, for loading onto the CpkC-ACP domain prior to incubation with CpkI. Alternatively, $\Delta cpkE$ and $\Delta cpkI$ deletion mutants could be generated, to probe the roles of the corresponding enzymes *in vivo*.

4.4 *In vitro* investigations with CpkD, ScF and CpkH

The oxidative genes *cpkD*, *scF* and *cpkH* were cloned, overexpressed and the corresponding fusion proteins purified to homogeneity. All were characterised as flavoproteins by UV and HRMS analysis; His₆-CpkD was found to bind both FAD and FMN non-covalently, while His₆-ScF and His₆-CpkH were observed to bind FAD covalently. A multiple sequence alignment of His₆-ScF and His₆-CpkH with TamL revealed that the mode of binding may be bicovalent, *via* conserved cysteine and histidine residues. This could be probed through crystallisation studies.

His₆-CpkH was demonstrated to oxidise synthetic racemic 1-aminododecan-5-ol **60** to 6-heptyl-2,3,4,5-tetrahydropyridine **81**, and thereby characterised as an FAD-dependent dehydrogenase, in line with data obtained from *in vivo* analysis of a *cpkH* deletion mutant. To investigate the stereo- and configurational selectivity of CpkH, four further prospective substrates ((*R*)-**60**, (*S*)-**60**, (*R,E*)-**84** and (*S,E*)-**84**) were synthesised over 10-12 steps in excellent yield. The enzyme was observed to exclusively oxidise (*R*)-**60**, rather than (*S*)-**60**, consistent with KR stereochemical predictions, but could not tolerate compounds with a *trans*-double bond. Future work could explore the synthesis of analogues containing a 3'-hydroxyl functional group or *cis*-double bond, to gain further insight into the nature of the pathway intermediates and the order of the biosynthetic steps.

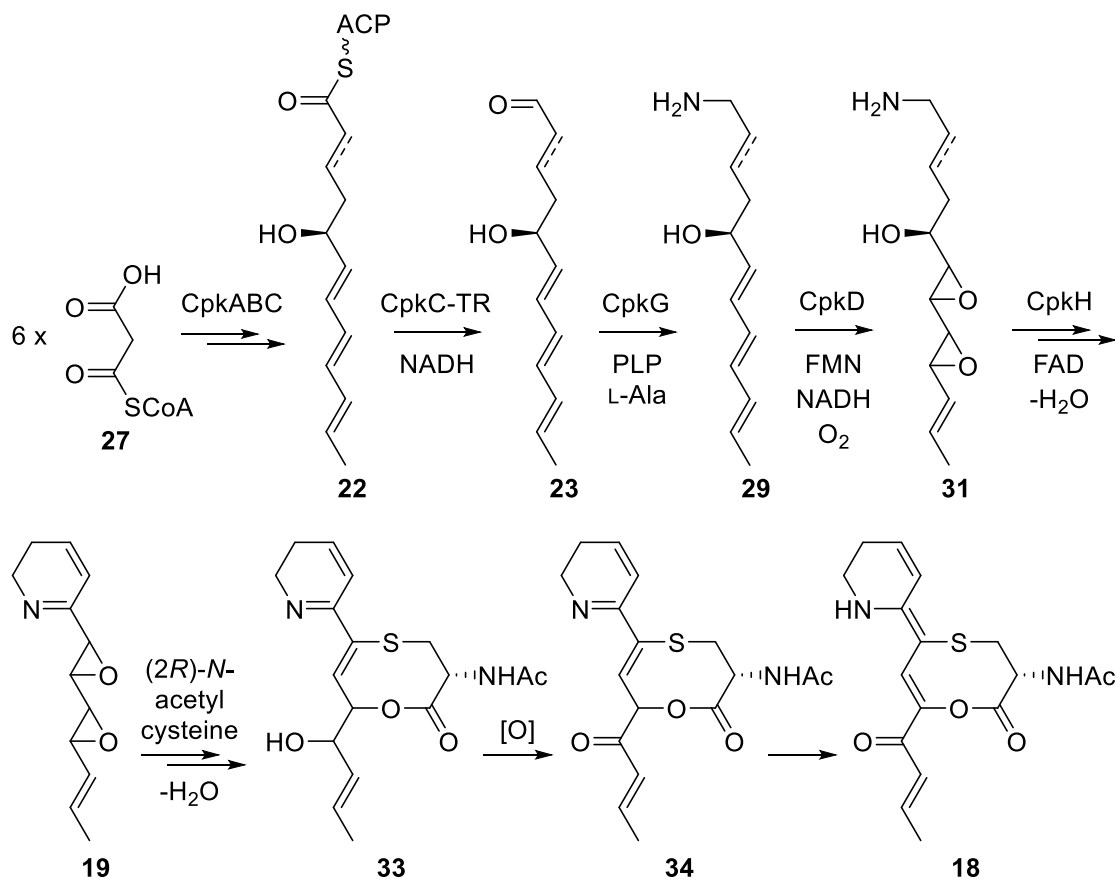
Initial investigations with His₆-CpkD and triene ester **99** suggested that the enzyme catalyses the formation of several mono- and bis-epoxides, utilising FMN as a preferred cofactor. Fragmentation studies and ¹⁸O₂ incorporation experiments supported the apparent monooxygenase activity observed. Further investigations could focus on the purification and characterisation of the enzymatic products, the synthesis of various standards for LC-MS comparisons, or the synthesis of more accurate substrate mimics.

As CpkD appears to catalyse both epoxidation reactions in coelimycin biosynthesis, the role of ScF is somewhat unclear. However, because the protein is secreted, and its closest homologue is CpkH, it may catalyse a similar, late-stage dehydrogenation reaction. This would be difficult to investigate, as the reaction is likely not observable due to the interception of the pathway by amino acid adducts (*N*-acetylcysteine and glutamate). It remains possible that ScF catalyses the oxidation of the 9'-hydroxyl group of **33** (or a similar intermediate) to the corresponding ketone (scheme 4.1). This could be probed by synthesising a 9'-hydroxyl-containing substrate analogue.

4.5 Updated biosynthetic pathway

Following *in vitro* investigations with the post-PKS tailoring enzymes, the coelimycin biosynthetic pathway was updated (scheme 4.1). Initially, a hexaketide thioester **22** is assembled by modular PKS enzymes CpkA, CpkB and CpkC, from six molecules of malonyl-CoA **27**. This thioester may contain a *trans*-2,3-double bond, a 3'-hydroxyl group or a saturated carbon-carbon bond, depending on the activity of the final DH domain and the actions of any external ERs (such as CpkI). The ACP-bound thioester is then reductively cleaved by CpkC-TR and NADH to yield aldehyde **23**, which may be further reduced and re-oxidised (possibly catalysed by CpkI or an alternative enzyme). The aldehyde is then subject to reductive amination by CpkG, which utilises PLP and an amino acid (e.g. L-alanine) as co-substrates. Amine **29** then undergoes two epoxidations, catalysed by CpkD and necessary cofactors O₂, FMN and NADH. Subsequently, CpkH catalyses an FAD-mediated oxidation of the hydroxyl group of **31**, and a spontaneous intramolecular cyclisation yields coelimycin A **19**. At some stage prior to the formation of **19**, a further transformation is necessary, depending on the nature of the dashed bond.

Addition of *N*-acetylcysteine and loss of water affords intermediate **33**, which undergoes a final oxidation (possibly catalysed by ScF) and tautomerisation to form coelimycin P1 **18** (scheme 4.1).



Scheme 4.1: Updated proposed biosynthesis of coelimycin A **19** and coelimycin P1 **18**. [O] represents an unknown dehydrogenase.

Further investigations are necessary to determine the structure of polyketide thioester **22**, the order of the oxidative steps and the enzymatic roles of CpkI, CpkE and ScF. Complete elucidation of coelimycin biosynthesis could allow for the re-engineering of the pathway in the future, for the generation of novel and stable bioactive analogues.

5 EXPERIMENTAL

5.1 Instruments and equipment

Autoclaving was performed at 121 °C and 1 bar for 10 min. Liquid and solid cultures were incubated in New Brunswick Scientific Innova shakers or Thermo Scientific Heraeus static incubators, respectively. Optical densities were measured using a Thermo Scientific BioMate 3 spectrophotometer. Centrifugations were performed using Eppendorf models 5415D, 5424, 5804R and 5810R, or a Sorvall RC 6 Plus equipped with an SS34 or SLA3000 rotor. *Streptomyces* strains were handled in a Bassaire laminar flow hood. PCR reactions were performed using an Eppendorf Mastercycler Gradient or Eppendorf Mastercycler Personal. DNA and protein concentrations were determined using a Thermo Scientific NanoDrop Lite spectrophotometer. Agarose gel electrophoresis was performed using a BioRad tank in conjunction with either a PowerPac™ Basic or a PowerPac™ 300. Visualisation was achieved with a UVP BioDoc-It® Imaging System 2UV Transilluminator. Electroporations were conducted with a BioRad Gene Pulser® II (200 Ω, 25 μF, 1.8 kV) in combination with a BioRad Pulse Controller Plus. Cell lysis was performed using a Constant Systems Ltd TS-Series Cabinet cell disruptor (One-Shot mode, 20 KPSI). Proteins were analysed by SDS-PAGE gel electrophoresis using a BioRad Mini-PROTEAN® Tetra cell. UV-Vis measurements of solutions in 3 mm path length sub-micro quartz cuvettes (Starna scientific) were recorded using a Varian Cary 50 Bio UV-Visible Spectrophotometer (single wavelength) or a Perkin Elmer® Lambda 3S UV/Vis Spectrophotometer (multiple wavelengths). UV-Vis measurements of solutions in 96-well plates were recorded using a Tecan GENios microplate reader.

Solvents were evaporated using a BUCHI Rotavapor R-200 or R-210 connected to a BUCHI Vacuum Pump V-700. Melting points were obtained by a Stuart Scientific SMP10 instrument. NMR spectra were measured using a Bruker DPX-300 (¹H: 300 MHz, ¹³C: 75 MHz) or DPX-400 (¹H: 400 MHz, ¹³C: 100 MHz) instrument, both equipped with a proton carbon dual probe. High field NMR spectra were obtained following submission to the University of Warwick NMR service, using a Bruker AV-500 (¹H: 500 MHz, ¹³C: 125 MHz), AV III-600 (¹H: 600 MHz, ¹³C: 150 MHz) or AV II-700 (¹H: 700 MHz, ¹³C: 175 MHz) instrument. ¹H NMR data are reported as follows: chemical shift (δ, ppm), multiplicity (singlet (s); doublet (d); triplet (t); quartet (q); quintet (quint); doublet of doublets (dd); doublet of doublet of doublet of doublets (dddd); doublet of triplets (dt); triplet of doublets (td); broad (br.); multiplet (m)), integration, assignment and coupling

constant (Hz). Assignments were assisted by use of COSY, HMQC and HMBC spectra. All low-resolution mass spectra were obtained on an Agilent 6130B Single Quad ESI spectrometer while high resolution mass spectra were recorded on a Bruker MaXis ESI spectrometer by the University of Warwick Mass Spectrometry service.

Analytical reverse-phase high performance liquid chromatography (HPLC) was performed on an Agilent 1100 HPLC instrument connected with a variable wavelength detector and an Agilent ZORBAX Eclipse Plus C18 analytical column (4.6×150 mm, $5 \mu\text{m}$). Preparative reverse-phase HPLC was performed on an Agilent 1200 or 1260 Infinity HPLC instrument connected with a diode array detector and an Agilent PrepHT XDB-C18 or BetaSil C18 preparative column (21.2×150 mm, $5 \mu\text{m}$), respectively. Liquid chromatography-high resolution mass spectrometry (LC-HRMS) was performed using a Thermo Scientific Dionex Ultimate 3000 RS UHPLC coupled to a Bruker MaXis Impact ESI-TOF mass spectrometer using an Agilent ZORBAX Eclipse XDB-C18 analytical column (2.1×5 mm, $1.8 \mu\text{m}$) for small molecules, or an Advanced Chromatography Technologies ACE 3 C4-300 reverse phase column (100×2.1 mm) for proteins. Sodium formate (10 mM) was used for internal calibration.

5.2 Materials

5.2.1 General

Bacto yeast extract, Bacto agar, Bacto Tryptic Soy Broth and tryptone were all purchased from Becton Dickinson and Company (USA). GelRed Nucleic Acid Gel Stain and InstantBlue were purchased from Cambridge Bioscience (UK) and Expedeon (UK), respectively. All other chemical reagents were purchased from Sigma Aldrich (USA), Thermo Fisher Scientific (UK) or Alfa Aesar (USA) and used without further purification unless otherwise stated. Solvents were purchased from Thermo Fisher Scientific unless otherwise stated. Anhydrous solvents were obtained by distillation under argon using calcium hydride as a drying reagent, and stored over 4 \AA molecular sieves. TLC was performed using aluminium plates pre-coated with silica gel 60 F254 (Merck, Germany) and flash chromatography was performed on silica gel (Aldrich, $40\text{-}63 \mu\text{m}$, 60 \AA). UV light or potassium permanganate staining were used for compound detection. All water used was deionised, unless stated otherwise.

5.2.2 Kits and enzymes

Kits used for DNA purification, plasmid DNA extraction, TOPO® cloning and ligations were all purchased from Thermo Fisher Scientific (UK). Restriction enzymes, corresponding buffers and *Taq* polymerase were also purchased from Thermo Fisher Scientific. Kits used for trypsin profile in-gel digests were purchased from Sigma Aldrich (USA). Phusion polymerase was obtained from New England Biolabs (USA). T4 DNA ligase and buffers were acquired from Roche (Germany).

5.2.3 Growth and production media

Luria-Bertani Medium (LB): LB Broth powder (Miller's) was dissolved in water to a concentration of 25 g/L and sterilised by autoclaving. To prepare LB agar, 15 g/L Bacto agar was added prior to sterilisation.

Tryptone Soya Broth (TSB): Bacto Tryptic Soy Broth (30 g/L) was dissolved in water and sterilised by autoclaving.⁹

Soya Flour Mannitol Medium (SFM): Soya flour (20 g/L) and D-mannitol (20 g/L) were dissolved in tap water. Bacto agar (20 g/L) was added and the contents were sterilised by autoclaving.⁹

International *Streptomyces* Project Medium 2 (ISP2): Difco yeast extract (4 g/L), Merck malt extract (10 g/L) and glucose (4 g/L) were dissolved in water and the pH of the solution adjusted to 7.3. Bacto agar (20 g/L) was added and the contents were sterilised by autoclaving.²¹⁰

International *Streptomyces* Project Medium 4 (ISP4): ISP4 powder was dissolved in water to a concentration of 37 g/L and sterilised by autoclaving. To prepare ISP4 agar, 20 g/L Bacto agar was added prior to sterilisation.

GYM *Streptomyces* Medium: Glucose (4 g/L), Difco yeast extract (4 g/L), Merck malt extract (10 g/L) and CaCO₃ (2 g/L) were dissolved in water and the pH of the solution

adjusted to 7.2. Bacto agar (12 g/L) was added and the contents were sterilised by autoclaving.²¹¹

R3 Medium: Glucose (10 g/L), Difco yeast extract (5 g/L), Difco casamino acids (0.1 g/L), L-proline (3 g/L), MgCl₂.6H₂O (10 g/L), CaCl₂.2H₂O (4 g/L), K₂SO₄ (0.2 g/L), KH₂PO₄ (0.05 g/L), TES (5.6 g/L) and trace element solution* (2 mL/L) were dissolved in water. The pH of the solution was adjusted to 7.2 with 1 M NaOH and the contents were sterilised by autoclaving. To prepare R3 agar, 22 g/L Bacto agar was added prior to sterilisation.²¹²

R5 Medium: Sucrose (103 g/L), K₂SO₄ (0.25 g/L), MgCl₂.6H₂O (10.12 g/L), glucose (10 g/L), Difco casamino acids (0.1 g/L), trace element solution* (2 mL/L), Difco yeast extract (5 g/L) and TES (5.73 g/L) were dissolved in water and sterilised by autoclaving. To prepare R5 agar, 22 g/L Bacto agar was added prior to sterilisation. At the time of use, 0.5 % KH₂PO₄ (10 mL/L), 5 M CaCl₂.2H₂O (4 mL/L), 20 % L-proline (15 mL/L) and 1 N NaOH (7 mL/L) were added in the order listed.⁹

Alkaloid Medium 1 (AM1): Glucose (20 g/L), peptone (5 g/L), meat extract (5 g/L), yeast extract (3 g/L), CaCO₃ (3 g/L) and NaCl (5 g/L) were dissolved in water. The pH of the solution was adjusted to 7.0 and the contents were sterilised by autoclaving.²¹³

Supplemented Minimal Medium (SMM): Difco casamino acids (2 g/L) and TES (5.73 g/L) were dissolved in water and the pH of the solution adjusted to 7.2 with 5 M NaOH. The contents were then sterilised by autoclaving. At the time of use, 50 mM NaH₂PO₄/K₂HPO₄ (10 mL/L), 1 M MgSO₄ (5 mL/L), glucose (50 % w/v, 18 mL/L) and trace element solution* (1 mL/L) were added in the order listed.⁹

*Trace element solution: ZnCl₂ (40 mg/L), FeCl₃.6H₂O (200 mg/L), CuCl₂.2H₂O (10 mg/L), MnCl₂.4H₂O (10 mg/L), Na₂B₄O₇.10H₂O (10 mg/L) and (NH₄)₆Mo₇O₂₄.4H₂O (10 mg/L) were dissolved in water.⁹

5.2.4 Buffers

STE buffer: 100 mM NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA

TBE buffer (10 ×): 108 g Tris-HCl (pH 8.0-8.3), 55.0 g boric acid, 11.7 g EDTA, water to 1 L

Ni²⁺-affinity washing buffer: 100 mM NaCl, 20 mM Tris-HCl (pH 8.0), 20 mM imidazole

Ni²⁺-affinity elution buffers: 100 mM NaCl, 20 mM Tris-HCl (pH 8.0), 50-300 mM imidazole

Concentration buffer: 100 mM NaCl, 20 mM Tris-HCl (pH 8.0)

Concentration buffer (CpkG): 100 mM NaCl, 50 mM HEPES (pH 7.6), 20 μM PLP

Crystallisation buffer (CpkG): 20 mM NaCl, 50 mM HEPES (pH 7.8)

Storage/gel filtration buffer: 100 mM NaCl, 20 mM Tris-HCl (pH 8.0), 10 % glycerol

Phosphate buffer (100 mM, pH 7.0): 1 M K₂HPO₄ (61.5 mL), 1 M KH₂PO₄ (38.5 mL), water to 1 L

SDS-PAGE sample buffer (2 ×)*: 500 mM Tris-HCl (2.5 mL, pH 6.8), 100 % glycerol (2.0 mL), β-mercaptoethanol (400 μL), bromophenol blue (20.0 mg), SDS (400 mg), water to 10 mL

SDS-PAGE running buffer (10 ×)*: 250 mM Tris-HCl (pH 8.8), 2 M glycine, 1 % SDS

*For preparation of native-PAGE sample and running buffers, β-mercaptoethanol and SDS were omitted

5.2.5 Antibiotics

Table 5.1: Antibiotics used within this project

Antibiotic	Stock conc. (mg/mL)	Working conc. (µg/mL)	Solvent	Supplier
Ampicillin	100	100	H ₂ O	Melford (UK)
Hygromycin	50	25	H ₂ O	Roche (Germany)
Kanamycin	50	50	H ₂ O	Melford (UK)
Chloramphenicol	25	25	MeOH	Duchefa Biochemie (Holland)
Nalidixic acid	25	20	0.3 M NaOH	Duchefa Biochemie (Holland)
Apramycin	50	25	H ₂ O	PanReac AppliChem (Germany)

5.2.6 Bacterial strains

Table 5.2: Bacterial strains used within this project

Bacterial strain	Use	Supplier
<i>Escherichia coli</i>		
One Shot® TOP10 chemically competent cells	Transformations, cloning and plasmid DNA isolation	Thermo Fisher Scientific (UK)
One Shot® BL21 Star™ (DE3) chemically competent cells	High level recombinant protein expression	Thermo Fisher Scientific (UK)
ET12567/pUZ8002	Conjugations	Lab stock
<i>Streptomyces</i>		
<i>coelicolor</i> M145	PCR amplification of <i>cpk</i> genes	Lab stock
<i>cyaneofuscatus</i> NRRL B-2570	Attempted detection of novel polyketide alkaloid	ARS culture (NRRL) collection (USA)
<i>cyaneofuscatus</i> ::SARP	Attempted detection of novel polyketide alkaloid	Genetically engineered in this study
<i>cyaneofuscatus</i> PKS::am	Attempted detection of novel polyketide alkaloid	Genetically engineered in this study
<i>cyaneofuscatus</i> ::SARP-PKS::am	Attempted detection of novel polyketide alkaloid	Genetically engineered in this study
<i>abikoensis</i> DSM 40831	Genome sequencing for identification of latumcidin gene cluster	DSMZ culture collection (Germany)

5.2.7 Vectors

Table 5.3: Vectors used within this project

Vector	Use	Resistance	Supplier
pET151	Overexpression of proteins with an N-terminal His-tag	Ampicillin	Thermo Fisher Scientific (UK)
pOSV556t	Integrative plasmid for genetic complementation in <i>S. cyaneofuscatu</i> s NRRL-2570	Ampicillin and hygromycin	Pernodet Group (University of Paris-Sud, France)
pKC1132	Suicide plasmid for insertional inactivation of <i>IF17_RS0112385</i> via a single crossover	Apramycin	Lab stock

5.2.8 PCR Primers

PCR primers were designed as recommended by the TOPO® cloning manual for pET151 constructs, or to include restriction sites and/or ribosomal binding sites for pOSV556t and pKC1132 constructs. All primers were purchased from Sigma-Aldrich.

Table 5.4: PCR primers used within this project

Primer ID	Sequence (5' → 3')	Use
<i>cpkE_fw</i>	CACCATGACGGCGACCCGGATC	Amplification of <i>cpkE</i> from <i>S. coelicolor</i> M145 and ligation with pET151
<i>cpkE_rv</i>	TCAGTGGGTCTGCCGGTACGCG	
<i>cpkI_fw</i>	CACCATGACTGAGAGATTGAGGAGTTT	Amplification of <i>cpkI</i> from <i>S. coelicolor</i> M145 and ligation with pET151
<i>cpkI_rv</i>	TCAGAGGCCGAGGCCG	
<i>cpkD_fw</i>	CACCATGAGAAGAAAGCGCGAAGGCCG	Amplification of <i>cpkD</i> from <i>S. coelicolor</i> M145 and ligation with pET151
<i>cpkD_rv</i>	TCACCGCTGGGCTCCGGCGATC	
<i>cpkH_fw</i>	CACCGCACAGCCCTCACCCGCC	Amplification of <i>cpkH</i> from <i>S. coelicolor</i> M145 and ligation with pET151
<i>cpkH_rv</i>	TCAGGCCGGCTCGATACC	
<i>scF_fw</i>	CACCGCAGGCCCGAGGACCTC	Amplification of <i>scF</i> from <i>S. coelicolor</i> M145 and ligation with pET151
<i>scF_rv</i>	CTACGGCTTGCGGATGCC	

<i>IF17_RS0112435_fw</i>	AAGGGAAGCTTAGGAGGATATACATGCTGTTCC	Amplification of SARP (<i>IF17_RS0112435</i>) from <i>S. cyaneofuscatus</i> NRRL B-2570
<i>IF17_RS0112435_rv</i>	AAAGGGCTCGAGCTACCCGTACGCGGTCTCC	
<i>IF17_RS0112435_fw_seq_1</i>	CCGCGATGGCCGACCGGCT	Sequencing of middle region of pOSV556t:: <i>IF17_RS0112435</i>
<i>IF17_RS0112435_rv_seq_1</i>	CCGCAGCAGGCCAGCAGGAGG	
<i>IF17_RS0112435_fw_seq_2</i>	TACAAGCATAAAGCTGGCATGC	Sequencing of outer regions of pOSV556t:: <i>IF17_RS0112435</i>
<i>IF17_RS0112435_rv_seq_2</i>	AGGATCTGACCGACGCGGT	
<i>IF17_RS0112385_fw</i>	CACAAGCTTGGAGGACCTGTGGAACCT	Amplification of PKS region (<i>IF17_RS0112385</i>) from <i>S. cyaneofuscatus</i> NRRL B-2570
<i>IF17_RS0112385_rv</i>	CAGTCTAGAGGAGAGCCGTTCCACCAG	
<i>IF17_RS0112385_fw_seq</i>	GCTGCAAGGCGATTAAGTTG	Sequencing of pKC1132:: <i>IF17_RS0112385</i>
<i>IF17_RS0112385_rv_seq</i>	TGTGGAATTGTGAGCGGATAAC	

5.3 Biological procedures

5.3.1 Growth, preparation and storage of microbial cells

5.3.1.1 General culture conditions

E. coli and *Streptomyces* strains were grown at 37 °C and 30 °C, respectively, in selected media (see section 5.2.3) containing an appropriate antibiotic (see sections 5.2.5 and 5.2.7), if necessary. Liquid cultures were incubated with shaking at 180 rpm.

5.3.1.2 Preparation and storage of *Streptomyces* spore stock solutions

Single colonies of *S. coelicolor*, *S. cyaneofuscatus* or *S. abikoensis* were selected and spread onto GYM, SFM, ISP2 or ISP4 agar plates. The cultures were then incubated at 30 °C for one week. Sterile water (3 mL) was added to each plate and the surface scraped using sterile plastic spreaders to suspend the spores. The spore suspension was then filtered through sterile non-absorbent glass wool and the filtrate centrifuged (6000 rpm, 20 °C, 5 min) to pellet the spores. The majority of the supernatant was then decanted and to the remainder was added an approximate equal volume of 50 % glycerol. The spore stock solution was then stored at -78 °C.

For preparation of a spore stock of ARS-acquired *S. cyaneofuscatus* or DSMZ-acquired *S. abikoensis*, the freeze-dried cells were first resuspended in TSB media (500 μ L) and aliquots (100 μ L) were spread onto agar plates. The plates were then incubated at 30 °C for one week before selection of single colonies.

5.3.1.3 Preparation and storage of chemically competent *E. coli* cells

LB medium (5 mL) was inoculated with an *E. coli* BL21Star or TOP10 glycerol stock solution (100 μ L) and incubated with shaking at 37 °C overnight. Fresh LB medium (500 mL) was then inoculated with overnight culture (5 mL), and the cells incubated with shaking at 37 °C until $OD_{600} = 0.4$. The cells were subsequently incubated on ice for 20 min and harvested by centrifugation (3000 rpm, 4 °C, 10 min). The media was decanted, and the cells resuspended in cold 0.1 M $CaCl_2$ (50 mL) then left on ice for 30 min. The cells were then centrifuged as before, the media decanted, and the cells resuspended in cold $CaCl_2$ (5 mL) containing 15 % glycerol. The cells were divided into aliquots (100 μ L), left on ice for 1 h, flash frozen in liquid N_2 and stored at -78 °C.

5.3.1.4 Preparation and storage of electrocompetent *E. coli* cells

LB medium (5 mL) containing the appropriate antibiotic(s) (see sections 5.2.5 and 5.2.7) was inoculated with an *E. coli* ET12567/pUZ8002 glycerol stock solution (100 μ L) and incubated with shaking at 37 °C overnight. Fresh LB medium (5 mL) containing the appropriate antibiotic(s) was inoculated with overnight culture (100 μ L) and incubated with shaking at 37 °C until the $OD_{600} = 0.6$. *E. coli* cells were then harvested by centrifugation (4000 rpm, 4 °C, 8 min), the medium decanted and the cell pellet resuspended in sterile 10 % glycerol. The cells were centrifuged as before and washed again with 10 % glycerol. Cells were then resuspended in 10 % glycerol (200 μ L) and aliquots (100 μ L) were flash frozen in liquid N_2 and stored at -78 °C.

5.3.2 DNA isolation and purification

5.3.2.1 Isolation of genomic DNA from *Streptomyces*

TSB media (25 mL) was inoculated with the appropriate spore stock solution (50 μ L) and incubated with shaking at 30 °C for 2-3 days. The culture was centrifuged (4000 rpm,

20 °C, 10 min) and the supernatant discarded. The cells were washed twice with STE buffer (5 mL) then resuspended in STE buffer (2.5 mL). Lysozyme (2.5 mL, 10 mg/mL STE buffer) was added and the cells were incubated at room temperature for 30 min with gentle rocking. 10 % SDS (250 µL) was then added and the mixture was incubated for a further 5 min with gentle rocking. The mixture was then incubated at 70 °C for 5 min and subsequently left on ice for 5 min. 5 M potassium acetate (630 µL) was added and the mixture was incubated on ice for 5 min. Phenol-chloroform (2.5 mL) was added and the mixture was mixed gently by inverting for 5 min at room temperature. The contents were then centrifuged (7000 rpm, 20 °C, 10 min) and the supernatant decanted into a fresh tube. The supernatant was then washed again with phenol-chloroform (2.5 mL). 3 M sodium acetate (500 µL) and isopropanol (3 mL) were added, the mixture centrifuged (7000 rpm, 4 °C, 10 min) and the supernatant discarded. The DNA pellet was washed with 70 % EtOH, left to dry overnight and dissolved in RNase water (250 µL). The concentration of the genomic DNA was measured using a NanoDrop spectrophotometer.

5.3.2.2 Isolation of plasmid DNA from *E. coli*

LB media (5 mL) supplemented with ampicillin was inoculated with a single colony or stock solution (1 µL) of *E. coli* One Shot® TOP10 cells harbouring pET151 derivatives, and incubated with shaking at 37 °C overnight. The culture was then centrifuged (2000 rpm, 20 °C, 10 min) and the supernatant discarded. Plasmid DNA was extracted using a GeneJET Plasmid Miniprep kit according to the manufacturer's instructions, and the DNA concentration measured using a NanoDrop spectrophotometer.

5.3.3 DNA manipulation

5.3.3.1 Polymerase chain reaction (PCR)

Following the design of primers, reactions were performed using the following reagent volumes (table 5.5).

Table 5.5: Volumes of reagents used for 20 μ L PCR reactions

Reagents and quantities for use with <i>Taq</i> polymerase		Reagents and quantities for use with Phusion polymerase	
Reagent	Volume / μ L	Reagent	Volume / μ L
Template DNA*	0.8	Template DNA*	2.0
Buffer 2	2.0	5 \times Phusion HF/GC buffer	4.0
dNTP mix (10 mM)	0.4	dNTP mix (10 mM)	0.4
Forward primer (10 μ M)	0.6	Forward primer (10 μ M)	1.0
Reverse primer (10 μ M)	0.6	Reverse primer (10 μ M)	1.0
<i>Taq</i> polymerase	0.2	Phusion polymerase	0.2
DMSO	2.0	DMSO	0.6
Sterile H ₂ O	13.4	Sterile H ₂ O	10.8

*Initially diluted to 10-50 ng/ μ L (genomic DNA) or 0.2-2000 pg/ μ L (plasmid DNA)

In some cases, a higher concentration of DMSO or MgCl₂ (present within the buffer) was added. Reactions were then performed using the following PCR cycle conditions:

1. Initial denaturation: 98 $^{\circ}$ C, 30 s
 2. Denaturation: 98 $^{\circ}$ C, 10 s
 3. Annealing: $T_m - 5$ $^{\circ}$ C*, 30 s
 4. Extension: 72 $^{\circ}$ C, 1 min**
 5. Final extension: 72 $^{\circ}$ C, 10 min
 6. Resting: 4 $^{\circ}$ C, ∞
- } $\times 30$

*Annealing temperatures were approximately 5 $^{\circ}$ C below the melting temperatures of the primers (estimated using Integrated DNA Technologies Oligo Analyzer) or optimised from gradient PCR trials.

**Extension times were increased for larger regions of DNA (1 min/kb using *Taq* polymerase or 30 s/kb using Phusion polymerase).

5.3.3.2 Agarose gel electrophoresis and gel extraction

Agarose (1 g) was added to 1 × TBE buffer (100 mL) and dissolved by heating. The solution was then cooled to room temperature and GelRed (2.5 µL) was added. The contents were poured into a mould and a comb put in place. Once set, the gel was transferred to a tank filled with 1 × TBE buffer and samples containing 1 × DNA loading dye were loaded alongside an appropriate DNA size marker. The DNA fragments were then electrophoresed at 80-110 V for 0.5-2 h depending on the separation required.

Following electrophoresis, DNA fragments were visualised using a UV-transilluminator. If required, PCR products were excised from the gel using a GeneJET Gel Extraction kit according to the manufacturer's instructions, and the DNA concentration measured using a NanoDrop spectrophotometer.

5.3.3.3 Restriction digests

5.3.3.3.1 Restriction digests for analysis of recombinant vectors

To double-stranded DNA (3 µL) was added the appropriate restriction enzyme(s) (1-3 µL), buffer (2 µL) and water (up to 20 µL). The buffer was selected by supplier recommendation (single digests) or by using the Thermo Scientific DoubleDigest Calculator (double digests). Reaction mixtures were incubated at 37 °C for 15 min (for fast digest enzymes) to 2 h. Fragments were then analysed by agarose gel electrophoresis.

5.3.3.3.2 Restriction digests for subsequent ligation with pOSV556t and pKC1132

To the PCR product (48 µL) were added the appropriate restriction enzymes (1.5-3 µL), buffer (6 µL) and water, if necessary (up to 60 µL), and the mixture incubated at 37 °C for 75 min. The digested DNA was then purified using the GeneJET PCR Purification kit according to the manufacturer's instructions.

Meanwhile, to the plasmid (40 µL) were added the appropriate restriction enzymes (1.5-3 µL), buffer (5 µL) and water, if necessary (up to 50 µL), and the mixture incubated at 37 °C for 70 min. At this time, shrimp alkaline phosphatase (rSAP) (1-2 µL) was added

to dephosphorylate linearised plasmid DNA and the mixture was incubated at 37 °C for a further 20 min. The mixture was then electrophoresed (see section 5.3.3.2) and the linear plasmid DNA excised and purified using the GeneJET Gel Extraction kit according to the manufacturer's instructions.

5.3.3.4 Ligations

5.3.3.4.1 Ligations with pET151

Following purification of the PCR product, ligations were performed using the TOPO cloning kit according to the manufacturer's instructions, although higher insert:vector ratios (7:1) were typically used.

5.3.3.4.2 Ligations with pOSV556t and pKC1132

Following purification of the PCR product and either pOSV556t or pKC1132, ligations were performed using the Rapid DNA Ligation kit according to the manufacturer's instructions. Insert:vector ratios were calculated using the *in silico* Ligation Calculator.

5.3.4 DNA transfer

5.3.4.1 Transformation of chemically competent *E. coli* cells

To chemically competent cells (50 µL) was added either plasmid DNA (2 µL) or ligation mixture (3 µL), as required, under flame. Cells were incubated on ice for 30 min, transferred to a 42 °C water bath for 30 s then returned to ice for 1 min to recover. Sterile LB medium (250 µL) was added and the transformation mixture shaken horizontally at 37 °C for 1 h. Different volumes of the transformation mixture were then spread onto two pre-warmed LB agar plates supplemented with ampicillin, and the plates were incubated at 37 °C overnight.

5.3.4.2 Transformation of electrocompetent *E. coli* cells

To electrocompetent ET12567/pUZ8002 cells (100 µL) was added either recombinant pOSV556t or pKC1132 (1.5 µL), under flame. The contents were then transferred into an

ice-cold 0.2 cm electroporation cuvette and shocked at 1.8 kV for approximately 4.9 ms. Sterile LB (800 μ L) was added and the transformation mixture was shaken horizontally at 37 °C for 1 h. Different volumes of the transformation mixture were then spread onto two pre-warmed LB agar plates supplemented with chloramphenicol, kanamycin and either ampicillin (pOSV556t) or apramycin (pKC1132), and the plates were incubated at 37 °C overnight.

5.3.4.3 DNA transfer from *E. coli* to *Streptomyces* via conjugation

Following the transformation of electrocompetent ET12567/pUZ8002 cells with pOSV556t or pKC1132 derivatives (see section 5.3.4.2), LB medium (5 mL) supplemented with chloramphenicol, kanamycin and either ampicillin (pOSV556t) or apramycin (pKC1132) was inoculated with a single colony and incubated with shaking at 37 °C overnight. Fresh LB medium (5 mL) supplemented with the antibiotics described above was then inoculated with overnight culture (100 μ L) and incubated with shaking at 37 °C for 4-5 h. The cells were then harvested by centrifugation (4000 rpm, 4 °C, 8 min), washed twice with LB medium (5 mL) to remove trace antibiotics and then resuspended in LB medium (500 μ L).

Meanwhile, to fresh LB medium (450 μ L) was added the appropriate *Streptomyces* spore stock solution (50 μ L). The spores were incubated at 55 °C for 15 min then cooled on ice for 2 min. Subsequently, the spores (500 μ L) were mixed with the *E. coli* cells (500 μ L) and centrifuged (6000 rpm, 2 min). Differing aliquots were then spread onto two pre-warmed SFM plates containing 10 mM MgCl₂, and incubated at 30 °C for 20 h. The plates were then overlaid with sterile water (1 mL) containing nalidixic acid (20 μ L) to selectively kill *E. coli* cells, and hygromycin or apramycin (25 μ L) to select for exconjugants containing pOSV556t or pKC1132 derivatives, respectively. The plates were then incubated for a further 4 days at 30 °C.

5.3.5 Protein overexpression, purification and analysis

5.3.5.1 IPTG-induced gene expression

To sterile LB medium (10 mL) supplemented with ampicillin was added a single BL21 Star(DE3) transformant and the culture incubated with shaking at 37 °C overnight. The following day, sterile LB medium (1 L) supplemented with ampicillin was inoculated with overnight culture (10 mL/L) and incubated with shaking at 37 °C until $OD_{600} = 0.6$. At this time, sterile isopropyl- β -thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM to induce expression of the cloned gene and the culture was incubated with shaking at 15 °C overnight.

5.3.5.2 Recombinant fusion protein purification

Cells were harvested by centrifugation (5000 rpm, 4 °C, 15 min), the supernatant decanted and the pellet resuspended in washing buffer (20 mL/L culture) containing PMSF (10 μ M). Cells were lysed with a cell disrupter and the cellular debris removed by centrifugation (11000 rpm, 4 °C, 30 min). The cell-free extract was then passed through a 0.45 μ m syringe filter (GE Healthcare) and loaded onto a 1 mL HiTrapTM HP affinity column (GE Healthcare), which was washed with washing buffer (15 mL). The protein was then eluted in a stepwise approach with elution buffer containing 50 mM (5 mL), 100 mM (5 mL), 200 mM (3 mL) and 300 mM (3 mL) imidazole. Fractions containing the fusion protein (as determined by SDS-PAGE analysis, see section 5.3.5.3) were pooled and concentrated (6000 rpm, 4 °C) using Amicon® Ultra-15 centrifugal filters with a 10 or 30 kDa molecular weight cut-off membrane (Merck Millipore) then exchanged twice into concentration buffer*. Glycerol was added to a final concentration of 25 % and the protein was aliquoted, flash frozen in liquid N₂ and stored at -78 °C.

*For purification of CpkG, all buffers were supplemented with PLP (20 μ M). For concentration and storage of CpkG, a HEPES-based buffer was used (see section 5.2.4).

5.3.5.3 SDS-PAGE gel electrophoresis

SDS-PAGE gels were prepared as shown in table 5.6.

Table 5.6: Volumes of reagents used for the preparation of 8 %, 10 % and 12 % gels

Resolving gel				Stacking gel	
Reagent	Volume (mL)			Reagent	Volume (mL)
	8 %	10 %	12 %		
H ₂ O	4.6	3.8	3.3	H ₂ O	1.40
30 % acrylamide mix	2.7	3.4	4.0	30 % acrylamide mix	0.33
1.5 M Tris-HCl (pH 8.8)	2.5	2.5	2.5	1.0 M Tris-HCl (pH 6.8)	0.25
10 % SDS	0.1	0.1	0.1	10 % SDS	0.02
10 % ammonium persulfate	0.1	0.1	0.1	10 % ammonium persulfate	0.02
TEMED	0.006	0.005	0.004	TEMED	0.002

Protein samples containing 1× SDS-PAGE loading dye (10 µL) were then loaded onto the gel alongside a protein molecular weight marker (5 µL) and electrophoresed at 200 V for 35-45 min. Gels were subsequently stained with InstantBlue and gently rocked for 15 min to effect visualisation.

5.3.5.4 Determining protein concentration

Protein absorbance at 280 nm was measured using a NanoDrop spectrophotometer. Molecular weights were obtained from NCBI and extinction coefficients from the ExPASy ProtParam tool. Concentrations were then calculated using the equation:

$$c = \frac{A}{\epsilon l}$$

5.3.5.5 Determining protein molecular weight

Proteins were diluted to 10 µM in water and analysed by high-resolution LC-MS according to the elution profile in table 5.7. The mass spectrometer was operated in positive ion mode with a scan range of 50-3000 *m/z*.

Table 5.7: Elution conditions used for LC-MS analyses of purified recombinant proteins

Time / min	H₂O / % (0.1 % FA)	MeOH / % (0.1 % FA)	Flow rate / mL
0.0	90	10	0.1
8.0	90	10	0.1
47.5	0	100	0.1
52.5	0	100	0.1
57.5	90	10	0.1

5.3.5.6 CD spectroscopic analysis of proteins

Proteins were diluted to 0.1 mg/mL in 5 mM phosphate buffer (pH 7.0) and the absorbance measured from 180-250 nm using a Jasco J-815 CD Spectrometer. Blank samples containing only buffer were used as calibrants. Instrument settings were as follows: sensitivity: standard; data pitch: 0.2 nm; bandwidth: 1.0 nm; data integration time: 1 s; scanning speed: 100 nm/min; accumulation: 4.

5.3.5.7 Tryptic digestion of proteins

The protein of interest (30 µg) was electrophoresed and the desired band excised from the gel. The protein was then digested using the Trypsin Profile IGD Kit, according to the manufacturer's instruction. The peptide fragments were subsequently analysed by LC-MS according to the elution profile in table 5.7.

5.3.5.8 Analysis of putative protein-protein interactions

5.3.5.8.1 Native-PAGE analysis

His₆-CpkC-TR (1 mg/mL) and His₆-CpkG (1 mg/mL) were incubated for 1 h at 4 °C in gel filtration buffer, in a total volume of 30 µL. The sample was diluted in native-PAGE loading buffer (7.5 µL) and electrophoresed (90 V, 4 °C, 2 h) alongside His₆-CpkC-TR and His₆-CpkG standards.

5.3.5.8.2 Pull-down assays

His₆-CpkC-TR (2 mg/mL, 800 µL) was incubated with TEV protease (lab stock, 40 µL) for 3 h at 37 °C. The mixture was subsequently loaded onto a 1 mL HiTrapTM HP affinity column (GE Healthcare) and washing buffer (10 mL) and elution buffer (300 mM imidazole, 5 mL) were sequentially applied. Following confirmation of cleaved CpkC-TR in the wash fraction by SDS-PAGE analysis, the protein was concentrated (6000 rpm, 4 °C) to 1 mg/mL and incubated with CpkG (1 mg/mL) for 1 h at 4 °C. The mixture was passed through a second 1 mL HiTrapTM HP affinity column and washing buffer and elution buffer were applied as previously described. CpkC-TR and His₆-CpkG were detected in the wash and elution fractions, respectively.

5.3.6 Attempted detection of putative polyketide alkaloid metabolites

5.3.6.1 Extraction of metabolites from agar

ISP2, R3 and R5 agar were inoculated with *Streptomyces cyaneofuscatus* spore stock solutions (10 µL). The cultures were then incubated at 30 °C for 7 days. The agar was then cut with a glass vial and to ~5 cm² portions were added EtOAc (3 mL) and 0.1 M HCl (1 mL). The contents were shaken for ~2 min to allow phase separation and 500 µL of the organic layer was transferred to a vial and left to evaporate.

5.3.6.2 Extraction of metabolites from liquid media

TSB media (10 mL) was inoculated with *Streptomyces cyaneofuscatus* spore stock solutions (10 µL). The cultures were then incubated at 30 °C with shaking for 2 days. ISP2, R5, SMM and AM1 production media (50 mL) were then inoculated with preculture (25 µL), and incubated at 30 °C with shaking for 7 days. The cells were centrifuged (5000 rpm, 20 °C, 15 min) and to the supernatant (500 µL) was added EtOAc (500 µL). The layers were then mixed and centrifuged (3000 rpm, 20 °C, 3 min) to allow phase separation and 300 µL of the organic layer was transferred to a vial and left to evaporate.

5.3.6.3 Sample preparation for LC-MS

The metabolite residues were dissolved in MeOH (0.5 mL), then transferred to a PVDF micro spin filter (Grace Davison Discovery Sciences) and centrifuged (12000 rpm, 2 min). The samples were analysed by high-resolution LC-MS according to the elution profile in table 5.8 and the absorbance was monitored at 210 nm.

Table 5.8: Elution conditions used for LC-MS detection of putative polyketide alkaloid metabolites *in vivo*, and to assess enzymatic activity *in vitro*

Time / min	H ₂ O / % (0.1 % FA)	Acetonitrile / % (0.1 % FA)	Flow rate / mL/min
0.0	95	5	0.2
5.3	95	5	0.2
18.3	0	100	0.2
22.3	0	100	0.2
25.3	95	5	0.2
34.0	95	5	0.2

5.3.7 Overproduction and purification of valinomycin fragments

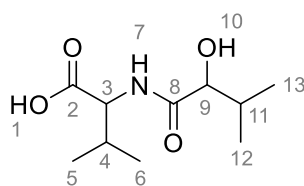
42 R3 agar plates were inoculated with *S. cyaneofuscatu*s::SARP spore stock solutions (10 µL) and incubated at 30 °C for 6 days. The agar was then cut with a scalpel and washed with 0.1 M HCl (300 mL). Metabolites were then extracted with EtOAc (3 × 300 mL), dried over MgSO₄ and concentrated under vacuum. The crude extract was redissolved in MeOH (3 mL), pre-adsorbed to dental cotton then packed into a cartridge (manufactured by University of Warwick), and purified by preparative HPLC according to the elution profile in table 5.9. Compounds **64** and **65** eluted at 23 and 28 min, respectively.

Table 5.9: Elution conditions used for HPLC purification detection **64** and **65**

Time / min	H ₂ O / %	Acetonitrile / %	Flow rate / mL/min
0.0	95	5	9.0
5.0	95	5	9.0
50.0	0	100	9.0
55.0	0	100	9.0
56.0	95	5	9.0
60.0	95	5	9.0

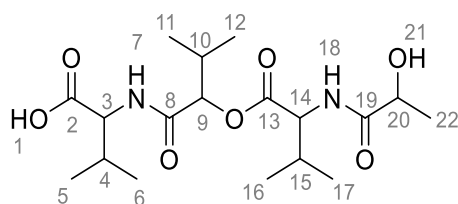
5.3.7.1 Characterisation data

5.3.7.1.1 (2-Hydroxy-3-methylbutanoyl)valine (**64**)



δ_{H} ((CD₃)₂CO, 500 MHz): 0.78 (d, 3H, H-12, $J = 7.0$ Hz), 0.86 (d, 3H, H-5/6, $J = 7.0$ Hz), 0.88 (d, 3H, H-5/6, $J = 7.0$ Hz), 0.91 (d, 3H, H-13, $J = 7.0$ Hz), 1.95-2.03 (m, 1H, H-11), 2.05-2.14 (m, 1H, H-4), 3.74 (t, 1H, H-9, $J = 4.5$ Hz), 4.21 (dd, 1H, H-3, $J = 9.0$ Hz, 5.0 Hz), 5.56, (d, 1H, H-10, $J = 5.5$ Hz), 7.48 (d, 1H, H-7, $J = 8.5$ Hz), 12.65 (br. s, 1H, H-1); δ_{C} ((CD₃)₂CO, 125 MHz): 16.1 (C-12), 17.7 (C-5), 19.1 (C-6), 19.2 (C-13), 30.3 (C-4), 31.3 (C-11), 56.3 (C-3), 75.1 (C-9), 172.9 (C-1), 173.1 (C-8); HRMS: Calculated for [M+Na]⁺ 240.1206, found 240.1202.

5.3.7.1.2 (2-(((2-Hydroxypropanoyl)valyl)oxy)-3-methylbutanoyl)valine (**65**)



δ_{H} ($(\text{CD}_3)_2\text{CO}$, 500 MHz): 0.82-0.93 (6 \times dd, 12H, H-5, H-6, H-11, H-12, H-16, H-17), 1.23 (d, 3H, H-22, $J = 6.5$ Hz), 2.02-2.18 (m, 3H, H-4, H-10, H-15), 4.04 (q, 1H, H-20, $J = 6.5$ Hz), 4.12 (dd, 1H, H-3, $J = 8.0$ Hz, 6.0 Hz), 4.29 (dd, 1H, H-14, $J = 8.5$ Hz, 6.0 Hz), 4.81 (d, 1H, H-9, $J = 5.0$ Hz), 5.63 (br. s, 1H, H-21), 7.62 (d, 1H, H-18, $J = 8.5$ Hz), 8.08 (d, 1H, H-7, $J = 8.5$ Hz); $(\text{CD}_3)_2\text{CO}$, 125 MHz): 17.1, 17.8, 18.1, 18.7, 18.9, 19.1, 19.2 (C-5, C-6, C-11, C-12, C-16, C-17), 21.2 (C-22), 29.8 (C-4), 30.1 (C-10), 30.4 (C-15), 56.8 (C-14), 57.2 (C-3), 67.2 (C-20), 77.9 (C-9), 168.6 (C-8), 170.7 (C-13), 172.7 (C-2), 174.6 (C-19); Calculated for $[\text{M}+\text{Na}]^+$ 411.2102, found 411.2103.

5.4 Biochemical assays

5.4.1 *In vitro* investigations with CpkG

5.4.1.1 UV-Vis spectroscopic analysis of PLP binding to CpkG

UV-Vis absorbance spectra were obtained for His₆-CpkG (60 μM) and PLP (60 μM) independently. Prior to analysis, His₆-CpkG was exchanged from HEPES buffer (50 mM HEPES (pH 7.6), 100 mM NaCl, 20 μM PLP) into Tris buffer (20 mM Tris-HCl, (pH 8.0), 100 mM NaCl) to remove unbound PLP. Absorbance measurements were recorded from 300-600 nm and performed in triplicate.

5.4.1.2 Spectrophotometric analysis of the substrate specificity of CpkG

The enzymatic activity and substrate specificity of CpkG were characterised by employing a spectrophotometric assay that relies on the detection of a blue α -amino acid-copper (II) complex. 20 μM His₆-CpkG was combined in separate reactions with 10 mM amino donor (benzylamine **45**, cyclohexylamine **46**, isobutylamine **47**, butylamine **48** or octylamine **49**) and 10 mM amino acceptor (pyruvate **25** or α -ketoglutarate **50**) in 100 mM potassium phosphate buffer (pH 7.0) in a total volume of 160 μL . Reactions were carried out in 96-well plates and incubated overnight at room temperature. Following overnight incubation, 40 μL of a $\text{CuSO}_4/\text{MeOH}$ solution (prepared as described by Hwang and Kim¹³⁰) was added to each reaction and the resulting mixtures were filtered by centrifugation (12000 rpm, 2 min) using PVDF micro spin filters (Grace Davison Discovery Sciences). Supernatants were subsequently analysed by measuring the

absorbance at 595 nm and reactions were performed in triplicate. Reaction mixtures incubated without enzyme served as negative controls. Blank samples consisted of 100 mM potassium phosphate buffer (pH 7.0) with and without 20 μM His₆-CpkG. The results were expressed as relative specificities calculated by first subtracting the mean absorbance value of the blank samples from each measured absorbance value, to obtain background-corrected values. Subsequently, the mean absorbance values of the test samples were divided by the mean absorbance value of the control samples. Relative specificities were then determined by comparison to the specificity of CpkG for octylamine and pyruvate as amino donor and acceptor.

5.4.1.3 Determination of the absolute stereochemistry of alanine

The absolute stereochemistry of alanine **51**, formed by the CpkG-catalysed transamination of pyruvate **25** with octylamine **49**, was determined by derivatisation with Marfey's reagent. His₆-CpkG (20 μM) was combined with octylamine (10 mM), pyruvate (10 mM) and phosphate buffer (100 mM, pH 7.0) in a final volume of 80 μL . After incubation for 2 h at room temperature, proteins were removed by centrifugation (12000 rpm, 15 min) using a Vivaspin 500 centrifugal filter with a 5 kDa MWCO (Sartorius Stedim Biotech). Acetone (65 μL), NaHCO₃ (1 M, 7.5 μL) and 1 % Marfey's reagent in acetone (17 μL) were then added to 40 μL of the filtrate and the resulting mixture was incubated at 37 °C for 1.5 h. Authentic standards of the Marfey's derivatives of L- and D-alanine were prepared similarly by adding acetone (65 μL), 1 M NaHCO₃ (7.5 μL) and 1 % Marfey's reagent in acetone (17 μL) to solutions of L- and D-alanine (10 mM) in phosphate buffer (100 mM, pH 7.0) in a final volume of 40 μL . These mixtures were incubated as described above. Reactions were quenched with 1 M HCl (7.5 μL) and the remaining acetone was removed by evaporation. Deionised water (200 μL) and acetonitrile (200 μL) were then added to each sample. Precipitates were removed by centrifugation (12000 rpm, 1 min) using spin filters. The samples were subsequently analysed by LC-MS according to the elution profile in table 5.8 and the absorbance was monitored at 210 nm. Independent experiments were performed in duplicate.

5.4.1.4 Detection of aldehyde

Production of octanal **57** in the CpkG-catalysed reaction with pyruvate **25** and octylamine **49** was detected by derivatisation with D-cysteine and subsequent high-resolution LC-MS analysis. His₆-CpkG (90 μM) was combined with octylamine (200 μM), pyruvate (200 μM) and phosphate buffer (25 mM, pH 7.0) in a final volume of 125 μL. After incubation for 15 min at 37 °C, D-cysteine (400 μM) was added and the resulting mixture incubated for 10 min at 50 °C. MeOH (125 μL) was then added to precipitate the enzyme. Precipitates were removed by centrifugation (12000 rpm, 1 min) using PVDF micro spin filters (Grace Davison Discovery Sciences) and the supernatants were analysed by LC-MS according to the elution profile in table 5.8. An authentic octanal standard (200 μM) was derivatised and analysed similarly. Independent assays were performed in duplicate.

Production of 5-hydroxydodecanal **59** was detected similarly, from the CpkG-catalysed reaction with pyruvate and 1-aminododecan-5-ol **60**.

5.4.1.5 Detection of pyruvate and amine

Production of pyruvate **25** in the CpkG-catalysed reaction with alanine **51** and octanal **57** was analysed using a lactate dehydrogenase (LDH)-coupled assay, while production of octylamine **49** was analysed by LC-MS. Reactions contained His₆-CpkG (90 μM), octanal (200 μM), L- or D-alanine (200 μM), lactate dehydrogenase (from *Lactobacillus leichmanii*, 1 μM), NADH (200 μM) and phosphate buffer (25 mM, pH 7.0) in a total volume of 250 μL. The absorbance at 340 nm was monitored for 15 min at 25 °C, at which time MeOH (450 μL) was added to precipitate the enzymes. Precipitates were removed by centrifugation (12000 rpm, 1 min) using PVDF micro spin filters (Grace Davison Discovery Sciences) and the supernatants were analysed by LC-MS according to the elution profile in table 5.8. An authentic octylamine standard (200 μM) was analysed similarly. Independent assays were performed in duplicate.

Production of 1-aminododecan-5-ol **60** was detected similarly, from the CpkG-catalysed reaction with alanine and 5-hydroxydodecanal **59**.

5.4.2 *In vitro* investigations with CpkE and CpkI

5.4.2.1 Incubation of *trans*-octen-2-al with CpkE

To investigate the putative isomerase activity of CpkE, His₆-CpkE (100 μM) was combined with *trans*-octen-2-al **78** (200 μM) and storage buffer in a total volume of 60 μL. After incubation for 3 h at room temperature, MeOH (120 μL) was added and precipitants were removed by centrifugation (12000 rpm, 15 min). The supernatant was analysed by LC-MS according to the elution profile in table 5.8. A control reaction and an authentic standard of *cis*-octen-3-al **79** (200 μM) in storage buffer were analysed similarly.

5.4.2.2 Incubation of dodecane-1,5-diol with CpkI

To investigate the putative oxidoreductase activity of CpkI, His₆-CpkI (90 μM) was combined with dodecane-1,5-diol **76** (200 μM), NAD⁺ or NADP⁺ (200 μM), phosphate buffer (pH 7.0) and water in a final volume of 100 μL. After incubation for 1 h at room temperature, D-cysteine (400 μM) was added and the mixture incubated for 10 min at 50 °C. MeOH (200 μL) was then added and precipitants were removed by centrifugation (12000 rpm, 15 min). The supernatant was analysed by LC-MS according to the elution profile in table 5.8. A control reaction and an authentic standard of 5-hydroxydodecanal **59** (200 μM) were derivatised and analysed similarly.

5.4.3 *In vitro* investigations with CpkD, ScF and CpkH

5.4.3.1 UV-Vis spectroscopic analysis of flavin binding

His₆-CpkD, His₆-ScF and His₆-CpkH (without added flavin) were diluted to 50 μM with concentration buffer. Absorbance spectra were recorded from 300-600 nm and values obtained for a blank (concentration buffer) were subtracted. Measurements were performed in triplicate and absorbance maxima were compared to literature values for FAD (450 nm) and FMN (446 nm).^{176,177}

5.4.3.2 UV-Vis spectroscopic analysis of NAD(P)H oxidation by flavoprotein

His₆-CpkD, His₆-ScF and His₆-CpkH (10 nmol) were separately incubated with either NADH or NADPH (25 nmol) and concentration buffer in a total volume of 140 μ L. The absorbance at 340 nm was monitored for 25 min at 25 °C and measurements were performed in duplicate.

5.4.3.3 CpkH-catalysed oxidation of 1-aminododecan-5-ol

To investigate putative flavin-dependent dehydrogenase activity, racemic 1-aminododecan-5-ol **60** (200 μ M) was incubated with either His₆-CpkD, His₆-ScF or His₆-CpkH (20 μ M) in a total volume of 62.5 μ L. After incubation for 3 h at 30 °C, MeOH (125 μ L) was added to precipitate the proteins and the mixtures were centrifuged (12000 rpm, 1 min) using PVDF micro spin filters (Grace Davison Discovery Sciences). The supernatants were analysed alongside a synthetic authentic standard, 6-heptyl-2,3,4,5-tetrahydropyridine **81** (200 μ M), by LC-MS according to the elution profile in table 5.8. Independent assays were performed in duplicate. To investigate the enantioselectivity of CpkH, (*R*)- and (*S*)-**60** were incubated separately with His₆-CpkH as described above. Independent assays were performed in duplicate.

5.4.3.4 Incubation of (*R,E*)- or (*S,E*)-1-aminododec-2-en-5-ol with coelimycin biosynthetic enzymes

To investigate the configurational selectivity of CpkH, (*R,E*)- or (*S,E*)-1-aminododec-2-en-5-ol **84** (200 μ M) were incubated with His₆-CpkH (20 μ M) in a total volume of 62.5 μ L. To investigate the potential epimerase or oxidoreductase activities of CpkE, CpkD, ScF and CpkI, the His₆-fusion proteins (20 μ M) were separately added to the above assays. After incubation for 3 h at 30 °C, MeOH (125 μ L) was added to precipitate the proteins and the mixtures were centrifuged (12000 rpm, 15 min). The supernatants were analysed by LC-MS according to the elution profile in table 5.8. Independent assays were performed in duplicate.

5.4.3.5 CpkD-catalysed epoxidation of ethyl (2E,4E,6E)-octa-2,4,6-trienoate

To investigate putative flavin-dependent monooxygenase activity, ethyl(2E,4E,6E)-octa-2,4,6-trienoate **99** (200 μM) was incubated with FAD or FMN (40 μM), NADH (40 μM), NAC (400 μM) and either His₆-CpkD or His₆-ScF (20 μM) in a total volume of 62.5 μL . After incubation for 3 h at room temperature, MeOH (125 μL) was added to precipitate proteins and the mixtures were centrifuged (12000 rpm, 15 min). The supernatants were then analysed by LC-MS according to the elution profile in table 5.8. MS/MS was performed to obtain fragmentation data for the ions at 324 and 340 m/z present within the His₆-CpkD sample (Dr. Lijiang Song, University of Warwick). Independent assays were performed in triplicate.

Performed in collaboration with Dr. Lijiang Song, University of Warwick

For analysis of ¹⁸O₂ incorporation, ¹⁸O₂ (2.64 atm, 97 %) was bubbled separately through a solution of substrate, NADH and NAC for 5 min, and a solution of His₆-CpkD and FMN for 5 min. The two solutions were then combined and ¹⁸O₂ was bubbled through for a further 5 min. The tube was then sealed and the mixture incubated and analysed as described above. Independent assays were performed in duplicate.

5.5 Crystallisation of CpkG

Performed in collaboration with Dr. Lona Alkhalaf, University of Warwick

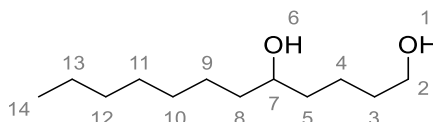
CpkG was overproduced in *E. coli* as an N-terminal hexahistidine fusion protein and purified to homogeneity by nickel affinity chromatography (see section 5.3.1). It was then concentrated to 10 mg / mL for sitting drop crystallisation trials, which were conducted with 100 nL aliquots, a Honeybee 963 robot and three 96-well commercially-available screens (JCSG-plus, Structure and PACT premier).

For co-crystallisation with substrates, conditions were optimised (0.2 M sodium malonate dibasic monohydrate, 0.15 M bis-tris propane, 20 % w/v PEG, pH 6.5) and the hanging-drop method was used. To obtain CpkG_ala crystals, the enzyme was mixed with L-alanine **51** at a final concentration of 2 mM. To obtain CpkG_ala_lactol crystals, saturating conditions of 5-hydroxydodecanal **59** were used.

5.6 Chemical procedures

5.6.1 Synthesis of putative CpkI substrate and CpkE standard

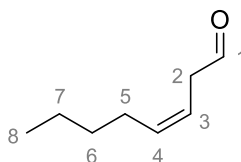
5.6.1.1 Dodecane-1,5-diol (**76**)



To δ -dodecalactone **61** (100 μ L, 475 μ mol, 1.0 eq.) in dry THF (2.5 mL) at 0 $^{\circ}$ C under argon was slowly added LiAlH₄ (1 M, 713 μ L, 713 μ mol, 1.5 eq.). The reaction was stirred for 40 min at room temperature, then cooled to 0 $^{\circ}$ C and quenched with 2 M HCl. The product was extracted with EtOAc (3 \times 15 mL), then washed with H₂O (20 mL), sat. aq. NaHCO₃ (20 mL) and brine (20 mL). The organics were dried over anhydrous MgSO₄ and concentrated under reduced pressure to afford **76** as a white solid (64.1 mg, 67 %).

m.p.: 42-45 $^{\circ}$ C; δ_{H} (CDCl₃, 400 MHz): 0.86 (t, 3H, H-14, J = 6.5 Hz), 1.19-1.34 (m, 10H, H-9-13), 1.35-1.66 (m, 8H, H-3-5, H-8), 2.37 (br. s, 2H, H-1, H-6), 3.54-3.58 (m, 1H, H-7), 3.62 (t, 2H, H-2, J = 6.0 Hz); δ_{C} (CDCl₃, 100 MHz): 14.2 (C-14), 21.9 (C-4), 22.8 (C-13), 25.8 (C-9), 29.4 (C-11), 29.8 (C-10), 32.0 (C-12), 32.6 (C-3), 37.0 (C-5), 37.7 (C-8), 62.6 (C-2), 71.9 (C-7); HRMS: Calculated for [M+Na]⁺ 225.1825, found 225.1827. Data are consistent with those reported by Wiesner, Soai and their respective coworkers.^{214,215}

5.6.1.2 *cis*-Octen-3-al (**79**)



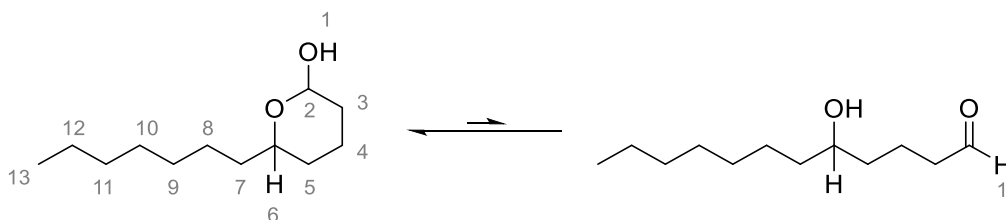
To *cis*-octen-3-ol **80** (119 μ L, 792 μ mol, 1.0 eq.) in dry DCM (7 mL) at 0 $^{\circ}$ C under argon was added DMP (440 mg, 1.04 mmol, 1.3 eq.) and the reaction stirred for 1 h. The reaction mixture was then diluted with pentane (20 mL) and washed with H₂O (6 \times 10 mL). The

organics were dried over anhydrous MgSO_4 and concentrated under reduced pressure to afford **79** as a colourless oil (96.3 mg, 96 %), which was used immediately and without purification due to rapid isomerisation.

δ_{H} (CDCl_3 , 400 MHz): 0.85 (t, 3H, H-8, $J = 7.0$ Hz), 1.23-1.35 (m, 4H, H-6, H-7), 1.96-2.09 (m, 2H, H-5), 3.14 (d, 2H, H-2, $J = 7.0$ Hz), 5.45-5.52 (m, 1H, H-3), 5.62-5.69 (m, 1H, H-4), 9.61 (br. t, 1H, H-1); HRMS: Derivatised with tris for detection; calculated for $[[\text{M}+\text{tris}]+\text{H}]^+$ 230.1751, found 230.1741.

5.6.2 Synthesis of racemic CpkG and CpkH substrates and standards

5.6.2.1 5-Hydroxydodecanal (**59**)

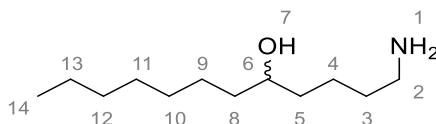


To δ -dodecalactone **61** (210 μL , 1.00 mmol, 1.0 eq.) in dry DCM (2.5 mL) at -78 $^{\circ}\text{C}$ under argon was slowly added DIBAL-H (1 M, 1.00 mL, 1.00 mmol, 1.0 eq.). The reaction was stirred at room temperature for 1 h, at which time DCM (10 mL) and potassium sodium tartrate (10 mL) were added. The resulting suspension was stirred for 1 h, the organic phase separated and the aqueous phase extracted with DCM (2×20 mL). The combined organics were washed with brine (20 mL), dried over anhydrous MgSO_4 and concentrated under reduced pressure. The crude material was purified on silica gel (1 : 3 EtOAc : pet. ether) to afford **59** (132 mg, 66 %) as a colourless oil.

Note: compound exists as mixture of hemiacetal diastereomers (99 %, A:B \approx 0.62:0.38) and open-chain aldehyde (1 %); δ_{H} (CDCl_3 , 400 MHz): 0.87 (t, 3H, H-13, $J = 6.5$ Hz), 1.12-1.71 (m, 14.38H, H-3/4b, H-5, H-7-12), 1.81-1.87 (m, 1.62H, H-3a, H-3/4b, H-4a), 2.81-2.92 (m, 0.38H, H-1b), 3.36-3.55 (m, 1.24H, H-1a, H-6a), 3.91 (m, 0.38H, H-6b), 4.68 (t, 0.62H, H-2a, $J = 7.0$ Hz), 5.29 (s, 0.38H, H-2b), 9.77 (s, 0.01H, H-1' (open chain aldehyde)); δ_{C} (CDCl_3 , 100 MHz): 14.2 (C-13), 17.6 (C-4b), 22.3 (C-4a), 22.8 (C-12), 25.5 (C-8a), 25.6 (C-8b), 29.4 (C-10a), 29.4, (C-10b), 29.8 (C-9a), 29.8 (C-9b), 30.0

(C-5b), 30.5 (C-5a), 31.3 (C-3b), 32.0 (C-11), 33.0 (C-3a), 36.2 (C-7b), 36.3 (C-7a), 68.9 (C-6b), 76.7 (C-6a), 92.0 (C-2b), 96.6 (C-2a); HRMS: Calculated for $[M+Na]^+$ 223.1669, found 223.1667. Data are consistent with those reported by Li and coworkers.²¹⁶

5.6.2.2 1-Aminododecan-5-ol (60)



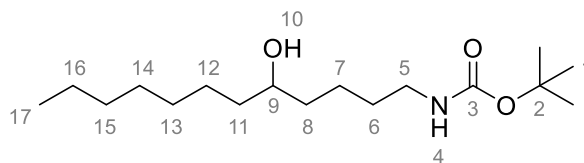
To 5-hydroxydodecanal **59** (39.5 mg, 197 μ mol, 1.0 eq.) in MeOH (1.5 mL) was added benzylamine (28.1 μ L, 257 μ mol, 1.3 eq.) and the solution stirred for 4 h at room temperature. 10 % Pd/C (84.0 mg) was added and the flask purged with $H_2 \times 5$ (1 atm) before being left to stir for a further 12 h. The reaction mixture was filtered through celite and concentrated under reduced pressure to afford **60** as a white solid in quantitative yield. 5 mg was purified by analytical HPLC (table 5.10, elution time = 8.8 min) for use in enzymatic assays.

m.p.: 82-85 °C; δ_H (CDCl₃, 500 MHz): 0.88 (t, 3H, H-14, $J = 7.0$ Hz), 1.20-1.33 (m, 10H, H-9-13), 1.35-1.80 (m, 10H, H-1, H-3-5, H-8), 2.84-2.99 (m, 2H, H-2), 3.49-3.60 (m, 1H, H-6); δ_C (CDCl₃, 125 MHz): 14.3 (C-14), 22.4 (C-4), 22.8 (C-13), 26.1 (C-9), 27.6 (C-3), 29.5 (C-11), 29.9 (C-10), 32.0 (C-12), 36.4 (C-5), 38.0 (C-8), 39.3 (C-2), 71.1 (C-6); HRMS: Calculated for $[M+H]^+$ 202.2165, found 202.2169.

Table 5.10: Elution conditions used for HPLC purification of 1-aminododecan-5-ol **60**

Time / min	H ₂ O / % (0.1 % FA)	MeOH / % (0.1 % FA)	Flow rate / mL/min
0.0	40	60	5.0
5.0	40	60	5.0
25.0	0	100	5.0
28.0	0	100	5.0
30.0	40	60	5.0

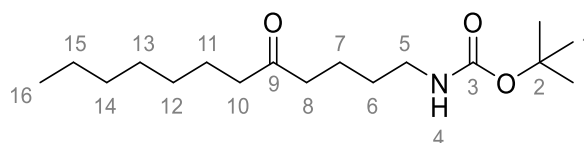
5.6.2.3 *tert*-Butyl (5-hydroxydodecyl)carbamate (**82**)



To 1-aminododecan-5-ol **60** (45.0 mg, 223 μmol , 1.0 eq.) in anhydrous DCM (1 mL) under argon was added di-*tert*-butyl dicarbonate (72.8 mg, 334 μmol , 1.5 eq.) and the solution stirred for 2 h at room temperature. H_2O (10 mL) was added and the product extracted with DCM (3×10 mL). The combined organic layers were washed sequentially with sat. aq. NaHCO_3 (15 mL) and brine (15 mL), dried over anhydrous MgSO_4 and concentrated under reduced pressure to afford **82** as a pale yellow oil, which was used without purification in the next step.

δ_{H} (CDCl_3 , 400 MHz): 0.81 (t, 3H, H-17, $J = 6.5$ Hz), 1.15-1.30 (m, 10H, H-12–16), 1.31-1.41 (m, 8H, H-6–8, H-11), 1.45 (s, 9H, H-1), 3.06 (m, 2H, H-5), 3.50 (m, 1H, H-9), 4.56 (br. s, 1H, H-4); δ_{C} (CDCl_3 , 100 MHz): 14.1 (C-17), 22.7 (C-16), 22.8 (C-7), 25.7 (C-12), 27.5 (C-1), 29.4 (C-14), 29.7 (C-13), 30.2 (C-6), 31.9 (C-15), 37.0 (C-8), 37.6 (C-11), 40.4 (C-5), 71.7 (C-9), 85.2 (C-2), 146.8 (C-3); HRMS: Calculated for $[\text{M}+\text{Na}]^+$ 324.2509, found 324.2510.

5.6.2.4 *tert*-Butyl (5-oxododecyl)carbamate (**83**)

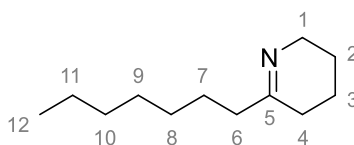


To *tert*-butyl (5-hydroxydodecyl)carbamate **82** (63.7 mg, 211 μmol , 1.0 eq.) in anhydrous DCM (1 mL) under argon was added Dess-Martin periodinane (DMP) (108 mg, 254 μmol , 1.2 eq.) and the mixture stirred for 2.5 h at room temperature. 20 % $\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O}$ (3 mL) and sat. aq. NaHCO_3 (3 mL) were added sequentially and the mixture stirred for a further 10 min. The product was extracted with EtOAc (3×10 mL) and the combined organics washed with brine (15 mL) and dried over anhydrous MgSO_4 .

The solvent was removed under reduced pressure to afford **83** as a colourless oil, which was used without purification in the next step.

δ_{H} (CDCl_3 , 400 MHz): 0.80 (t, 3H, H-16, $J = 6.5$ Hz), 1.13-1.28 (m, 6H, H-13–15), 1.34-1.41 (m, 4H, H-6, H-12), 1.42-1.56 (m, 13H, H-1, H-7, H-11), 2.31 (t, 2H, H-10, $J = 7.5$ Hz), 2.35 (t, 2H, H-8, $J = 7.0$ Hz), 3.04 (q, 2H, H-5, $J = 6.5$ Hz), 4.54 (br. s, 1H, H-4); δ_{C} (CDCl_3 , 100 MHz): 14.1 (C-16), 20.8 (C-7), 22.7 (C-15), 24.0 (C-11), 27.5 (C-1), 29.2 (C-13), 29.3 (C-12), 29.6 (C-6), 31.8 (C-14), 40.2 (C-5), 42.2 (C-8), 43.0 (C-10), 85.2 (C-2), 146.8 (C-3), 210.9 (C-9); HRMS: Calculated for $[\text{M}+\text{Na}]^+$ 322.2353, found 322.2355.

5.6.2.5 6-Heptyl-2,3,4,5-tetrahydropyridine (**81**)

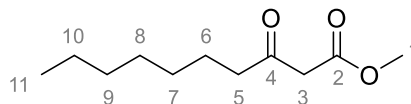


To *tert*-butyl (5-oxododecyl)carbamate **83** (67.0 mg, 222 μmol , 1.0 eq.) in DCM (1 mL) at -10 $^{\circ}\text{C}$ was added TFA (0.30 mL) over 5 min. The solution was allowed to warm to room temperature and stirred for 5 h. The solution was basified with 2 M NaOH (pH > 12), diluted with H_2O (5 mL) and extracted with DCM (3×10 mL). The combined organics were washed with sat. aq. NaHCO_3 (15 mL), H_2O (15 mL) and brine (15 mL), dried over anhydrous MgSO_4 and concentrated under reduced pressure. The crude oil was purified by silica gel chromatography (1 : 19 MeOH : DCM) to afford **81** as an orange semi-solid (9.50 mg, 27 % over 4 steps).

δ_{H} (CDCl_3 , 500 MHz): 0.87 (t, 3H, H-12, $J = 7.0$ Hz), 1.25-1.42 (m, 8H, H-8–11), 1.65 (quint, 2H, H-7, $J = 7.0$ Hz), 1.81-1.84 (m, 4H, H-2, H-3), 2.52-2.58 (br. s, 2H, H-4), 2.71 (t, 2H, H-6, $J = 7.0$ Hz), 3.67-3.73 (br. s, 2H, H-1); δ_{C} (CDCl_3 , 125 MHz): 14.2 (C-12), 17.8 (C-2), 20.2 (C-3), 22.7 (C-11), 26.9 (C-7), 29.3 (C-9), 29.4 (C-4), 29.9 (C-8), 31.7 (C-10), 38.4 (C-6), 45.4 (C-1), 167.9 (C-5); HRMS: Calculated for $[\text{M}+\text{H}]^+$ 182.1903, found 182.1904. Data are consistent with those reported by Fukuda and coworkers.²¹⁷

5.6.3 Synthesis of enantiopure CpkH substrates

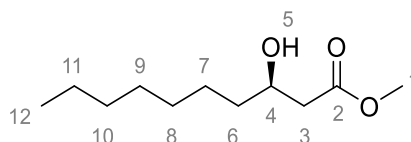
5.6.3.1 Methyl-3-oxodecanoate (**87**)



To octanoic acid **86** (1.10 mL, 6.93 mmol, 1.0 eq.) in anhydrous THF (14 mL) under argon was added CDI (1.35 g, 8.32 mmol, 1.2 eq.) and the reaction stirred at room temperature for 2 h. Monomethyl malonate potassium salt (1.62 g, 10.4 mmol, 1.5 eq.) and anhydrous MgCl_2 (792 mg, 8.32 mmol, 1.2 eq.) were subsequently added and the mixture stirred at room temperature overnight. The mixture was then quenched with 0.1 M HCl and the product extracted with EtOAc (3×75 mL). The combined organic layers were washed sequentially with 5 % NaHCO_3 (100 mL), H_2O (100 mL) and brine (100 mL) and dried over anhydrous MgSO_4 . The solvent was removed under reduced pressure to afford **87** as a yellow oil (1.41 g, 100 %).

δ_{H} (CDCl_3 , 300 MHz): 0.81 (t, 3H, H-11, $J = 6.5$ Hz), 1.12-1.35 (m, 8H, H-7-10), 1.54 (quint, 2H, H-6, $J = 7.0$ Hz), 2.49 (t, 2H, H-5, $J = 7.5$ Hz), 3.40 (s, 2H, H-3), 3.68 (s, 3H, H-1); δ_{C} (CDCl_3 , 75 MHz): 14.0 (C-11), 22.5 (C-10), 23.4 (C-6), 28.9 (C-8), 28.9 (C-7), 31.6 (C-9), 43.0 (C-5), 48.9 (C-3), 52.2 (C-1), 167.6 (C-2), 202.8 (C-4); HRMS: Calculated for $[\text{M}+\text{Na}]^+$ 223.1308, found 225.1305. Data are consistent with those reported by Sieben and coworkers.²¹⁸

5.6.3.2 Methyl (*R*)-3-hydroxydecanoate (**88**)

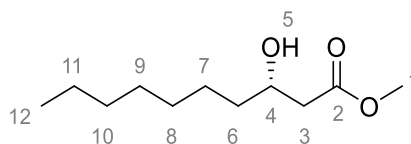


To $[\text{Ru}(\text{OAc})_2(\text{R})\text{-BINAP}]$ (200 mg, 238 μmol , 1.0 eq.) in dry MeOH (10 mL) was added 1 M HCl in MeOH (238 μL , 238 μmol , 1.0 eq.). The suspension was stirred at ambient temperature for 30 min, at which time methyl-3-oxodecanoate **87** (4.98 g, 24.9 mmol, 105 eq.) was added. The flask was purged with H_2 (49 atm) three times over

5 min and stirred under H₂ at 50 °C overnight. The reaction mixture was cooled, concentrated under reduced pressure and the resulting residue purified by silica gel chromatography (4 : 1 EtOAc : hexane) to afford (*R*)-**88** as a colourless oil (4.03 g, 80 %).

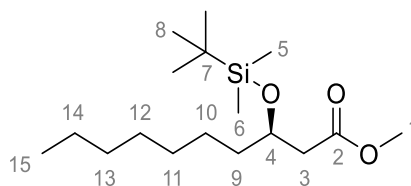
δ_{H} (CDCl₃, 300 MHz): 0.85 (t, 3H, H-12, $J = 6.5$ Hz), 1.16-1.57 (m, 12H, H-6–11), 2.39 (dd, 1H, H-3a, $J = 16.5$ Hz, 9.0 Hz), 2.50 (dd, 1H, H-3b, $J = 16.5$ Hz, 3.5 Hz), 2.71 (br. s, 1H, H-5), 3.69 (s, 3H, H-1), 3.98 (m, 1H, H-4); δ_{C} (CDCl₃, 75 MHz): 14.2 (C-12), 22.7 (C-11), 25.6 (C-7), 29.3 (C-9), 29.5 (C-8), 31.9 (C-10), 36.6 (C-6), 41.2 (C-3), 51.8 (C-1), 68.1 (C-4), 173.6 (C-2); HRMS: Calculated for [M+Na]⁺ 225.1460, found 225.1461. Data are consistent with those reported by Wu and coworkers.²¹⁹

5.6.3.3 Methyl (*S*)-3-hydroxydecanoate (**88**)



Compound (*S*)-**88** was synthesised following the procedure described for (*R*)-**88**, using [Ru[(OAc)₂((*S*)-BINAP)] as the catalyst. Yield: 4.69 g, 93 %. NMR and LRMS data are consistent with those for (*R*)-**88**.

5.6.3.4 Methyl (*R*)-3-((*tert*-butyldimethylsilyloxy)decanoate (**89**)

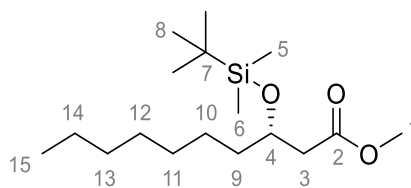


To a solution of methyl-(*R*)-3-hydroxydecanoate **88** (3.86 g, 19.1 mmol, 1.0 eq.) in anhydrous DMF (30 mL) under argon were added imidazole (2.60 g, 38.2 mmol, 2.0 eq.) and *tert*-butyl-dimethylsilyl chloride (4.31 g, 28.6 mmol, 1.5 eq.) sequentially at 0 °C. The mixture was stirred at room temperature for 20 h then quenched with H₂O (30 mL) and stirred for a further 30 min. The product was extracted with Et₂O (3 × 30 mL) and the combined organic layers were washed with brine, dried over anhydrous MgSO₄ and

concentrated under reduced pressure. The residue was purified by silica gel chromatography (1 : 99 Et₂O : pet. ether) to afford (*R*)-**89** as a colourless oil (5.58 g, 92 %).

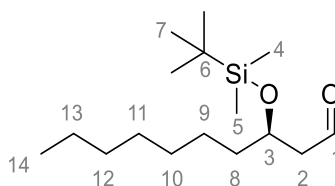
δ_{H} (CDCl₃, 300 MHz): 0.02 (s, 3H, H-5/6), 0.05 (s, 3H, H-5/6), 0.84-0.91 (m, 12H, H-8, H-15), 1.19-1.35 (m, 10H, H-10–14), 1.42-1.53 (m, 2H, H-9), 2.41 (s, 1H, H-3a), 2.44 (d, 1H, H-3b, $J = 1.5$ Hz), 3.65 (s, 3H, H-1), 4.11 (quint, 1H, H-4, $J = 6.0$ Hz); δ_{C} (CDCl₃, 75 MHz): -4.7 (C-5/6), -4.4 (C-5/6), 14.2 (C-15), 18.1 (C-7), 22.8 (C-14), 25.1 (C-10), 25.9 (C-8), 29.4 (C-12), 29.8 (C-11), 31.9 (C-13), 37.8 (C-9), 42.7 (C-3), 51.6 (C-1), 69.7 (C-4), 172.5 (C-2); HRMS: Calculated for [M+Na]⁺ 339.2326, found 339.2325. Data are consistent with those reported by Kociński and coworkers.²²⁰

5.6.3.5 Methyl (*S*)-3-((*tert*-butyldimethylsilyl)oxy)decanoate (**89**)



Compound (*S*)-**89** was synthesised from (*S*)-**88**, following the procedure described for (*R*)-**89**. Yield: 6.02 g, 88 %. NMR and LRMS data are consistent with those for (*R*)-**89**.

5.6.3.6 (*R*)-3-((*tert*-Butyldimethylsilyl)oxy)decanal (**90**)

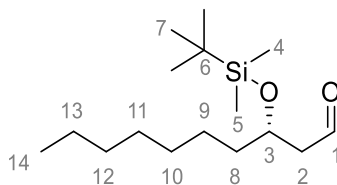


To a solution of methyl (*R*)-3-((*tert*-butyldimethylsilyl)oxy)decanoate **89** (5.41 g, 17.1 mmol, 1.0 eq.) in anhydrous DCM (100 mL) at -78 °C under argon was added DIBAL-H (18.8 mL, 1 M in hexanes, 18.8 mmol, 1.1 eq.), dropwise. After 1 h, DCM (100 mL) and sat. aq. sodium potassium tartrate (200 mL) were added. The resulting suspension was stirred for 1 h at room temperature, the organics separated and the

aqueous layer extracted with DCM (2 × 100 mL). The combined organics were washed with brine (200 mL), dried over anhydrous MgSO₄ and concentrated under reduced pressure. The crude material was purified by silica gel chromatography (1 : 19 Et₂O : pet. ether) to afford (*R*)-**90** as a colourless oil (4.48 g, 92 %).

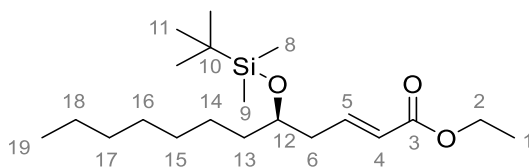
δ_{H} (CDCl₃, 300 MHz): 0.06 (s, 3H, H-4/5), 0.08 (s, 3H, H-4/5), 0.84-0.91 (m, 12H, H-7, H-14), 1.21-1.37 (m, 10H, H-9–13), 1.42-1.53 (m, 2H, H-8), 2.51 (dd, 2H, H-2, $J = 5.5$ Hz, 2.5 Hz), 4.18 (quint, 1H, H-3, $J = 5.5$ Hz), 9.82 (t, 1H, H-1, $J = 2.5$ Hz); δ_{C} (CDCl₃, 75 MHz): -4.5 (C-4/5), -4.3 (C-4/5), 14.2 (C-14), 18.2 (C-6), 22.8 (C-13), 25.3 (C-9), 26.0 (C-7), 29.4 (C-11), 29.8 (C-10), 31.9 (C-12), 38.1 (C-8), 51.0 (C-2), 68.5 (C-3), 202.3 (C-1); HRMS: Calculated for [M+Na]⁺ 309.2220, found 309.2224. Data are consistent with those reported by Kocieński and coworkers.²²⁰

5.6.3.7 (*S*)-3-((*tert*-Butyldimethylsilyl)oxy)decanal (**90**)



Compound (*S*)-**90** was synthesised from (*S*)-**89**, following the procedure described for (*R*)-**90**. Yield: 4.76 g, 90 %. NMR and LRMS data are consistent with those for (*R*)-**90**.

5.6.3.8 Ethyl (*R,E*)-5-((*tert*-butyldimethylsilyl)oxy)dodec-2-enoate (**91**)

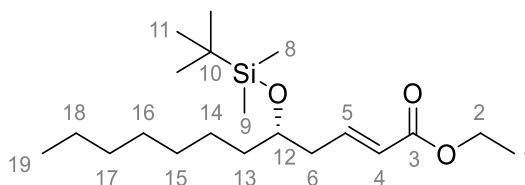


To a solution of (*R*)-3-((*tert*-butyldimethylsilyl)oxy)decanal **90** (4.42 g, 15.4 mmol, 1.0 eq.) in anhydrous THF (85 mL) under argon was added ethyl(triphenyl phosphoranylidene)acetate (8.06 g, 23.1 mmol, 1.5 eq.) and the reaction stirred at room temperature overnight. After completion, the solvent was removed under reduced

pressure and the crude material purified by silica gel chromatography (2 : 23 EtOAc : pet. ether) to afford (*R*)-**91** as a colourless oil (5.09, 93 %).

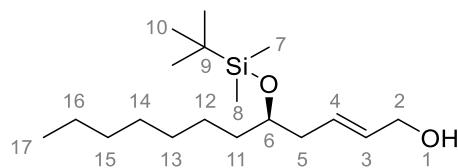
δ_{H} (CDCl₃, 300 MHz): 0.04 (s, 6H, H-8–9), 0.84-0.92 (m, 12H, H-11, H-18), 1.21-1.36 (m, 13H, H-1, H-13–17), 1.37-1.49 (m, 2H, H-12), 2.25-2.40 (m, 2H, H-6), 3.76 (quint, 1H, H-7, $J = 5.5$ Hz), 4.19 (q, 2H, H-2, $J = 7.0$ Hz), 5.83 (d, 1H, H-4, $J = 15.5$ Hz), 6.96 (dt, 1H, H-5, $J = 15.5$ Hz, 7.5 Hz); δ_{C} (CDCl₃, 75 MHz): -4.4 (C-8, C-9), 14.2 (C-18), 14.4 (C-1), 18.2 (C-10), 22.8 (C-17), 25.4 (C-13), 26.0 (C-11), 29.4 (C-15), 29.8 (C-14), 32.0 (C-16), 37.4 (C-12), 40.3 (C-6), 60.3 (C-2), 71.5 (C-7), 123.3 (C-4), 146.3 (C-5), 166.7 (C-3); HRMS: Calculated for [M+Na]⁺ 379.2641, found 379.2639. Data are consistent with those reported by Nemoto and coworkers.²²¹

5.6.3.9 Ethyl (*S,E*)-5-((*tert*-butyldimethylsilyl)oxy)dodec-2-enoate (**91**)



Compound (*S*)-**91** was synthesised from (*S*)-**90**, following the procedure described for (*R*)-**91**. Yield: 5.20 g, 91 %. NMR and LRMS data are consistent with those for (*R*)-**91**.

5.6.3.10 (*R,E*)-5-((*tert*-Butyldimethylsilyl)oxy)dodec-2-en-1-ol (**92**)

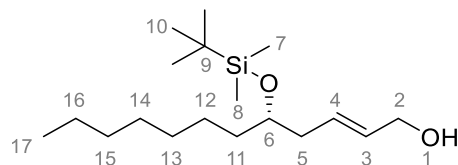


To a solution of ethyl (*R,E*)-5-((*tert*-butyldimethylsilyl)oxy)dodec-2-enoate **91** (5.01 g, 14.1 mmol, 1.0 eq.) in anhydrous DCM (98 mL) at -78 °C under argon was added DIBAL-H (35.1 mL, 1 M in hexane, 35.1 mmol, 2.5 eq.), dropwise. The reaction mixture was stirred for 3 h at -78 °C then quenched by the slow addition of MeOH. The mixture was diluted with DCM (100 mL) and stirred with sat. aq. sodium potassium tartrate (200 mL) for 2 h at room temperature. The organics were separated and the aqueous layer extracted with DCM (2 × 100 mL). The combined organics were washed with brine

(200 mL), dried over anhydrous MgSO_4 and concentrated under reduced pressure to afford (*R*)-**92** as a colourless oil (4.38 g, 99 %).

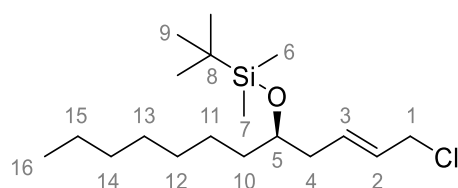
δ_{H} (CDCl_3 , 300 MHz): 0.03 (s, 6H, H-7–8), 0.83–0.93 (m, 12H, H-10, H-17), 1.18–1.34 (m, 10H, H-12–16), 1.35–1.47 (m, 2H, H-11), 2.25–2.40 (m, 2H, H-5), 3.76 (quint, 1H, H-6, $J = 5.5$ Hz), 4.19 (d, 2H, H-2, $J = 4.0$ Hz), 5.83 (m, $2 \times 1\text{H}$, H-3–4); δ_{C} (CDCl_3 , 75 MHz): -4.3 (C-7/8), -4.4 (C-7/8), 14.2 (C-17), 18.3 (C-9), 22.8 (C-16), 25.5 (C-12), 26.0 (C-10), 29.4 (C-14), 29.9 (C-13), 32.0 (C-15), 37.0 (C-12), 40.3 (C-5), 63.9 (C-2), 72.2 (C-6), 129.8 (C-3), 131.3 (C-4); HRMS: Calculated for $[\text{M}+\text{Na}]^+$ 337.2533, found 337.2529. Data are consistent with those reported by Kirsch and coworkers.²²²

5.6.3.11 (*S,E*)-5-((*tert*-Butyldimethylsilyl)oxy)dodec-2-en-1-ol (**92**)



Compound (*S*)-**92** was synthesised from (*S*)-**91**, following the procedure described for (*R*)-**92**. Yield: 4.35 g, 96 %. NMR and LRMS data are consistent with those for (*R*)-**92**.

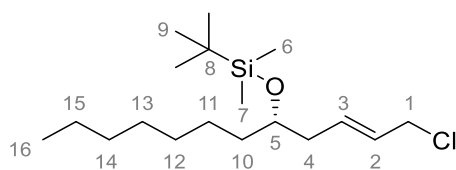
5.6.3.12 (*R,E*)-*tert*-Butyl((1-chlorododec-2-en-5-yl)oxy)dimethylsilane (**93**)



To a solution of (*R,E*)-*tert*-butyl((1-chlorododec-2-en-5-yl)oxy)dimethylsilane **92** (100 mg, 318 μmol , 1.0 eq.) in anhydrous THF (2 mL) under argon were added triphenylphosphine (120 mg, 445 μmol , 1.4 eq.) and *N*-chlorosuccinamide (63.7 mg, 477 μmol , 1.5 eq.), sequentially. The reaction was stirred at room temperature for 3.5 h, then diluted with ice-cold hexane (20 mL) and filtered. The filtrate was concentrated under reduced pressure and the crude material purified by silica gel chromatography (1 : 99 EtOAc : pet. ether) to afford (*R*)-**93** as a colourless oil (96.2 mg, 91 %).

δ_{H} (CDCl_3 , 300 MHz): 0.04 (s, 6H, H-6–7), 0.81-0.93 (m, 12H, H-9, H-16), 1.18-1.34 (m, 10H, H-11–15), 1.35-1.47 (m, 2H, H-10), 2.13-2.28 (m, 2H, H-4), 3.68 (quint, 1H, H-5, $J = 5.5$ Hz), 4.19 (d, 2H, H-1, $J = 7.0$ Hz), 5.63 (dt, 1H, H-2, $J = 14.5$ Hz, 7.0 Hz), 5.79 (dt, 1H, H-3, $J = 14.5$ Hz, 7.0 Hz); δ_{C} (CDCl_3 , 75 MHz): -4.4 (C-6/7), -4.3 (C-6/7), 14.3 (C-16), 18.3 (C-8), 22.8 (C-15), 25.5 (C-14), 26.0 (C-9), 29.4 (C-13), 29.9 (C-12), 32.0 (C-14), 37.1 (C-11), 40.1 (C-4), 45.5 (C-1), 72.0 (C-5), 128.1 (C-2), 132.8 (C-3); HRMS: Calculated for $[\text{M}+\text{Na}]^+$ 355.2194, found 355.2192.

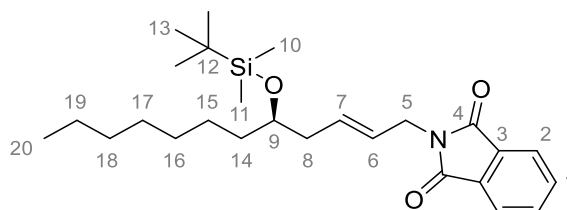
5.6.3.13 (*S,E*)-*tert*-Butyl((1-chlorododec-2-en-5-yl)oxy)dimethylsilane (**93**)



To a solution of (*S,E*)-*tert*-butyl(1-chlorododec-2-en-5-yl)oxydimethylsilane **92** (4.34 g, 13.8 mmol, 1.0 eq.) in anhydrous DCM (100 mL) at 0 °C under argon were slowly added triethylamine (2.88 mL, 20.7 mmol, 1.5 eq.) and a solution of *p*-toluenesulfonyl chloride (2.89 g, 15.2 mmol, 1.1 eq.) in DCM (10 mL). The reaction mixture was stirred at 0 °C for 3 h, then allowed to warm to room temperature and stirred overnight. The mixture was washed sequentially with 1 M HCl (100 mL), H_2O (100 mL) and 5 % aq. Na_2SO_4 (100 mL), and the organics dried over anhydrous MgSO_4 and concentrated under reduced pressure. The crude material was purified by silica gel chromatography (1 : 99 EtOAc : pet. ether) to afford (*S*)-**93** as a colourless oil (1.11 g, 24 %).

δ_{H} (CDCl_3 , 300 MHz): 0.04 (s, 6H, H-6–7), 0.81-0.93 (m, 12H, H-9, H-16), 1.18-1.34 (m, 10H, H-11–15), 1.35-1.47 (m, 2H, H-10), 2.13-2.28 (m, 2H, H-4a, H-4b), 3.68 (quint, 1H, H-5, $J = 5.5$ Hz), 4.19 (d, 2H, H-1, $J = 7.0$ Hz), 5.63 (dt, 1H, H-2, $J = 14.5$ Hz, 7.0 Hz), 5.79 (dt, 1H, H-3, $J = 14.5$ Hz, 7.0 Hz); δ_{C} (CDCl_3 , 75 MHz): -4.4 (C-6/7), -4.3 (C-6/7), 14.3 (C-16), 18.3 (C-8), 22.8 (C-15), 25.5 (C-14), 26.0 (C-9), 29.4 (C-13), 29.9 (C-12), 32.0 (C-14), 37.1 (C-11), 40.1 (C-4), 45.5 (C-1), 72.0 (C-5), 128.1 (C-2), 132.8 (C-3); HRMS: Calculated for $[\text{M}+\text{Na}]^+$ 355.2194, found 355.2192.

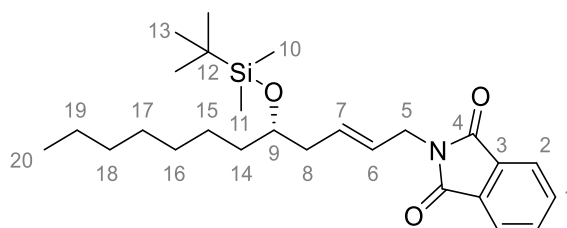
5.6.3.14 (*R,E*)-2-(5-((*tert*-Butyldimethylsilyl)oxy)dodec-2-en-1-yl)isoindoline-1,3-dione (**94**)



To a solution of (*R,E*)-*tert*-butyl((1-chlorododec-2-en-5-yl)oxy)dimethylsilane **93** (1.00 g, 3.01 mmol, 1.0 eq.) in anhydrous DMF (10 mL) under argon was added potassium phthalimide (837 mg, 4.52 mmol, 1.5 eq.), and the reaction stirred at room temperature for 72 h. The mixture was diluted with Et₂O (50 mL) and brine (50 mL) and the organics separated. The aqueous layer was then extracted with Et₂O (2 × 50 mL) and the combined organics washed with brine (2 × 50 mL), dried over anhydrous MgSO₄ and concentrated under reduced pressure. The crude material was purified by silica gel chromatography (7 : 93 EtOAc : pet. ether) to afford (*R*)-**94** as a colourless oil (1.29 g, 97 %).

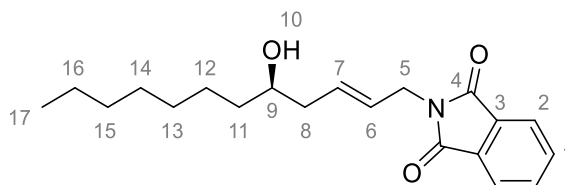
δ_{H} (CDCl₃, 300 MHz): -0.01 (s, 6H, H-10–11), 0.82-0.92 (m, 12H, H-13, H-20), 1.17-1.42 (m, 10H, H-15–19), 1.35-1.47 (m, 2H, H-14), 2.16 (t, 2H, H-8, $J = 6.0$ Hz), 3.68 (quint, 1H, H-9, $J = 5.5$ Hz), 4.19 (d, 2H, H-5, $J = 6.0$ Hz), 5.54 (dt, 1H, H-6, $J = 15.5$ Hz, 6.0 Hz), 5.74 (dt, 1H, H-7, $J = 15.5$ Hz, 7.0 Hz), 7.71 (dd, 2H, H-1, $J = 5.5$ Hz, 3.0 Hz), 7.84 (dd, 2H, H-2, $J = 5.5$ Hz, 3.0 Hz); δ_{C} (CDCl₃, 75 MHz): -4.4 (C-10/11), -4.3 (C-10/11), 14.2 (C-20), 18.2 (C-12), 22.8 (C-19), 25.4 (C-15), 26.0 (C-13), 29.4 (C-17), 29.8 (C-16), 32.0 (C-18), 37.0 (C-14), 39.7 (C-5), 40.3 (C-8), 72.0 (C-9), 123.4 (C-2), 125.4 (C-6), 131.8 (C-7), 132.4 (C-3), 134.0 (C-1), 168.1 (C-4); HRMS: Calculated for [M+Na]⁺ 466.2748, found 466.2750.

5.6.3.15 (*S,E*)-2-(5-((*tert*-Butyldimethylsilyl)oxy)dodec-2-en-1-yl)isoindoline-1,3-dione (**94**)



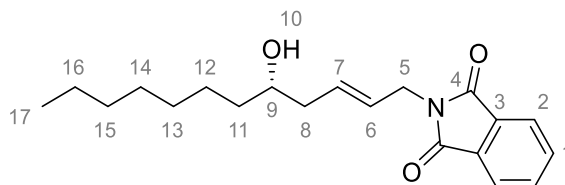
Compound (*S*)-**94** was synthesised from (*S*)-**93**, following the procedure described for (*R*)-**94**. Yield: 1.12 g, 87 %. NMR and LRMS data are consistent with those for (*R*)-**94**.

5.6.3.16 (*R,E*)-2-(5-Hydroxydodec-2-en-1-yl)isoindoline-1,3-dione (**95**)

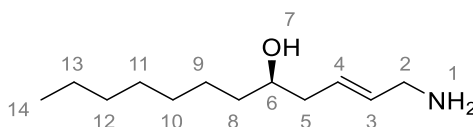


To a solution of (*R,E*)-2-(5-((*tert*-butyldimethylsilyl)oxy)dodec-2-en-1-yl)isoindoline-1,3-dione **94** (50.0 mg, 113 μmol , 1.0 eq.) in a 1 : 1 mixture of MeOH : DCM (1.6 mL) was added (*1S*)-(+)-10-camphorsulfonic acid (27.5 mg, 118 μmol , 1.1 eq.) and the reaction stirred at room temperature for 1.5 h. The mixture was diluted with Et₂O (20 mL), quenched with H₂O (10 mL) and the organics separated. The aqueous layer was then extracted with Et₂O (2 \times 10 mL) and the combined organics washed with H₂O (10 mL), brine (10 mL), dried over anhydrous MgSO₄ and concentrated under reduced pressure. The crude material was purified by silica gel chromatography (1 : 3 EtOAc : pet. ether) to afford (*R*)-**95** as a white solid (37.0 mg, 99 %).

m.p.: 52-54 °C; δ_{H} (CDCl₃, 300 MHz): 0.86 (t, 3H, H-17, $J = 7.0$ Hz), 1.16-1.34 (m, 10H, H-12-16), 1.35-1.47 (m, 2H, H-11), 1.79 (br. s, 1H, H-10), 2.09 (dt, 1H, H-8a, $J = 14.0$, 7.5 Hz), 2.24 (dtd, 1H, H-8b, $J = 14.0$ Hz, 6.0 Hz, 1.0 Hz), 3.57-3.65 (m, 1H, H-9), 4.23 (dd, 1H, H-5a, $J = 15.0$ Hz, 6.0 Hz), 4.30 (dd, 1H, H-5b, 2H, H-5, $J = 15.0$ Hz, 6.0 Hz), 5.60 (dt, 1H, H-6, $J = 15.5$ Hz, 6.0 Hz), 5.76 (dt, 1H, H-7, $J = 15.5$ Hz, 7.5 Hz), 7.71 (dd, 2H, H-1, $J = 5.5$ Hz, 3.0 Hz), 7.84 (dd, 2H, H-2, $J = 5.5$ Hz, 3.0 Hz); δ_{C} (CDCl₃, 75 MHz): 14.2 (C-17), 22.8 (C-16), 25.8 (C-12), 29.4 (C-14), 29.7 (C-13), 31.9 (C-15), 36.9 (C-11), 39.8 (C-5), 40.4 (C-8), 70.8 (C-9), 123.4 (C-2), 126.8 (C-6), 131.3 (C-7), 132.3 (C-3), 134.1 (C-1), 168.2 (C-4); HRMS: Calculated for [M+Na]⁺ 352.1883, found 352.1888.

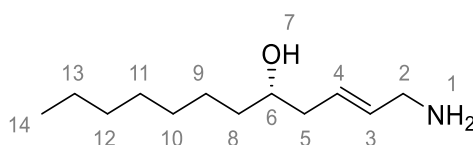
5.6.3.17 (*S,E*)-2-(5-Hydroxydodec-2-en-1-yl)isoindoline-1,3-dione (95**)**

Compound (*S*)-**95** was synthesised from (*S*)-**94**, following the procedure described for (*R*)-**95**. Yield: 292 mg, 95 %. NMR and LRMS data are consistent with those for (*R*)-**95**.

5.6.3.18 (*R,E*)-1-Aminododec-2-en-5-ol (84**)**

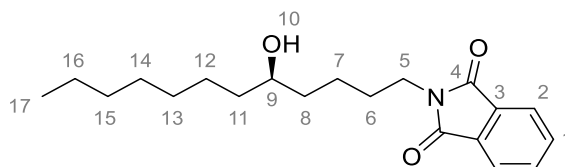
To a solution of (*R,E*)-2-(5-hydroxydodec-2-en-1-yl)isoindoline-1,3-dione **95** (37.0 mg, 112 μ mol, 1.0 eq.) in EtOH (2 mL) was added 40 % aq. MeNH₂ (150 μ L, excess) and the solution stirred at room temperature for 2 h. The reaction was quenched with H₂O (2 mL), basified with 1 M NaOH (pH > 10) and the product extracted with Et₂O (3 \times 10 mL). The combined organics were washed with H₂O (20 mL) and brine (20 mL), dried over anhydrous MgSO₄ and concentrated under reduced pressure to afford (*R*)-**84** as a white solid (21.2 mg, 95 %).

m.p.: 53-56°C; δ_{H} (MeOD, 300 MHz): 0.83 (t, 3H, H-14, $J = 6.5$ Hz), 1.17-1.45 (m, 12H, H-8-13), 2.02-2.22 (m, 2H, H-5), 3.15-3.21 (m, 2H, H-2), 3.45-3.46 (m, 1H, H-6), 5.47-5.77 (m, 2H, H-3-4); δ_{C} (MeOD, 75 MHz): 14.5 (C-14), 23.7 (C-13), 26.8 (C-9), 30.5 (C-11), 30.8 (C-10), 33.0 (C-12), 37.8 (C-8), 41.4 (C-5), 44.0 (C-2), 72.2 (C-6), 130.1 (C-4), 132.1 (C-3); HRMS: Calculated for [M+H]⁺ 200.2009, found 200.2014.

5.6.3.19 (*S,E*)-1-Aminododec-2-en-5-ol (84**)**

Compound (*S*)-**84** was synthesised from (*S*)-**95**, following the procedure described for (*R*)-**84**. Yield: 31.1 mg, 99 %. NMR and LRMS data are consistent with those for (*R*)-**84**.

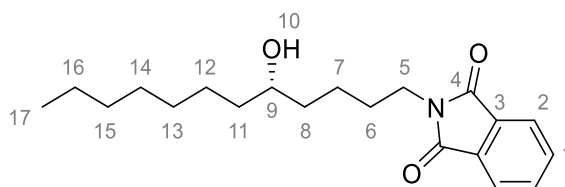
5.6.3.20 (*R*)-2-(5-Hydroxydodecyl)isoindoline-1,3-dione (**98**)



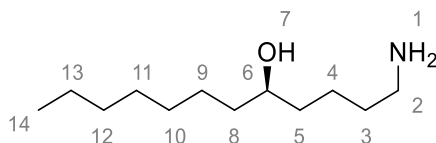
To a stirred solution of (*R*)-**95** (180 mg, 516 μ mol, 1.0 eq.) in MeOH (2.1 mL) was added 10 % Pd/C (60.0 mg). The flask was purged with H₂ \times 5 (1 atm) and stirred at room temperature for 5 h. The catalyst was removed by filtration through celite, the solvent evaporated and the crude material purified by silica gel chromatography (1 : 3 EtOAc : pet. ether) to afford (*R*)-**98** as a white solid (112 mg, 62 %).

m.p.: 62-63 °C; δ_{H} (CDCl₃, 300 MHz): 0.87 (t, 3H, H-17, J = 7.0 Hz), 1.20-1.57 (m, 16H, H-7-8, H-11-16), 1.63-1.77 (m, 2H, H-6), 3.53-3.61 (m, 1H, H-9), 3.69 (t, 2H, H-5, J = 7.0 Hz), 7.70 (dd, 2H, H-1, J = 5.0 Hz, 3.0 Hz), 7.84 (dd, 2H, H-2, J = 5.0 Hz, 3.0 Hz); δ_{C} (CDCl₃, 75 MHz): 14.2 (C-17), 22.8 (C-16), 23.0 (C-7), 25.8 (C-12), 28.7 (C-6), 29.4 (C-14), 29.8 (C-13), 32.0 (C-15), 37.0 (C-8), 37.7 (C-11), 38.0 (C-5), 71.8 (C-9), 123.3 (C-2), 132.3 (C-3), 134.0 (C-1), 168.6 (C-4); HRMS: Calculated for [M+Na]⁺ 354.2040, found 354.2043.

5.6.3.21 (*S*)-2-(5-Hydroxydodecyl)isoindoline-1,3-dione (**98**)

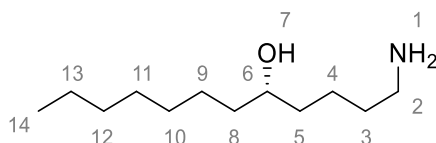


Compound (*S*)-**98** was synthesised from (*S*)-**95**, following the procedure described for (*R*)-**98**. Yield: 97.8 mg, 57 %. NMR and LRMS data are consistent with those for (*R*)-**98**.

5.6.3.22 (R)-1-Aminododecan-5-ol (60)

To a solution of (*R*)-2-(5-hydroxydodecyl)isoindoline-1,3-dione **98** (55.0 mg, 166 μmol , 1.0 eq.) in EtOH (2.8 mL) was added 40 % aq. MeNH₂ (223 μL , excess) and the solution stirred at room temperature for 3 h. The reaction was quenched with H₂O (2 mL), basified with 1 M NaOH (pH > 10) and the product extracted with EtOAc (3 \times 10 mL). The combined organics were washed with H₂O (20 mL) and brine (20 mL), dried over anhydrous MgSO₄ and concentrated under reduced pressure to afford (*R*)-**60** as a white solid (30.2 mg, 90 %). 7 mg was purified by analytical HPLC (table 5.10, elution time = 8.8 min) for accurate determination of the enantioselectivity of CpkH.

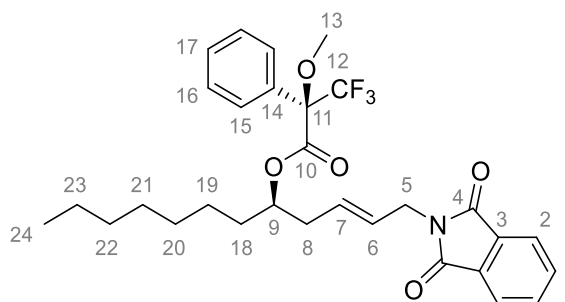
m.p.: 82-85 °C; δ_{H} (MeOD, 500 MHz): 0.90 (t, 3H, H-14, $J = 7.0$ Hz), 1.24-1.39 (m, 10H, H-9-13), 1.40-1.59 (m, 8H, H-3-5, H-8), 1.60-1.73 (m, 2H, H-3), 2.92 (t, 2H, H-2, $J = 7.5$ Hz), 3.50-3.56 (m, 1H, H-6); δ_{C} (MeOD, 125 MHz): 14.4 (C-14), 23.7 (C-4), 23.7 (C-13), 26.8 (C-9), 28.7 (C-3), 30.5 (C-11), 30.8 (C-10), 33.0 (C-12), 37.6 (C-5), 38.6 (C-8), 40.7 (C-2), 72.1 (C-6); HRMS: Calculated for [M+H]⁺ 202.2165, found 202.2170.

5.6.3.23 (S)-1-Aminododecan-5-ol (60)

Compound (*S*)-**60** was synthesised from (*S*)-**98**, following the procedure described for (*R*)-**60**. Yield: 12.6 mg, 94 %. NMR and LRMS data are consistent with those for (*R*)-**60**.

5.6.4 Verification of absolute stereochemistry of substrate intermediates

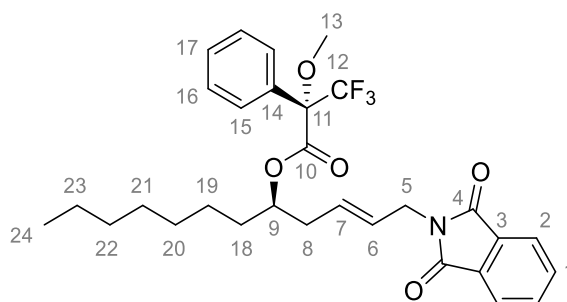
5.6.4.1 (*R,E*)-1-(1,3-Dioxoisindolin-2-yl)dodec-2-en-5-yl (*R*)-3,3,3-trifluoro-2-methoxy-2-phenylpropanoate (**96**)



To a solution of (*R,E*)-2-(5-hydroxydodec-2-en-1-yl)isoindoline-1,3-dione **95** (20.0 mg, 60.7 μmol , 1.0 eq.) in DCM (600 μL) was added a solution of (*R*)-(+)- α -methoxy- α -trifluoromethylphenylacetic acid (MTPA) (15.6 mg, 66.8 μmol , 1.1 eq.) in DCM (200 μL), followed by EDC (12.8 mg, 66.8 μmol , 1.1 eq.) and DMAP (820 μg , 6.07 μmol , 0.1 eq.). The reaction was stirred at room temperature for 16 h, at which time 2 M HCl (1 mL) and H₂O (10 mL) were added. The layers were separated and the aqueous extracted with DCM (2 \times 10 mL). The combined organics were dried over anhydrous MgSO₄, concentrated under reduced pressure and the crude residue purified by silica gel chromatography (1 : 3 EtOAc : pet. ether) to afford (*R,R*)-**96** as a colourless oil (18.0 mg, 54 %).

δ_{H} (CDCl₃, 500 MHz): 0.86 (t, 3H, H-24, $J = 7.0$ Hz), 1.07-1.31 (m, 10H, H-19–23), 1.47-1.60 (m, 2H, H-18), 2.37 (t, 2H, H-8, $J = 6.0$ Hz), 3.51 (s, 3H, H-13), 4.23 (d, 2H, H-5, $J = 5.5$ Hz), 5.09 (quint, 1H, H-9, $J = 6.0$ Hz), 5.61 (dt, 1H, H-6, $J = 15.5$ Hz, 6.0 Hz), 5.67 (dt, 1H, H-7, $J = 15.5$ Hz, 6.5 Hz), 7.37-7.41 (m, 3H, H-15, H-17), 7.49-7.54 (m, 2H, H-16), 7.71 (dd, 2H, H-1, $J = 5.5$ Hz, 3.0 Hz), 7.84 (dd, 2H, H-2, $J = 5.5$ Hz, 3.0 Hz); δ_{C} (CDCl₃, 125 MHz): 14.2 (C-24), 22.7 (C-23), 24.9 (C-21), 29.2 (C-20), 29.3 (C-19), 31.8 (C-22), 33.3 (C-18), 36.8 (C-8), 39.4 (C-5), 55.6 (C-13), 76.5 (C-9), 84.7 (C-11), 123.4 (C-2), 124.6 (C-12), 127.3 (C-6), 127.4 (C-16), 128.5 (C-15), 129.2 (C-17), 129.7 (C-7), 132.3 (C-14), 132.5 (C-3), 134.1 (C-1), 166.3 (C-10), 168.0 (C-4); HRMS: Calculated for [M+Na]⁺ 568.2281, found 568.2286.

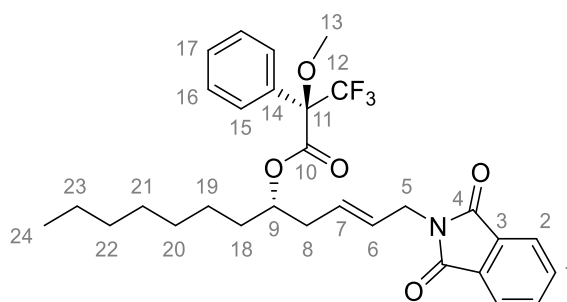
5.6.4.2 (R,E)-1-(1,3-Dioxoisindolin-2-yl)dodec-2-en-5-yl (S)-3,3,3-trifluoro-2-methoxy-2-phenylpropanoate (96)



Compound (*R,S*)-**96** was synthesised from (*R*)-**95** and (*S*)-(-)-MTPA, following the procedure described for the synthesis of (*R,R*)-**96**. Yield: 16.8 mg, 51 %.

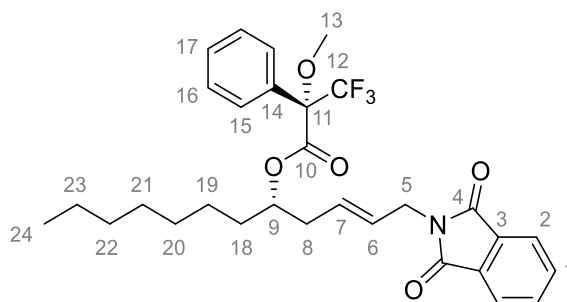
δ_{H} (CDCl_3 , 500 MHz): 0.87 (t, 3H, H-24, $J = 7.0$ Hz), 1.16-1.35 (m, 10H, H-19–23), 1.50-1.66 (m, 2H, H-18), 2.31 (t, 2H, H-8, $J = 6.0$ Hz), 3.51 (s, 3H, H-13), 4.16 (d, 2H, H-5, $J = 5.0$ Hz), 5.09 (quint, 1H, H-9, $J = 6.0$ Hz), 5.50 (dt, 1H, H-6, $J = 15.5$ Hz, 5.5 Hz), 5.57 (dt, 1H, H-7, $J = 15.5$ Hz, 6.5 Hz), 7.30-7.34 (m, 3H, H-15, H-17), 7.42-7.46 (m, 2H, H-16), 7.64 (m, 2H, H-1, $J = 5.5$ Hz, 3.0 Hz), 7.77 (dd, 2H, H-2, $J = 5.5$ Hz, 3.0 Hz); δ_{C} (CDCl_3 , 125 MHz): 14.2 (C-24), 22.7 (C-23), 25.3 (C-21), 29.2 (C-20), 29.4 (C-19), 31.8 (C-22), 33.3 (C-18), 36.4 (C-8), 39.4 (C-5), 55.6 (C-13), 76.6 (C-9), 84.7 (C-11), 123.4 (C-2), 127.2 (C-6), 127.5 (C-16), 128.5 (C-15), 128.9 (C-17), 129.7 (C-7), 132.3 (C-14), 132.5 (C-3), 134.1 (C-1), 166.3 (C-10), 168.0 (C-4); HRMS: Calculated for $[\text{M}+\text{Na}]^+$ 568.2281, found 568.2286.

5.6.4.3 (S,E)-1-(1,3-Dioxoisindolin-2-yl)dodec-2-en-5-yl (R)-3,3,3-trifluoro-2-methoxy-2-phenylpropanoate (96)



Compound (*S,R*)-**96** was synthesised from (*S*)-**95** and (*R*)-(+)-MTPA, following the procedure described for the synthesis of (*R,R*)-**96**. Yield: 21.4 mg, 65 %. NMR and LRMS data are consistent with those for (*R,S*)-**96**.

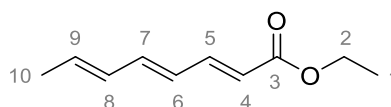
5.6.4.4 (*S,E*)-1-(1,3-Dioxoisindolin-2-yl)dodec-2-en-5-yl (*S*)-3,3,3-trifluoro-2-methoxy-2-phenylpropanoate (**96**)



Compound (*S,S*)-**96** was synthesised from (*S*)-**95** and (*S*)-(-)-MTPA, following the procedure described for the synthesis of (*R,R*)-**96**. Yield: 20.5 mg, 62 %. NMR and LRMS data are consistent with those for (*R,R*)-**96**.

5.6.5 Synthesis of CpkD substrate

5.6.5.1 Ethyl (*2E,4E,6E*)-octa-2,4,6-trienoate (**99**)



To a stirred suspension of NaH (75.2 mg, 3.13 mmol, 1.2 eq.) in anhydrous THF (5 mL) at 0 °C under argon was added triethyl-4-phosphonoacetate (621 μL, 3.13 mmol, 1.2 eq.), dropwise. The mixture was stirred for 1 h, at which time a solution of *trans,trans*-2,4-hexadienal **100** (288 μL, 2.61 mmol, 1 eq.) in THF (2.5 mL) was added dropwise. The reaction was allowed to warm to room temperature and stirred until completion. The reaction was then quenched with sat. aq. NH₄Cl (10 mL) and the aqueous phase extracted with Et₂O (3 × 15 mL). The combined organics were washed with H₂O (10 mL), brine (10 mL) and dried over anhydrous MgSO₄. The solvent was removed under reduced

pressure and the crude residue purified by silica gel chromatography (1 : 9 EtOAc : pet. ether) to afford **99** as a white solid (312 mg, 72 %).

Note: compound exists as mixture of geometric isomers (*E:Z* ≈ 9:1) and m.p. not obtained due to polymerisation; δ_{H} (CDCl₃, 300 MHz): 1.21 (t, 3H, H-1, *J* = 7.0 Hz), 1.75 (d, 3H, H-10, *J* = 7.0 Hz), 4.12 (q, 2H, H-2, *J* = 7.0 Hz), 5.76 (d, 1H, H-4, *J* = 15.5 Hz), 5.77-5.91 (m, 1H, H-9), 6.00-6.16 (m, 2H, H-6, H-8), 6.45 (dd, 1H, H-7, *J* = 15.0 Hz, 10.5 Hz), 7.22 (dd, 1H, H-5, *J* = 15.5 Hz, 11.5 Hz); δ_{C} (CDCl₃, 75 MHz): 14.4 (C-1), 16.6 (C-10), 60.3 (C-2), 120.2 (C-4), 127.7 (C-8), 131.3 (C-6), 135.1 (C-7), 141.1 (C-8), 144.9 (C-5), 167.3 (C-3); HRMS: Calculated for [M+Na]⁺ 189.0886; found 189.0888. Data are consistent with those reported by Goodreid and coworkers.²²³

6 REFERENCES

1. S. A. Waksman and A. T. Henrici, *J. Bacteriol.*, 1943, **46**, 337–341.
2. R. G. Buttery and J. A. Garibaldi, *J. Agr. Food. Chem.*, 1976, **24**, 1246–1247.
3. G. van Keulen and P. J. Dyson, *Adv. Appl. Microbiol.*, 2014, **89**, 217–266.
4. K. Flärdh and M. J. Buttner, *Nat. Rev. Microbiol.*, 2009, **7**, 36–49.
5. E. R. Angert, *Nat. Rev. Microbiol.*, 2005, **3**, 214–224.
6. M. J. Bush, N. Tschowri, S. Schlimpert and K. Flärdh, *Nat. Rev. Microbiol.*, 2015, **13**, 749–760.
7. S. D. Bentley, K. F. Chater and A. M. Cerdeno, *Nature*, 2002, **417**, 141–147.
8. D. Klementz, K. Döring and X. Lucas, *Nucleic Acids Res.*, 2016, **44**, D509–D514.
9. T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich, 2000.
10. D. A. Hopwood, *Microbiology*, 1999, **145**, 2183–2202.
11. M. Nett, H. Ikeda and B. S. Moore, *Nat. Prod. Rep.*, 2009, **26**, 1362–1384.
12. I. Haug, A. Weissenborn and D. Brolle, *Microbiology*, 2003, **149**, 505–513.
13. S. D. Bentley, S. Brown, L. D. Murphy, D. E. Harris, M. A. Quail, J. Parkhill, B. G. Barrell, J. R. MacCormick, R. I. Santamaria, R. Losick, M. Yamasaki, H. Kinashi, C. W. Chen, G. Chandra, D. Jakimowicz, H. M. Kieser, T. Kieser and K. F. Chater, *Mol. Microbiol.*, 2004, **51**, 1615–1628.
14. F. Malpartida and D. A. Hopwood, *Mol. Gen. Genet.*, 1986, **205**, 66–73.
15. J. S. Feitelson, F. Malpartida and D. A. Hopwood, *J. Gen. Microbiol.*, 1985, **131**, 2431–2441.
16. R. Kirby and D. A. Hopwood, *J. Gen. Microbiol.*, 1977, **98**, 239–252.
17. E. Takano, R. Chakraborty, T. Nihira, Y. Yamada and M. J. Bibb, *Mol. Microbiol.*, 2001, **41**, 1015–1028.
18. P. P. Chong, S. M. Podmore, H. M. Kieser, M. Redenbach, K. Turgay, M. Marahiel, D. A. Hopwood and C. P. Smith, *Microbiology*, 1998, **144**, 193–199.
19. N. K. Davis and K. F. Chater, *Mol. Microbiol.*, 1990, **4**, 1679–1691.
20. P. J. Rutledge and G. L. Challis, *Nat. Rev. Microbiol.*, 2015, **13**, 509–523.
21. L. Song, F. Barona-Gomez, C. Corre, L. Xiang, D. W. Udworthy, M. B. Austin, J. P. Noel, B. S. Moore and G. L. Challis, *J. Am. Chem. Soc.*, 2006, **128**, 14754–14755.
22. S. Lautru, R. J. Deeth, L. M. Bailey and G. L. Challis, *Nat. Chem. Biol.*, 2005, **1**, 265–269.
23. C. Corre, L. Song, S. O'Rourke, K. F. Chater and G. L. Challis, *Prot. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17510–17515.

24. K. Kuczek, M. Mordarski and M. Goodfellow, *FEMS Microbiol. Lett.*, 1994, **118**, 317–326.
25. K. Kuczek, K. Pawlik, M. Kotowska and M. Mordarski, *FEMS Microbiol. Lett.*, 1997, **157**, 195–200.
26. M. Gottelt, S. Kol, J. P. Gomez-Escribano, M. Bibb and E. Takano, *Microbiology*, 2010, **156**, 2343–2353.
27. K. Pawlik, M. Kotowska, and P. Kolesiński, *J. Mol. Microbiol. Biotechnol.*, 2010, **19**, 147–151.
28. H. Hu, Q. Zhang and K. Ochi, *J. Bacteriol.*, 2002, **184**, 3984–3991.
29. M. J. Bibb, *Curr. Opin. Microbiol.*, 2005, **8**, 208–215.
30. J. P. Gomez-Escribano, L. Song, D. J. Fox, V. Yeo, M. J. Bibb and G. L. Challis, *Chem. Sci.*, 2012, **3**, 2716–2720.
31. C. C. Ladner and G. J. Williams, *J. Ind. Microbiol. Biotechnol.*, 2016, **43**, 371–387.
32. I. Koryakina, C. Kasey, J. B. McArthur, A. N. Lowell, J. A. Chemler, S. Li, D. A. Hansen, D. H. Sherman and G. J. Williams, *ACS Chem. Biol.*, 2017, **12**, 114–123.
33. B. Shen, *Curr. Opin. Chem. Biol.*, 2003, **7**, 285–295.
34. Y. P. Lim, M. K. Go and W. S. Yew, *Molecules*, 2016, **21**, 806.
35. H. Chen and L. Du, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 541–557.
36. A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson and C. M. Thomas, *Chem. Biol.*, 2003, **10**, 419–430.
37. C. Olano, B. Wilkinson, S. J. Moss, A. F. Braña, C. Méndez, P. F. Leadlay and J. A. Salas, *Chem. Commun.*, 2003, 2780–2782.
38. W. Xu, K. Qiao and Y. Tang, *Crit. Rev. Biochem. Mol. Biol.*, 2013, **48**, 98–122.
39. D. J. Bevitt, J. Cortes, S. F. Haydock and P. F. Leadlay, *Eur. J. Biochem.*, 1992, **204**, 39–49.
40. A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, **29**, 1050–1073.
41. H. Park, B. M. Kevany, D. H. Dyer, M. G. Thomas and K. T. Forest, *PLoS One*, 2014, **9**, e110965.
42. B. S. Moore and C. Hertweck, *Nat. Prod. Rep.*, 2002, **19**, 70–99.
43. Y. A. Chan, A. M. Podevels, B. M. Kevany and M. G. Thomas, *Nat. Prod. Rep.*, 2009, **26**, 90–114.
44. S. Anand, M. V. Prasad, G. Yadav, N. Kumar, J. Shehara, M. Z. Ansari and D. Mohanty, *Nucleic Acids Res.*, 2010, **38**, W487–W496.

45. D. C. Gay, G. Gay, A. J. Axelrod, M. Jenner, C. Kojlhaas, A. Kampa, N. J. Oldham, J. Piel and A. T. Keatinge-Clay, *Structure*, 2014, **22**, 444–451.
46. T. Robbins, J. Kapilivsky, D. E. Cane and C. Khosla, *Biochemistry*, 2016, **55**, 4476–4484.
47. C. Bisang, P. F. Long, J. Cortés, J. Westcott, J. Crosby, A. L. Matharu, R. J. Cox, T. J. Simpson, J. Staunton and P. F. Leadlay, *Nature*, 1999, **401**, 502–505.
48. X. Xie, A. Garg, A. T. Keatinge-Clay, C. Khosla and D. E. Cane, *Biochemistry*, 2016, **55**, 1179–1186.
49. Y. Li, G. J. Dodge, W. D. Fiers, R. A. Fecik, J. L. Smith and C. C. Aldrich, *J. Am. Chem. Soc.*, 2015, **137**, 7003–7006.
50. D. H. Kwan and P. F. Leadlay, *ACS Chem. Biol.*, 2010, **5**, 829–838.
51. D. H. Kwan and F. Schulz, *Molecules*, 2011, **16**, 6092–6115.
52. L. Tang, S. Shah, L. Chung, J. Carney, L. Katz, C. Khosla and B. Julien, *Science*, 2000, **287**, 640–642.
53. M. A. Skiba, A. P. Sikkema, W. D. Fiers, W. H. Gerwick, D. H. Sherman, C. C. Aldrich and J. L. Smith, *ACS Chem. Biol.*, 2016, **11**, 3319–3327.
54. J. Young, D. C. Stevens, R. Carmichael, J. Tan, S. Rachid, C. N. Boddy, R. Müller and R. E. Taylor, *J. Nat. Prod.*, **76**, 2269–2276.
55. T. J. Paulus, J. S. Tuan, V. E. Luebke, G. T. Maine, J. P. DeWitt and L. Katz, *J. Bacteriol.*, 1990, **172**, 2541–2546.
56. P. Argyropoulos, F. Bergeret, C. Pardin, J. M. Reimer, A. Pinto, C. N. Boddy and T. M. Schmeing, *Biochim. Biophys. Acta*, 2016, **1860**, 486–497.
57. U. R. Awodi, J. L. Ronan, J. Masschelein, E. L. C. de los Santos and G. L. Challis, *Chem. Sci.*, 2017, **8**, 411–415.
58. M. Kotowska, K. Pawlik, A. Smulczyk-Krawczynszyn, H. Bartosz-Bechowski and K. Kuczek, *Appl. Environ. Microbiol.*, 2009, **75**, 887–896.
59. M. Kotowska, J. Ciekot and K. Pawlik, *Acta Biochim. Pol.*, 2014, **61**, 141–147.
88. S. Mathew and H. Yun, *ACS Catal.*, 2012, **2**, 993–1001.
60. N. Kessler, H. Schuhmann, S. Morneweg, U. Linne and M. A. Marahiel, *J. Biol. Chem.*, 2004, **279**, 7413–7419.
61. B. Silakowski, G. Nordsiek, B. Kunze, H. Blöcker and R. Müller, *Chem. Biol.*, 2001, **8**, 59–69.
62. G. Zhu, X. Shi and X. Cai, *BMC Biochem.*, 2010, **11**, 46.

63. A. M. Bailey, R. J. Cox, K. Harley, C. M. Lazarus, T. J. Simpson and E. Skellam, *Chem. Commun.*, 2007, **39**, 4053–4055.
64. S. W. Ragsdale, *Chem. Rev.*, 2003, **103**, 2333–2346.
65. K. L. Levert, G. L. Waldrop and J. M. Stephens, *J. Biol. Chem.*, 2002, **277**, 16347–16350.
66. S. Mathew and H. Yun, *ACS Catal.*, 2012, **2**, 993–1001.
67. H. Brundiek and M. Hohne, in *Applied Biocatalysis: From Fundamental Science to Industrial Applications*, ed. L. Hilterhaus, A. Liese, U. Kettling and G. Antranikian, Wiley-VCH, Germany, 1st edn, 2016, ch. 10, pp. 203–206.
68. R. B. Silverman, *The Organic Chemistry of Enzyme Catalyzed Reactions*, Academic Press, San Diego, 2nd edn, 2002.
69. A. Łyskowski, C. Gruber, G. Steinkellner, M. Schürmann, H. Schwab, K. Gruber and K. Steiner, *PloS One*, 2014, **9**, e87450.
70. T. D. H. Bugg, *Introduction to Enzyme and Coenzyme Chemistry*, John Wiley & Sons, Chichester, 3rd edn, 2012, ch. 9, pp. 197–212.
71. L. Skalden, M. Thomsen, M. Höhne, U. T. Bornscheuer and W. Hinrichs, *FEBS J.*, 2015, **282**, 407–415.
72. S. Shätzle, F. Steffen-Munsberg, A. Thontowi, M. Höhne, K. Robins and U. T. Bornscheuer, *Adv. Synth. Catal.*, 2011, **353**, 2439–2445.
73. S.-W. Han, E.-S. Park, J.-Y. Dong, and J.-S. Shin, *Appl. Environ. Microbiol.*, 2015, **81**, 6994–7002.
74. R. L. Fagan and B. A. Palfey, in *Comprehensive Natural Products II: Chemistry and Biology*, ed. L. Mander and H-W. Liu, Elsevier Ltd, Oxford, 1st edn, 2010, vol. 7, pp. 37–113.
75. V. Piano, B. A. Palfey and A. Mattevi, *Trends Biochem. Sci.*, 2017, **42**, 457–459.
76. P. Macheroux, B. Kappes, and S. E. Ealick, *FEBS J.*, 2011, **278**, 2625–2634.
77. T. D. H. Bugg, *Introduction to Enzyme and Coenzyme Chemistry*, John Wiley & Sons, Chichester, 3rd edn, 2012, ch. 6, pp. 115–147.
78. P. F. Fitzpatrick, *Acc. Chem. Res.*, 2001, **34**, 299–307.
79. C. T. Walsh and T. A. Wenczewicz, *Nat. Prod. Rep.*, 2013, **30**, 175–200.
80. P. F. Fitzpatrick, *Arch. Biochem. Biophys.*, 2010, **493**, 13–25.
81. V. Joosten and W. J. H. van Berkel, *Curr. Opin. Chem. Biol.*, 2007, **11**, 195–202.

82. A. Alfieri, F. Fersini, N. Ruangchan, M. Prongjit, P. Chaiyen and A. Mattevi, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 1177–1182.
83. M. M. E Huijbers, S. Montersino, A. H. Westphal, D. Tischler and W. J. H. van Berkel, *Arch. Biochem. Biophys.*, 2014, **544**, 2–17.
84. K. L. Kavanagh, H. Jörnvall, B. Persson and U. Oppermann, *Cell Mol. Life Sci.*, 2008, **65**, 3895–3906.
85. A. Koumanov, J. Benach, S. Atrian, R. González-Duarte, A. Karshikoff and R. Ladenstein, *Proteins*, 2003, **51**, 289–298.
86. J. B. Thoden, T. M. Wohlers, J. L. Fridovich-Keil and H. M. Holden, *Biochemistry*, 2000, **39**, 5691–5701.
87. Y. Liu, J. B. Thoden, J. Kim, E. Berger, A. M. Gulick, F. J. Ruzicka, H. M. Holden and P. A. Frey, *Biochemistry*, 1997, **36**, 10675–10684.
88. U. C. Oppermann, C. Filing, K. D. Berndt, B. Persson, J. Benach, R. Ladenstein and H. Jörnvall, *Biochemistry*, 1997, **36**, 34–40.
89. T. D. H. Bugg, *Introduction to Enzyme and Coenzyme Chemistry*, John Wiley & Sons, Chichester, 3rd edn, 2012, ch. 10, pp. 213–224.
90. L. Liu, K. Iwata, A. Kita, Y. Kawarabayasi, M. Yohda and K. Miki, *J. Mol. Biol.*, 2002, **319**, 479–489.
91. R. Mazumder, T. Sasakawa, Y. Kaziro and S. Ochoa, *J. Biol. Chem.*, 1962, **237**, 3065–3068.
92. C. J. Jeffery, B. J. Bahnson, W. Chien, D. Ringe and G. A. Petsko, *Biochemistry*, 2000, **39**, 955–964.
93. J. M. Schwab and J. B. Klassen, *J. Am. Chem. Soc.*, 1984, **106**, 7217–7227.
94. S. M. Cuesta, N. Furnham, S. A. Rahman, I. Sillitoe and J. M. Thornton, *Curr. Opin. Struct. Biol.*, 2014, **26**, 121–130.
95. D. A. Widdick, K. Dilks, G. Chandra, A. Bottrill, M. Naldrett, M. Pohlschröder and T. Palmer, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 17927–17932.
96. P. Natale, T. Brüser and A. J. M. Driessen, *Biochim. Biophys. Acta*, 2008, **1778**, 1735–1756.
97. J-y. Kato, N. Funai, H. Watanabe, Y. Ohnishi and S. Horinouchi, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 2378–2383.
98. L. Martin Sanchez, PhD thesis, University of Groningen, 2016.
99. N. H. Hsiao, J. Söding, D. Linke, C. Lange, C. Hertweck, W. Wohlleben and E. Takano, *Microbiology*, 2007, **153**, 1394–1404.

100. J. D. Sidda, Poon V, L. Song, W. Wang, K. Yang and C. Corre, *Org. Biomol. Chem.*, 2016, **14**, 6390–6393.
101. J. L. Ramos, M. Martínez-Bueno, A. J. Molina-Henares, W. Terán, K. Watanabe, X. Zhang, M. T. Gallegos, R. Breenan and R. Tobes, *Microbiol. Mol. Biol. Rev.*, 2005, **69**, 326–356.
102. R. Natsume, Y. Ohnishi, T. Senda and S. Horinouchi, *J. Mol. Biol.*, 2004, **336**, 409–419.
103. E. Takano, H. Kinoshita, V. Mersinias, G. Bucca, G. Hotchkiss, T. Nihira, C. P. Smith, M. Bibb, W. Wohlleben and K. Chater, *Mol. Microbiol.*, 2005, **56**, 465–479.
104. X. Li, J. Wang, S. Li, J. Ji, W. Wang and K. Yang, *Sci. Rep.*, 2015, **5**, 1–12.
105. G. Xu, J. Wang, L. Wang, X. Tian, H. Yang, K. Fan, K. Yang and H. Tan, *J. Biol. Chem.*, 2010, **285**, 27440–27448.
106. J. Wang, W. Wang, L. Wang, G. Zhang, K. Fan, H. Tan and K. Yang, *Mol. Microbiol.*, 2011, **82**, 236–250.
107. A. Tanaka, Y. Takano, Y. Ohnishi and S. Horinouchi, *J. Mol. Biol.*, 2007, **369**, 322–333.
108. G. Liu, K. F. Chater, G. Chandra, G. Niu and H. Tan, *Microbiol. Mol. Biol. Rev.*, 2013, **77**, 112–143.
109. M. Nowaczyk, K. Pawlik and M. Kotowska, *The 3rd Workshop on Microbiology in Health and Environment Protection*, II-49, Poland, 2013.
110. B. Karniol and R. D. Vierstra, *J. Bacteriol.*, 2004, **186**, 445–453.
111. Z. Yu, H. Zhu, F. Dang, W. Zhang, Z. Qin, S. Yang, H. Tan, Y. Lu and W. Jiang, *Mol. Microbiol.*, 2012, **85**, 535–556.
112. R. Wang, Y. Mast, J. Wang, W. Zhang, G. Zhao, W. Wohlleben, Y. Lu and W. Jiang, *Mol. Microbiol.*, 2013, **87**, 30–48.
113. S. Chen, G. Zheng, H. Zhu, H. He, L. Chen, W. Zhang, W. Jiang and Y. Lu, *FEMS Microbiol. Lett.*, 2016, **363**, fnw160.
114. S. Rigali, F. Titgemeyer, S. Barends, S. Mulder, A. W. Thomae, D. A. Hopwood and G. P. van Wezel, *EMBO Rep.*, 2008, **9**, 670–675.
115. Y. H. Yang, E. Song, E. J. Kim, K. Lee, W. S. Kim, S. S. Park, J. S. Hahn and B. G. Kim, *Appl. Microbiol. Biotechnol.*, 2009, **82**, 501–511.
116. N. E. Allenby, E. Laing, G. Bucca, A. M. Kierzek and C. P. Smith, *Nucleic Acids Res.*, 2012, **40**, 9543–9546.

117. R. Pérez-Redondo, A. Rodríguez-García, A. Botas, I. Santamarta, J. F. Martín and P. Liras, *PloS One*, 2012, **7**, e32697.
118. E. Takano, T. Nihira, Y. Hara, J. J. Jones, C. J. Gershater, Y. Yamada and M. Bibb, *J. Biol. Chem.*, 2000, **275**, 11010–11016.
119. H. Seto, T. Sato, H. Yonehara and W. C. Jankowski, *J. Antibiot.*, 1973, **26**, 609–611.
120. T. Terashima, E. Idaka, Y. Kishi and T. Goto, *J. Chem. Soc., Chem. Commun.*, 1973, 75–76.
121. Y. Iwai, K. Kumano and S. Ōmura, *Chem. Pharm. Bull.*, 1978, **26**, 736–39.
122. M. Mayer and R. Thiericke, *J. Org. Chem.*, 1993, **58**, 3486–3489.
123. F. J. Leeper, P. Padmanabhan, G. W. Kirby and G. N. Sheldrake, *J. Chem. Soc., Chem. Commun.*, 1987, 505–506.
124. M. Izumikawa, T. Hosoya, M. Takagi and K. Shin-ya, *J. Antibiot. (Tokyo)*, 2012, **65**, 41–43.
125. U. Groenhagen, M. Maczka, J. S. Dickschat and S. Schulz, *Beilstein J. Org. Chem.*, 2014, **10**, 1421–1432.
126. Q. F. Liu, J. D. Wang, X. J. Wang, C. X. Liu, J. Zhang, Y. W. Pang, C. Yu and W. S. Xiang, *J. Asian Nat. Prod. Res.*, 2013, **15**, 221–224.
127. S. Gomi, D. Ikeda, H. Nakamura, H. Naganawa, F. Yamashita, K. Hotta, S. Kondo, Y. Okami, H. Umezawa and Y. Iitaka, *J. Antibiot. (Tokyo)*, 1984, **37**, 1491–1494.
128. T. Nakashima, R. Miyano, M. Iwatsuki, T. Shirahata, T. Kimura, Y. Asami, Y. Kobayashi, K. Shiomi, G. A. Petersson, Y. Takahashi and S. Ōmura, *J. Antibiot. (Tokyo)*, 2016, **69**, 611–615.
129. R. A. Vacca, S. Giannattasio, G. Capitani, E. Marra and P. Christen, *BMC Biochem.*, 2008, **9**, 17.
130. B-Y. Hwang and B-G. Kim, *Enzyme Microb. Technol.*, 2004, **35**, 429–436.
131. H. J. Kim and H. S. Shin, *Anal. Chim. Acta*, 2011, **702**, 225–232.
132. R. Bhushan and H. Brückner, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, 2011, 879, 3148–3161.
133. M. Sova, G. Čadež, S. Turk, V. Majce S. Polanc, S. Batson, A. J. Lloyd, D. I. Roper, C. W. Fishwick and S. Gobec, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 1376–1379.
134. Q. Zhou, J. Xie, X. Yang and L. Wang, *Chinese Patent Application*, WO/2016/082583, 2016.
135. Y. Zhu, J. Xu, X. Mei, Z. Feng, L. Zhang, Q. Zhang, G. Zhang, W. Zhu, J. Liu and C. Zhang, *ACS Chem. Biol.*, 2016, **11**, 943–952.

136. G. Schneider, H. Käck and Y. Lindqvist, *Structure*, 2000, **8**, R1–R6.
137. N. van Oosterwijk, S. Willies, J. Hekelaar, A. C. Terwisscha van Scheltinga, N. J. Turner and B. W. Dijkstra, *Biochemistry*, 2016, **55**, 4422–4431.
138. D. Schiroli and A. Peracchi, *Biochim. Biophys. Acta*, 2015, **1854**, 1200–1211.
139. Y. Li, K. J. Weissman and R. Müller, *J. Am. Chem. Soc.*, 2008, **130**, 7554–7555.
140. A. M. Kunjapur and K. L. J. Prather, *Appl. Environ. Microbiol.*, 2015, **81**, 1892–1901.
141. H. Käck, J. Sandmark, K. Gibson, G. Schneider and Y. Lindqvist, *J. Mol. Biol.*, 1999, **291**, 857–876.
142. H. Lee, J. I. Juncosa and R. B. Silverman, *Med. Res. Rev.*, 2015, **35**, 286–305.
143. M. Markova, C. Peneff, M. J. E. Hewlins, T. Schirmer and R. A. John, *J. Biol. Chem.*, 2005, **280**, 36409–36416.
144. P. Storici, G. Capitani, R. Müller, T. Schirmer and J. N. Jansonius, *J. Mol. Biol.*, 1999, **285**, 297–309.
145. K. Hirotsu, M. Goto, A. Okamoto and I. Miyahara, *Chem. Rec.*, 2005, **5**, 160–172.
146. W. Liu, P. E. Peterson, R. J. Carter, X. Zhou, J. A. Langston, A. J. Fisher and M. D. Toney, *Biochemistry*, 2004, **43**, 10896–10895.
147. T. Arakawa, J. S. Philo, D. Ejima, K. Tsumoto and F. Arisaka, *Bioprocess Int.*, 2006, **4**, 42–43.
148. R. Rapley and J. M. Walker, *Molecular Biomethods Handbook*, Humana Press, New Jersey, 1st ed., 1998.
149. T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Müller, W. Wohlleben, R. Breitling, E. Takano and M. H. Medema, *Nucleic Acids Res.*, 2015, **43**, W237–W243.
150. 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, *Nucleic Acids Res.*, 2014, **43**, D222–D226.
151. S. Donadio and L. Katz, *Gene*, 1992, **111**, 51–60.
152. S. F. Haydock, J. F. Aparicio, I. Molnár, T. Schwecke, L. E. Khaw, A. König, A. F. Marsden, I. S. Galloway, J. Staunton and P. F. Leadlay, *FEBS Lett.*, 1995, **374**, 246–248.
153. F. Del Vecchio, H. Petkovic, S. G. Kendrew, L. Low, B. Wilkinson, R. Lill, J. Cortés, B. A. Rudd, J. Staunton and P. F. Leadlay, *J. Ind. Microbiol. Biotechnol.*, 2003, **30**, 489–494.

154. P. Caffrey, *ChemBioChem*, 2003, **4**, 654–657.
155. A. T. Keatinge-Clay, *Chem. Biol.*, 2007, **14**, 898–908.
156. R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2003, **42**, 72–79.
157. A. T. Keatinge-Clay, *J. Mol. Biol.*, 2008, **384**, 941–953.
158. D. H. Kwan, Y. Sun, F. Schulz, H. Hong, B. Popovic, J. C. Sim-Stark, S. F. Haydock and P. F. Leadlay, *Chem. Biol.*, 2008, **15**, 1231–1240.
159. W. Huang, S. J. Kim, J. Liu and W. Zhang, *Org. Lett.*, 2015, **17**, 5344–5347.
160. S. Ohno, Y. Katsuyama, Y. Tajima, M. Izumikawa, M. Takagi, M. Fujie, N. Satoh, K. Shin-Ya and Y. Ohnishi, *ChemBioChem*, 2015, **16**, 2385–2391.
161. A. Wietzorrek and M. Bibb, *Mol. Microbiol.*, 1997, **25**, 1181–1184.
162. C. Chen, X. Zhao, L. Chen, Y. Jin, Z. K. Zhao and J-W. Suh, *RSC Adv.*, 2015, **5**, 15756–15762.
163. Hindra, P. Pak and M. A. Elliot, *J. Bacteriol*, 2010, **192**, 4973–4982.
164. H. C. Gramajo, E. Takano and M. J. Bibb, *Mol. Microbiol.*, 1993, **7**, 837–845.
165. Y. Chen, E. Wendt-Pienkowski and B. Shen, *J. Bacteriol.*, 2008, **190**, 5587–5596.
166. A. Raynal, A. Friedmann, K. Tuphile, M. Guerineau and J. L. Pernodet, *Microbiology*, 2002, **148**, 61–67.
167. M. Bierman, R. Logan, K. O'Brien, E. T. Seno, R. N. Rao and B. E. Schoner, *Gene*, 1992, **116**, 43–49.
168. A. M. Matter, S. B. Hoot, P. D. Anderson, S. S. Neves and Y-Q. Cheng, *PloS One*, 2009, **4**, e7194.
169. Y-Q. Cheng, *ChemBioChem*, 2006, **7**, 471–477.
170. J. Huang, J. Shi, V. Molle, B. Sohlberg, D. Weaver, M. J. Bibb, N. Karoonuthaisiri, C. J. Lih, C. M. Kao, M. J. Buttner and S. N. Cohen, *Mol. Microbiol.*, 2005, **58**, 1276–1287.
171. P. J. Rutledge and G. L. Challis, *Nat. Rev. Microbiol.*, 2015, **13**, 509–523.
172. D. J. C. Constable, P. J. Dunn, J. D. Hayler, G. R. Humphrey, J. L. Leazer Jr, R. J. Linderman, K. Lorenz, J. Manley, B. A. Pearlman, A. Wells, A. Zaks and T. Y. Zhang, *Green Chem.*, 2007, **9**, 411–420.
173. Invitrogen Life Technologies, *Champion™ pET Directional TOPO® Expression Kits*, 25-0400, 2010.
174. N. Lenfant, T. Hotelier, E. Velluet, Y. Bourne, P. Marchot and A. Chatonnet, *Nucleic Acids Res.*, 2013, **41**, D423–D429.

175. P. G. Bagos, E. P. Nikolaou, T. D. Liakopoulos and K. D. Tsirigos, *Bioinformatics*, 2010, **26**, 2811–2817.
176. J. Koziol, *Methods Enzymol.*, 1971, **18**, 253–285.
177. J. W. Hinkson, *Biochemistry*, 1968, **7**, 2666–2672.
178. G. Cecchini, *Annu. Rev. Biochem.*, 2003, **72**, 77–109.
179. L. Caldinelli, S. Iametti, A. Barbiroli, D. Fessas, F. Bonomi, L. Piubelli, G. Molla and L. Pollegioni, *Protein Sci.*, 2008, **17**, 409–419.
180. J. C. Carlson, S. Li, S. S. Gunatilleke, Y. Anzai, D. A. Burr, L. M. Podust and D. H. Sherman, *Nat. Chem.*, 2011, **3**, 628–633.
181. K. Wilson and J. Walker, *Principles and Techniques of Biochemistry and Molecular Biology*, Cambridge University Press, Cambridge, 7th edn, 2010.
182. R. Noyori, T. Ohkuma, M. Kitamura, H. Takaya, N. Sayo, H. Kumobayashi and S. Akutagawa, *J. Am. Chem. Soc.*, 1987, **109**, 5856–5858.
183. R. Noyori and T. Ohkuma, *Angew. Chem. Int. Ed. Engl.*, 2001, **40**, 40–73.
184. G. Wittig and U. Schöllkopf, *Chem. Ber.*, 1954, **87**, 1318–1330.
185. G. Wittig and W. Haag, *Chem. Ber.*, 1955, **88**, 1654–1666.
186. E. Vedejs and M. J. Peterson, *Top. Stereochem.*, 1994, **21**, 1–157.
187. R. Ding, Y. He, X. Wang, J. Xu, Y. Chen, M. Feng and C. Qi, *Molecules*, 2011, **16**, 5665–5673.
188. S. Gabriel, *Eur. J. Inorg. Chem.*, 1887, **20**, 2224–2236.
189. W. Zhao and J. Montgomery, *J. Am. Chem. Soc.*, 2016, **138**, 9763–9766.
190. I. G. Molnár and R. Gilmour, *J. Am. Chem. Soc.*, 2016, **138**, 5004–5007.
191. J. M. Chapman Jr, G. H. Cocolas and I. H. Hall, *J. Med. Chem.*, 1979, **22**, 1399–1402.
192. T. T. Steckler, M. J. Lee, Z. Chen, O. Fenwick, M. R. Andersson, F. Cacialli and H. Sirringhaus, *J. Mat. Chem. C*, 2014, **2**, 5133–5141.
193. U. Svensson, B. Ringdahl, S. Lindgren and U. Dahlbom, *Acta Pharm. Suec.*, 1975, 290–292.
194. N. Menges and M. Balci, *Synlett*, 2014, **25**, 671–676.
195. J. A. Dale, D. L. Dull and H. S. Mosher, *J. Org. Chem.*, 1969, **34**, 2543–2549.
196. J. A. Dale and H. S. Mosher, *J. Am. Chem. Soc.*, 1973, **95**, 512–519.
197. T. R. Hoye, C. S. Jeffrey and F. Shao, *Nat. Protoc.*, 2007, **2**, 2451–2458.
198. T. D. H. Bugg, *Bioorg. Chem.*, 2004, **32**, 367–375.
199. W. S. Wadsworth and W. D. Emmons, *J. Am. Chem. Soc.*, 1961, **83**, 1733–1738.

200. E. A. Argueta, A. N. Amoh, P. Kafle and T. L. Schneider, *FEBS Lett.*, 2015, **589**, 880–884.
201. E. L. Loechler and T. C. Hollocher, *J. Am. Chem. Soc.*, 1980, **102**, 7312–7321.
202. E. L. Loechler and T. C. Hollocher, *J. Am. Chem. Soc.*, **102**, 7322–7327.
203. K. L. Johnson and D. C. Muddiman, *J. Am. Soc. Mass Spectrom.*, 2004, **15**, 437–445.
204. H. M. Gertz-Hansen, N. Blom, A. M. Feist, S. Brunak and T. N. Petersen, *Proteins*, **82**, 1819–1828.
205. M. Eppink, H. A. Schreuder and W. J. van Berkel, *Protein Sci.*, 1997, **6**, 2454–2458.
206. W. J. van Berkel, N. M. Kamerbeek and M. W. Fraaije, *J. Biotechnol.*, 2006, **124**, 670–689.
207. B. P. Laden, Y. Tang and T. D. Porter, *Arch. Biochem. Biophys.*, 2000, **374**, 381–388.
208. A. D. Hieber, R. C. Bugos and H. Y. Yamamoto, *Biochim. Biophys. Acta*, 2000, **1482**, 84–91.
209. C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman and G. J. Hughes, *Science*, 2010, **329**, 305–309.
210. E. B. Shirling and D. Gottlieb, *Int. J. Syst. Evol. Microbiol.*, 1966, **15**, 313–340.
211. DSMZ GYM Streptomyces Medium, http://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium65.pdf (accessed June 2017).
212. J. Shima, A. Hesketh, S. Okamoto, S. Kawamoto and K. Ochi, *J. Bacteriol.*, 1996, **178**, 7276–7284.
213. S. Ōmura, H. Tanaka, J. Awaya, Y. Narimatsu, Y. Konda and T. Hata, *Agr. Biol. Chem.*, 1973, **38**, 899–906.
214. K. Wiesner, Z. Valenta, D. E. Orr, V. Liede and G. Kohan, *Can. J. Chem.*, 1968, **46**, 3617–3624.
215. K. Soai, H. Oyamada, M. Takase and A. Ookawa, *Bull. Chem. Soc. Jpn.*, 1984, **57**, 1948–1953.
216. H. Li, L. C. M. Castro, J. Zheng, T. Roisnel, V. Dorcet, J-B. Sortais and C. Darcel, *Angew. Chem. Int. Ed.*, 2013, **52**, 8045–8049.
217. Y. Fukuda, S. Matsubara and K. Utimoto, *J. Org. Chem.*, 1991, **56**, 5812–5816.
218. D. Sieben, A. Santana, P. Nowka, S. Weber, K. Funke and S. H. Hüttenhain, *Tetrahedron Lett.*, 2016, **57**, 808–810.

219. C. Wu, P. A. Miller and M. J. Miller, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 2611–2615.
220. P. J. Kocieński, B. Pelotier, J-M. Pons and H. Prideaux, *J. Chem. Soc., Perkin Trans. I*, 1998, **8**, 1373–1382.
221. H. Nemoto, W. Zhong, T. Kawamura, M. Kamiya, Y. Nakano and K. Sakamoto, *Synlett.*, 2007, **15**, 2343–2346.
222. S. F. Kirsch, P. Klahn and H. Menz, *Synthesis*, 2011, **22**, 3592–3603.
223. J. D. Goodreid, K. Wong, E. Leung, S. E. McCaw, S. D. Gray-Owen, A. Lough, W. A. Houry and R. A. Batey, *J. Nat. Prod.*, 2014, **77**, 2170–2181.
224. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. D Biol. Crystallogr.*, 2011, **67**, 235–242.
225. A. G. Leslie, *Acta Crystallogr. D Biol. Crystallogr.*, 2006, **62**, 48–57.
226. AIMLESS (CCP4: Supported Program), <http://www.ccp4.ac.uk/html/aimless.html> (accessed June 2017).
227. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.
228. P. Emsley, B. Lohkamp, W. G. Scott and K. D. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 486–501.
229. A. A. Vagin, R. A. Steiner, A. A. Lebedev, L. Potterton, S. J. McNicholas, F. Long and G. N. Murshudov, *Acta Crystallogr. D Biol. Crystallogr.*, 2004, **60**, 2184–2195.
230. H. Peng, E. Wei, J. Wang, Y. Zhang, L. Cheng, H. Ma, Z. Deng and X. Qu, *ACS Chem. Biol.*, 2016, **11**, 3278–3283.

7 APPENDICES

Appendix 1

Table 7.1: Proposed functions of proteins encoded within the coelimycin biosynthetic gene cluster in *Streptomyces coelicolor* A3(2).³⁰ Open reading frames (ORFs) are coloured as follows. Red: PKS component; blue: post-PKS tailoring; green: regulatory; purple: precursor supply; yellow: export; and grey: hypothetical/unknown function.

ORF	CDS number	Proposed function
ScbC	SCO6264	Ketoreductase involved in SCB1 biosynthesis
ScbR	SCO6265	Butyrolactone-responsive repressor protein
ScbA	SCO6266	Butenolide synthase involved in SCB1 biosynthesis
ScbB	SCO6267	Oxidoreductase involved in SCB1 biosynthesis
CpkM	SCO6268	Two-component system histidine kinase
CpkP β	SCO6269	α -Ketoacid-dependent ferredoxin reductase β -subunit
CpkP α	SCO6270	α -Ketoacid-dependent ferredoxin reductase α -subunit
AccA1	SCO6271	Acyl-CoA carboxylase α -subunit (biotinylated)
ScF	SCO6272	Secreted flavin-dependent epoxidase/dehydrogenase
CpkC	SCO6273	Polyketide synthase module 5
CpkB	SCO6274	Polyketide synthase modules 3 and 4
CpkA	SCO6275	Polyketide synthase loading module, and modules 1 and 2
CpkD	SCO6276	Flavin-dependent epoxidase/dehydrogenase
CpkE	SCO6277	Isomerase? (α , β -hydrolase fold)
CpkF	SCO6278	Transmembrane efflux protein
CpkG	SCO6279	PLP-dependent ω -transaminase
CpkO	SCO6280	SARP-family transcriptional activator
CpkH	SCO6281	Secreted flavin-dependent epoxidase/dehydrogenase
CpkI	SCO6282	Nicotinamide-dependent dehydrogenase
CpkJ	SCO6283	NmrA-family protein (unknown function)
CpkK	SCO6284	Acyl-CoA carboxylase β -subunit
CpkL	SCO6285	Hypothetical protein (unknown function)
ScbR2	SCO6286	Repressor protein
ScoT	SCO6287	Type II thioesterase
CpkN	SCO6288	SARP-family transcriptional activator

Appendix 2

Procedures for acquiring X-Ray diffraction data from CpkG crystals, performed by Dr. Lona Alkhalaf:

Single crystals of CpkG (PACT premierTM, position F12: 0.2 M sodium malonate dibasic monohydrate, 0.1 M bis-tris propane, 20 % w/v PEG 3350, pH 6.5), CpkG_ala and Cpk_ala_lactol (0.2 M sodium malonate dibasic monohydrate, 0.15 M bis-tris propane, 20 % w/v PEG, pH 6.5) were selected and flash frozen in liquid N₂. All X-ray data sets were collected at 100 K at the microfocus beam line I24 at the Diamond Light Source, U.K. All data handling was carried out using the CCP4 software package²²⁴. Data were indexed, integrated and scaled using mosflm²²⁵ and aimless²²⁶. The apo structure was solved by molecular replacement using the coordinates of 5dds¹³⁵ using Phaser²²⁷ and refined using alternate cycles of manual refitting using COOT²²⁸ and REFMAC²²⁹. Water molecules were added to the atomic model automatically at the positions of large positive peaks in the difference electron density, only at places where the resulting water molecule fell into an appropriate hydrogen bonding environment. CpkG_ala and CpkG_ala_lactol were then solved by molecular replacement using the structure of CpkG and refined similarly. The polypeptide chain could be unambiguously traced in all structures between residues 16 and 523. Data collection and refinement statistics are presented in table 7.2.

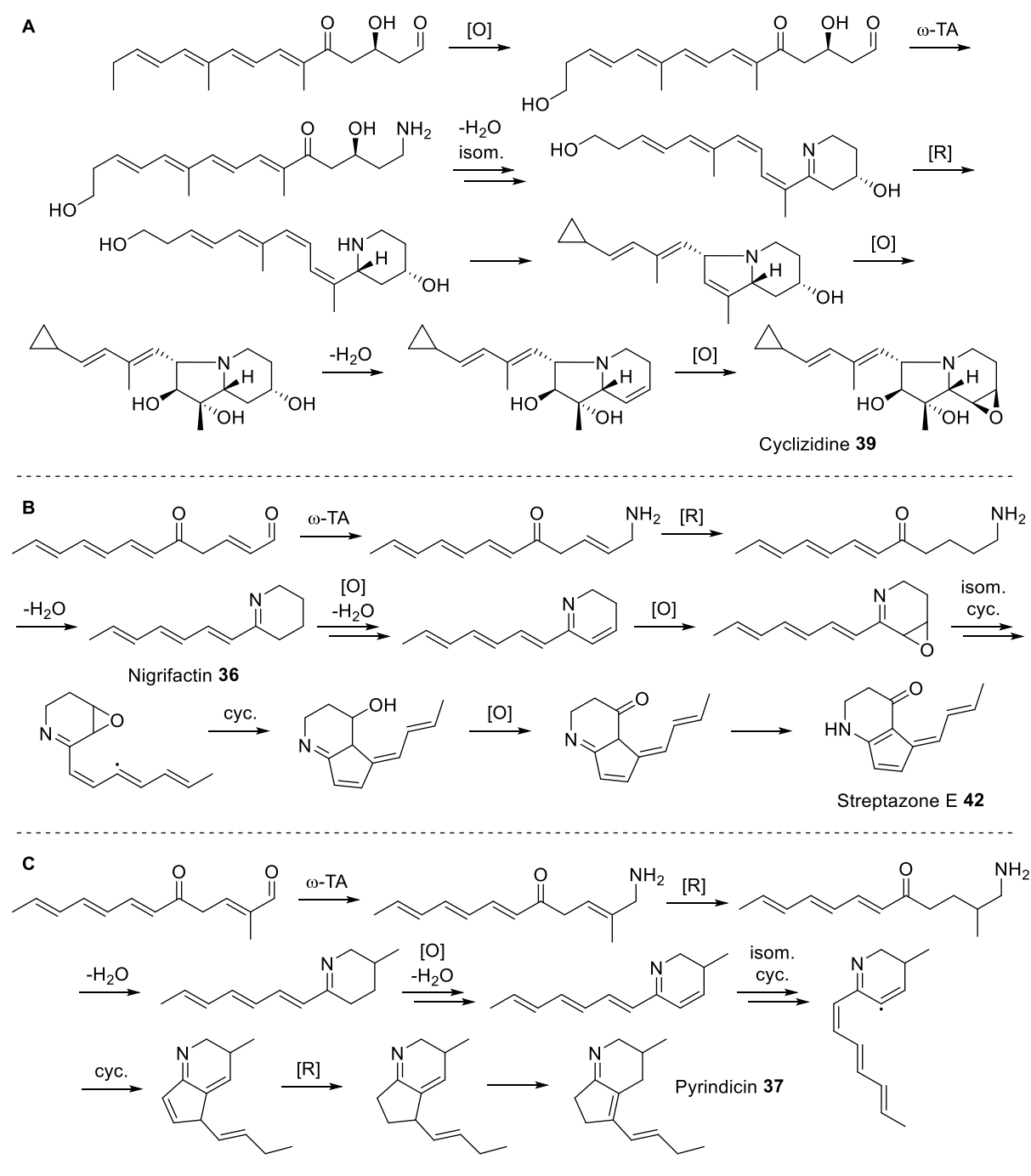
Appendix 3

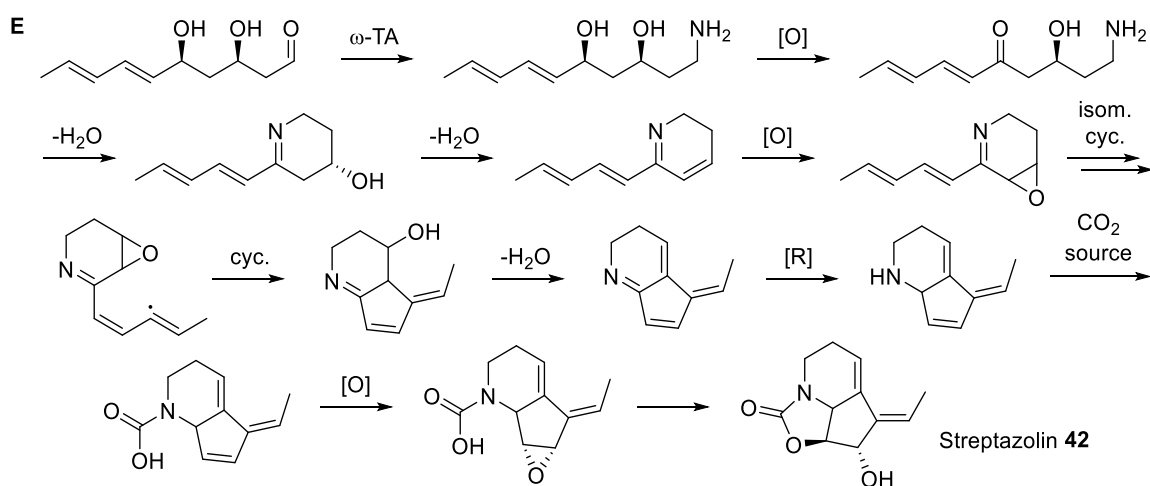
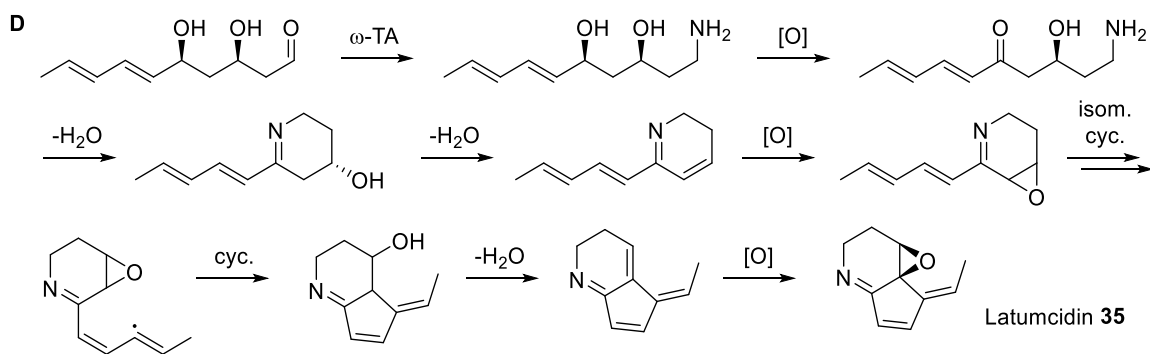
Table 7.2: Statistics from data collection and refinement of crystal structures of CpkG, CpkG_ala and Cpk_ala_lactol

	CpkG	CpkG_ala	CpkG_ala_lactol
Data collection			
Space group	P21 21 21	P21 21 21	P21 21 21
Cell dimensions			
a, b, c (Å)	84.69, 115.10, 126.04	84.30, 114.48, 124.8	84.62, 113.67, 123.99
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength	0.9763	0.9762	0.9762
Resolution (Å)	2.85	2.60	2.60
R_{pim}	0.092 (0.240)	0.100 (0.487)	0.066 (0.237)
$I / \sigma I$	7.6 (3.4)	10.3 (4.1)	12.7(6.0)
Completeness (%)	98.0 (98.5)	100 (100)	99.6 (99.5)
Redundancy	6.8 (6.6)	12.1 (12.5)	11.5 (11.9)
Refinement			
Resolution (Å)	2.85	2.60	2.60
No. reflections	28761	37891	37338
$R_{\text{work}} / R_{\text{free}}$	0.224 / 0.288	0.192 / 0.253	0.197 / 0.259
No. atoms			
Protein	7432	7734	7708
Ligand/ion	32	32	-
Water	2	125	139
B-factors			
Protein	45.4	37.5	37.8
Ligand/ion	55.3	40.6	-
Water	25.3	36.6	30.0
R.m.s deviations			
Bond lengths (Å)	0.011	0.013	0.012
Bond angles (°)	1.7	1.8	1.7

Appendix 4

Proposed biosyntheses of (A) cyclizidine **39**; (B) nigrifactin **36** and streptazone E **42**; (C) pyrindicin **37**; (D) latumcidin **35**; and (E) streptazolin **38**, from predicted polyketide aldehyde intermediates. [O] = Dehydrogenase or oxygenase; [R] = reductase; ω -TA = ω -transaminase; isom. = isomerase; cyc. = cyclase. Pathways are based on those provided by Ohno, Huang, Peng and their respective coworkers.^{160,159,230}





Appendix 5

Table 7.3: Proposed functions of proteins encoded within the putative polyketide alkaloid biosynthetic gene cluster in *Streptomyces cyaneofuscatus* NRRL B-2570. Closest protein homologues are displayed with sequence identities given in brackets. Open reading frames (ORFs) are coloured as follows. Red: PKS component; blue: post-PKS tailoring; and green: regulatory.

ORF	Closest protein homologue	Proposed function
IF17_ RS0112370	MerR family transcriptional regulator from <i>Streptomyces sp.</i> IB2014 011-1 (99 %)	MerR family transcriptional regulator
IF17_ RS0112375	PPOX class F420-dependent enzyme from <i>Streptomyces sp.</i> IB2014 011-1 (100 %)	Oxidoreductase (PMP oxidase?)
IF17_ RS0112380	Aspartate aminotransferase family protein from <i>Streptomyces sp.</i> IB2014 011-1 (99 %)	PLP-dependent ω -transaminase
IF17_ RS0112385	Polyketide synthase from <i>Streptomyces sp.</i> IB2014 011-1 (99 %)	PKS loading module, and modules 1 and 2
IF17_ RS0112390	Polyketide synthase from <i>Streptomyces sp.</i> IB2014 011-1 (99 %)	PKS module 3
IF17_ RS0112395	Polyketide synthase from <i>Streptomyces sp.</i> IB2014 011-1 (99 %)	PKS module 4
IF17_ RS0112400	FAD-binding monooxygenase from <i>Streptomyces sp.</i> IB2014 011-1 (99 %)	FAD-dependent monooxygenase
IF17_ RS0112405	NAD(P)-dependent oxidoreductase from <i>Streptomyces sp.</i> IB2014 011-1 (100 %)	NAD(P)-dependent oxidoreductase
IF17_ RS0112410	Hypothetical protein from <i>Streptomyces sp.</i> IB2014 011-1 (99 %); hydrolase from <i>Streptomyces sp.</i> NRRL F-5193 (64 %)	Hydrolase
IF17_ RS0112415	HTH-type transcriptional regulator CdhR from <i>Streptomyces sp.</i> IB2014 011-1 (100 %)	Transcriptional regulator
IF17_ RS0112420	6-Phosphogluconate dehydrogenase from <i>Streptomyces sp.</i> IB2014 011-1 (100 %)	Dehydrogenase
IF17_ RS0112425	Pseudooxynicotine oxidase from <i>Streptomyces sp.</i> IB2014 011-1 (99 %)	Amine oxidase/dehydrogenase/monooxygenase
IF17_ RS0112430	Cytochrome P450 from <i>Streptomyces sp.</i> IB2014 011-1 (99 %)	Cytochrome P450
IF17_ RS0112435	Transcriptional regulator from <i>Streptomyces sp.</i> IB2014 011-1 (99 %)	SARP-family transcriptional regulator

Appendix 6

Procedures and data for the whole genome sequencing, assembly and annotation of *Streptomyces abikoensis* DSM 40831, performed by Dr. Emmanuel de los Santos:

Genomic DNA was extracted and purified from *Streptomyces abikoensis* DSM 40831 using the phenol-chloroform method. It was used to construct a PacBio library that was sequenced on two SMRT cells of a PacBio RSII machine using C4-C6 chemistry. 1-15 kb reads were assembled using HGAP3 and PacBio SMRT analysis software, to generate a single contig consisting of 7519709 bp, with a GC content of 72.16 %. PROKKA and antiSMASH version 4.0 were used for genome annotation and mining, respectively. The 74 putative specialised metabolite gene clusters can be viewed at the following weblink:

<http://antismash.secondarymetabolites.org/upload/bacteria-f7924202-ff6a-4f98-a72b-5046d8cffa0d/index.html>

Appendix 7

Table 7.4: Proposed functions of proteins encoded within the putative latumcidin biosynthetic gene cluster in *Streptomyces abikoensis* DSM 40831. Closest protein homologues and those from the coelimycin (*cpk*)³⁰, streptazone E (*stz*)¹⁶⁰ and cyclizidine (*cyc*)¹⁵⁹ gene clusters are displayed. Sequence identities are shown in brackets. Open reading frames (ORFs) are coloured as follows. Red: PKS component; blue: post-PKS tailoring; green: regulatory; purple: precursor supply; and grey: hypothetical/unknown function.

ORF	Closest protein homologue	Proposed function	Protein homologue		
			In <i>cpk</i> cluster	In <i>stz</i> cluster	In <i>cyc</i> cluster
Abik_03606	Acyl-CoA dehydrogenase from <i>Streptomyces sp.</i> PAN_FS17 (64 %)	Short-chain dehydrogenase	-	StzJ (50 %)	CycK (27 %)
Abik_03607	6-Phosphogluconate dehydrogenase from <i>Streptomyces sp.</i> CNS615 (69 %)	Short-chain dehydrogenase	-	StzI (45 %)	-
Abik_03608	Mycothioliol maleyl pyruvate isomerase from <i>Streptomyces sp.</i> NRRL S-146 (62 %)	Unknown/hypothetical protein	-	-	-
Abik_03609	Transcriptional regulator AfsR from <i>Streptomyces sp.</i> IB2014 011-1 (68 %)	SARP-family transcriptional activator	CpkN (65 %)	StzR (33 %)	CycO (38 %)
Abik_03610	Methylmalonyl-CoA carboxyl-transferase from <i>Streptomyces coelicolor</i> A3(2) (86 %)	Acyl-CoA carboxylase β -subunit	CpkK (87 %)	-	-
Abik_03611	Flavin reductase domain-containing FMN-binding protein from <i>Catenulispora acidiphila</i> (55 %)	Flavin reductase	-	StzG (37 %)	CycL (37 %)
Abik_03612	Transposase from <i>Streptomyces sp.</i> TLI_105 (50 %)	Transposase?	-	-	-
Abik_03613	TetR family transcriptional regulator from <i>Streptomyces cinnamoneus</i> (80 %)	Butyrolactone-responsive repressor protein	ScbR (46 %)	-	-
Abik_03614	Hypothetical protein from <i>Streptomyces cinnamoneus</i> (59 %)	Butenoloid synthase involved in γ -butyrolactone biosynthesis	ScbA (32 %)	-	-

Abik_03615	Short-chain dehydrogenase from <i>Streptomyces cinnamoneus</i> (85 %)	Short-chain dehydrogenase	CpkI (32 %)	-	CycF (40 %)
Abik_03616	Thioester reductase-domain containing protein from <i>Streptomyces sp.</i> 2114.2 (65 %)	PKS module 4	CpkC (65 %)	StzB (57 %)	CycB (52 %)
Abik_03617	Polyketide synthase from <i>Streptomyces coelicolor</i> (58 %)	PKS module 3	CpkB (58 %)	StzC (52 %)	CycE (49 %)
Abik_03618	Polyketide synthase from <i>Streptomyces lushanensis</i> (68 %)	PKS loading module, and modules 1 and 2	CpkA (66 %)	StzD (65 %)	CycC (53 %)
Abik_03619	Hydrolase from <i>Streptomyces griseus</i> (79 %)	Cyclase	-	StzE (49 %)	CycJ (36 %)
Abik_03620	Monoamine oxidase from <i>Streptomyces albobiridis</i> (64 %)	Monooxygenase	-	StzK (41 %)	-
Abik_03621	Hypothetical protein from <i>Streptomyces sp.</i> NRRL F-5140 (75 %)	Cyclase	-	StzF (52 %)	-
Abik_03622	Transcriptional regulator from <i>Longispora albida</i> (62 %)	Transcriptional regulator	-	-	-
Abik_03623	Hypothetical protein from <i>Streptomyces venezuelae</i> (53 %)	SARP-family transcriptional activator	CpkO (50 %)	StzR (36 %)	-
Abik_03624	Class III aminotransferase from <i>Streptomyces exfoliatus</i> (66 %)	PLP-dependent ω -transaminase	CpkG (61 %)	StzH (59 %)	CycI (55 %)
Abik_03625	Oleoyl-ACP hydrolase from <i>Streptomyces sp.</i> WM4235 (61 %)	Type II thioesterase	ScoT (61 %)	StzA (53 %)	-
Abik_03626	Hypothetical protein from <i>Streptomyces sp.</i> MP131-18 (53 %)	Repressor protein	ScbR2 (33 %)	-	-
Abik_03627	Two-component system histidine kinase from <i>Streptomyces griseus</i> (47 %)	Two-component system histidine kinase	CpkM (36 %)	-	-
Abik_03628	Acetyl-/propionyl-CoA carboxylase subunit α from <i>Streptomyces sp.</i> CNT372 (90 %)	Acyl-CoA carboxylase α -subunit	AccA1 (86 %)	-	-

Appendix 8

Procedures and data for the confirmation of latumcidin **35** production in *Streptomyces abikoensis* DSM 40831, performed by Dr. Yousef Dashti and Dr. Daniel Zabala:

ISP2 agar was inoculated with *Streptomyces abikoensis* DSM 40831 spore stock solution (15 μ L) and incubated at 30 °C for 7 days. Metabolites were then extracted with EtOAc and redissolved in acetonitrile as described in sections 5.3.6.1 and 5.3.6.3. The samples were analysed by high-resolution LC-MS according to the elution profile in table 7.5, and the absorbance monitored at 210 and 254 nm.

Table 7.5: Elution conditions used for LC-MS detection of latumcidin **35**

Time / min	H ₂ O / % (0.1 % FA)	Acetonitrile / % (0.1 % FA)	Flow rate / mL/min
0.0	95	5	0.2
5.3	95	5	0.2
17.3	0	100	0.2
22.3	0	100	0.2
25.3	95	5	0.2
33.3	95	5	0.2

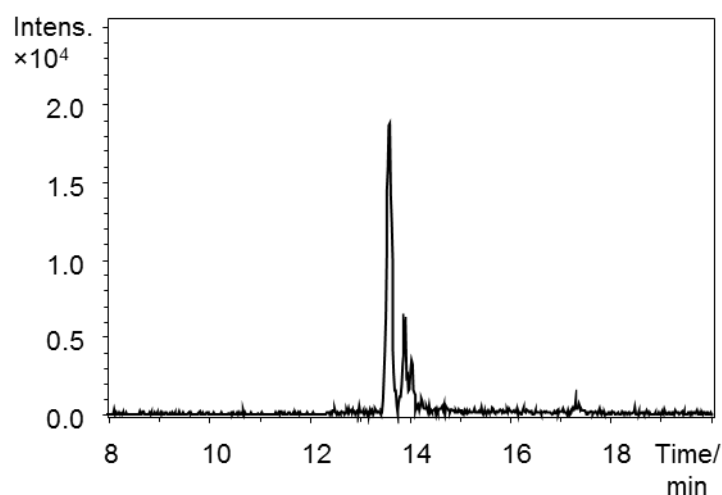


Figure 7.1: Extracted ion chromatogram at $m/z = 162.0913 \pm 0.05$ (corresponding to the $[M+H]^+$ ion for latumcidin **35**). The calculated molecular formula ($C_{10}H_{12}NO^+$) is also consistent with that of latumcidin.

Appendix 9

MS/MS data and calculated molecular formulae arising from the fragmentation of ions at $m/z = 324$ (top; corresponding to the NAC-derivatised mono-epoxide) and $m/z = 340$ (bottom; corresponding to the NAC-derivatised bis-epoxide), following incubation of triene ester **99** with His₆-CpkD, FMN, NAC and NADH. Performed by Dr. Lijiang Song.

