

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

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/146288>

Copyright and reuse:

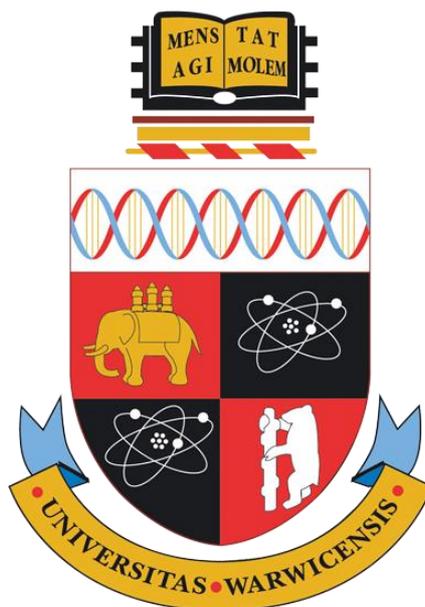
This thesis is made available online and is protected by original copyright.

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



Chemical Probing of Nonribosomal Peptide Assemblies

by

Daniel James Leng

A thesis submitted in partial fulfilment of the requirements for the degree
of

Doctor of Philosophy in Life Sciences

Department of Chemistry

University of Warwick

September 2019

Table of Contents

<i>List of Figures</i>	6
<i>List of Tables</i>	16
<i>List of Schemes</i>	17
<i>Acknowledgements</i>	18
<i>Declaration</i>	19
<i>Abstract</i>	20
<i>Abbreviations</i>	22
<i>1. Introduction</i>	26
1.1 Natural products	26
1.2 Nonribosomal peptides	28
1.3 NRPS modules	30
1.3.1 Adenylation domains	30
1.3.2 Thiolation domain.....	32
1.3.3 Condensation domain	34
1.3.4 Cyclisation domain	36
1.3.5 Epimerisation domain	38
1.3.6 X domain	39
1.3.7 Methyltransferase domain.....	40
1.3.8 Thioesterase domain	41
1.3.9 Post Extension Tailoring.....	43
1.4 Generation of structural diversity by NRPSs	44
1.4.1 Class A – Linear NRPSs	45
1.4.2 Class B – Iterative NRPSs	47
1.4.3 Class C – Nonlinear NRPSs.....	50
1.5 Approaches to study NRP biosynthesis	52
1.5.1 Isotopic labelling.....	52
1.5.2 <i>In vitro</i> enzyme characterisation.....	53
1.5.3 <i>In vivo</i> studies	55
1.5.4 Mass spectrometry	60
1.5.5 Chemical probes	62

1.6	Project Aims	68
2.	<i>Investigation of echinomycin biosynthesis</i>	70
2.1	Echinomycin bioactivity	70
2.1.1	Antitumour activity	70
2.1.2	Antibacterial <i>activity</i>	72
2.1.3	Antiviral activity	72
2.2	Echinomycin biosynthesis	73
2.2.1	Quinoxaline-2-carboxylic acid biosynthesis	74
2.2.2	Nonribosomal peptide synthetase	76
2.2.3	Post-assembly line tailoring	78
2.3	Chemically probing echinomycin biosynthesis	78
2.3.1	Thioamide-based chemical probes	80
2.3.2	<i>N</i> -methylated chemical probes	86
2.3.3	Cysteine-based chemical probe	89
2.4	Genetic manipulation of <i>S. lasaliensis</i> with CRISPR-Cas9	92
2.5	TAR capture of the echinomycin BGC	97
2.6	<i>In vitro</i> reconstitution of Ecm6 module 1	103
2.6.1	Protein constructs and expression	103
2.6.2	<i>In vitro</i> activity assays	111
3.1	<i>Chemical probing of colibactin assembly</i>	116
3.1.1	Discovery of colibactin	116
3.1.2	Unravelling colibactin biosynthesis	119
3.2	Chemical probes for the investigation of colibactin biosynthesis	129
3.2.1	Cysteine-based chemical probes	130
3.2.2	Aminomalonate-based chemical probes	136
3.3	Following work	146
4.	<i>Investigation of vancomycin bio-assembly via chemical probes</i>	152
4.1	Vancomycin discovery	152
4.2	Vancomycin activity	153
4.3	Antimicrobial resistance to vancomycin	155

4.4	Derivatisation of vancomycin	157
4.4.1	Glycopeptide biosynthetic engineering.....	157
4.4.2	Semi-synthetic derivatives of vancomycin	159
4.4.3	Total synthesis	161
4.5	Biosynthesis of vancomycin	162
4.5.1	NRPS assembly.....	164
4.5.2	Non-proteinogenic amino acid biosynthesis	165
4.5.3	Tailoring of the heptapeptide	167
4.5.4	Oxidative cross-linking.....	169
4.5.5	Halogenation.....	171
4.6	Chemical probes to investigate vancomycin biosynthesis	174
4.6.1	Synthesis of chemical probes.....	174
4.6.2	Interception and capture of NRPS intermediates in vancomycin bio-assembly	178
5.	<i>Investigation of scleric acid biosynthesis</i>	193
5.1	Genome mining of <i>Streptomyces</i>	193
5.2	Discovery of scleric acid	195
5.3	Proposed biosynthetic route to scleric acid	196
5.4	Synthesis and characterisation of a scleric acid isomer	197
5.5	Biosynthesis of scleric acid	199
5.6	Biological activity of scleric acid	206
5.7	Generation of scleric acid derivatives	206
6.	<i>Conclusions and future work</i>	211
6.1	Investigation of echinomycin biosynthesis	211
6.1.1	Chemical probing of echinomycin biosynthesis	211
6.1.2	<i>In vivo</i> editing of the echinomycin BGC	214
6.1.3	<i>In vitro</i> reconstitution of Ecm6 module 1	215
6.2	Chemical probing of the colibactin biosynthetic pathway	216
6.3	Chemical probing of vancomycin biosynthesis	219
6.4	Study of the biosynthesis of scleric acid	222
7.	<i>Experimental</i>	224

7.1	General chemical methods.....	224
7.1.1	Chemicals and solvents.....	224
7.1.2	Compound characterisation	224
7.1.3	Chemical synthesis	225
7.2	Feeding experiments of bacterial strains with chemical probes	301
7.2.1	General microbiology methods.....	301
7.2.2	Bacterial strains used with chemical probes	301
7.2.3	Bacterial culture media	303
7.2.4	Feeding experiments set up.....	305
7.3	General molecular biology and protein methods	308
7.3.1	Reagents and equipment	309
7.3.2	Bacterial strains used for molecular biology and protein biochemistry	309
7.3.3	Oligonucleotides	310
7.3.4	Plasmids	312
7.3.5	Buffers	314
7.3.6	Antibiotics.....	315
7.2.7	Procedures.....	316
8.	<i>Bibliography</i>	326

List of Figures

Figure 1 - Crystal structure of rapamycin bound to FKBP12 and mTor, and the chemical structure of rapamycin.	27
Figure 2 - Examples of the natural product families: alkaloids, terpenoids, polyketides, and nonribosomal peptides.	28
Figure 3 - Gramicidin S and Tyrocidine A, NRPs used in early investigations of biosynthesis.	29
Figure 4 - a) The adenylation of an amino acid	30
Figure 5 - Structure of PheA (GrsA).....	31
Figure 6 - Loading of the active site serine of thiolation (T) domains	33
Figure 7 - Amide bond formation catalysed by the condensation domain	34
Figure 8 - Crystal structure of CDA-C1(E17C).....	35
Figure 9 - Representative thiazole formation catalysed by the cyclisation domain...37	
Figure 10 - General mechanism of epimerisation to convert L-amino acids to D-amino acids.....	39
Figure 11 - Methylation in yersiniabactin biosynthesis by PKS-NRPS HWMP1.....	40
Figure 12 - Chain release of an NRPS product by the catalytic triad of a thioesterase domain and a nucleophile.....	41
Figure 13 - Homodimerisation and cyclisation by the formation of a peptide bond in the TE release of gramicidin S.....	42
Figure 14 - Teicoplanin with post extension tailoring with <i>N</i> -acetylglucosamine, mannose, acyl chain, and oxidative cross-linking.....	43
Figure 15 - Post extension tailoring steps to produce teicoplanin, including 4 cross-links, 3 glycosylations and acylation.	44
Figure 16 - Vancomycin binding the D-Ala-D-Ala motif of peptidoglycan through five hydrogen bonds.....	45
Figure 17 - Biosynthesis of class I NRP vancomycin heptapeptide by 3 NRPS.	46
Figure 18 - a) Colinear biosynthesis of ACV peptide by ACV synthetase. b) Further processing of ACV peptide to generate penicillin G.	47
Figure 19 - Echinomycin biosynthesis.....	49

Figure 20 - Gramicidin S showing the dimeric structure formed from two pentapeptide chains by an iterative NRPS.	50
Figure 21 - The domain organisation of the yersiniabactin PKS-NRPS, showing an unusual A domain which loads three T domains.	51
Figure 22 - Unusual domain structure in NRPS ClbH which contains two A domains.	51
Figure 23 - Common methods for <i>in vitro</i> study of A domain activity.	54
Figure 24 - Epimerisation of isonocardicin A, 17, to produce nocardicin A, 18, catalysed by NocJ.....	55
Figure 25 - Example of heptapeptides isolated from P450 monooxygenase knockout mutant strains of <i>Amycolatopsis balhimycina</i>	56
Figure 26 - TAR Workflow.....	58
Figure 27 - Single and double crossover mutations	59
Figure 28 - Targeting of the Cas9 nuclease by a single guide RNA (sgRNA) consisting of CRISPR RNA (crRNA) and trans-acting crRNA (tracrRNA)	60
Figure 29 - T domain-bound intermediates in the biosynthesis of pyrrolyl in clorobiocin, detected by ESI-FTMS.	61
Figure 30 - Proposed mechanism for the IRMPD or CAD induced ejection of the pantetheine moiety of NRPS.....	61
Figure 31 - Vinylsulfonamide based probes for A domains, which stabilise A-T domain interactions.	63
Figure 32 - Benzophenone and alkyne functionalized proteomic probes developed by Kakeya and co-workers.....	64
Figure 33 - Probes for the functionalization of natural products	64
Figure 34 - Mechanism of intermediate capture.	65
Figure 35 - a) The biosynthesis of 6-methylsalicylic acid by an iterative type I PKS. b) Offloaded intermediates which support the proposed timing of reduction in the biosynthetic pathway.....	66
Figure 36 – Chain termination probes for PKS biosynthesis.	67

Figure 37 - a) Chemical probes for NRPSs designed to mimic PPant-bound aminoacyl thioesters. b) Chemical probes compete with the acceptor amino acids of the pathway to offload NRP intermediates which can be analysed by LC-MS ⁿ	68
Figure 38 - Structure of the nonribosomal peptide echinomycin.....	70
Figure 39 - Echinomycin binding DNA by intercalation.....	71
Figure 40 - Structure of triostin A, 15, containing a disulfide cross-bridge rather than the thioacetal of echinomycin.	72
Figure 41 - The echinomycin gene cluster (<i>ecm</i>) and the standalone acyl carrier protein (ACP) from primary metabolism, <i>fabC</i>	73
Figure 42 - Proposed biosynthesis of quinoxaline-2-carboxylic acid. The ring opening catalysed by Ecm14 is the latest step to be experimentally characterised. ..	74
Figure 43 - Structure of quinomycin-type antibiotic SW163-D, containing a 3-hydroxyquinaldic acid chromophore.....	75
Figure 44 - Biosynthesis of echinomycin by an iterative NRPS, followed by post-assembly tailoring to install the central cross-bridge.	77
Figure 45 - Proposed mechanism of the conversion of the disulfide to the thioacetal of echinomycin catalysed by Ecm18, using SAM (shown in red).....	78
Figure 46 - Generic structure of a chemical probe for NRPS systems, in which the length of the acyl chain can be varied, and a range of amino acids used.....	78
Figure 47 - Offloading of peptide intermediates of the echinomycin NRPSs Ecm6 and Ecm7 using chemical probes.	79
Figure 48 - Structure of thioamide-containing natural product closthioamide, x, and its biologically inactive hexamide analogue, x.	81
Figure 49 - Feeding of the thioamide-glycine probe 54 to <i>S. lasaliensis</i> S970A at 0.2 mM, 1.0 mM, and 2.0 mM concentrations.....	84
Figure 50 - Detection of the offloading of the quinoxaline-2-carboxylic acid starter unit from echinomycin NRPS using thioamide-glycine probe 54.	85
Figure 51 - Feeding of thioamide-alanine probe 56 to <i>S. lasaliensis</i> S970A.....	86
Figure 52 - Feeding of thioamide-valine probe 56 to <i>S. lasaliensis</i> S970A.....	86
Figure 53 - Feedings of <i>S. lasaliensis</i> S970A with <i>N</i> -methylated Gly probe 67	88
Figure 54 - Feedings of <i>S. lasaliensis</i> S970A with <i>N</i> -methylated Ala probe 68	88

Figure 55 - Feedings of <i>S. lasaliensis</i> S970A with <i>N</i> -methylated Val probe 69	89
Figure 56 - Previously performed deprotection of a cysteine based chemical probe using TFA.	90
Figure 57 - Feedings of <i>S. lasaliensis</i> S970A with Cys probe 75.....	92
Figure 58 - Oligonucleotide sequences used to introduce the gRNA construct into pCRISPomyces2 for each condensation domain of <i>Ecm6</i>	93
Figure 59 - Golden Gate assembly process.....	94
Figure 60 - Sequencing confirms the correct insertion of the protospacer DNA for condensation domain 1 using Golden Gate assembly.....	95
Figure 61 - Gibson assembly process	95
Figure 62 - Sequence for the synthetic DNA used to create the pCAP03 capture vector. The 20 bp regions of homology required for Gibson assembly are shown in yellow, the homologous arms for recombination in yeast are shown in orange and grey, flanking the PmeI cut site.	97
Figure 63 - The action of URA3, converting 5-fluoroorotic acid, 77, to 5-fluorouracil, 78, an anti-metabolite of uracil, 79.	98
Figure 64 - Plasmid map of the pCAP03 capture vector, containing homology arms for the <i>Ecm</i> cluster between the promoter pADH and the URA3 gene.	99
Figure 65 - Plasmid map of the pCAP1000 capture vector, containing homology arms for the <i>Ecm</i> cluster flanking the pADH and URA3 genes.	99
Figure 66 - The echinomycin gene cluster (<i>Ecm</i>), flanked by the restriction sites of NsiI and MseI, which can be used to digest <i>S. lasaliensis</i> genomic DNA.	100
Figure 67 - Yeast transformants from the pCAP03 and pCAP1000 capture vectors	101
Figure 68 - Agarose gel of PCR products when screening for <i>ecm1</i>	102
Figure 69 - DNA fragments that were screened for in yeast transformants and their location within the <i>ecm</i> gene cluster.	103
Figure 70 - Vector contains DNA Topoisomerase I after a 5'-CCCTT sequence, with a GTGG-5' overhang which can insert between a PCR product containing a complementary sequence.	105
Figure 71 - - Plasmid map of pET151-TOPO-FabC.....	106

Figure 72 - SDS-PAGE gel from the expression of FabC	106
Figure 73 - Plasmid map of pET151-TOPO-Ecm1.	107
Figure 74 - SDS-PAGE gel of the fractions from the Nickel affinity purification of the adenylation domain Ecm1.....	107
Figure 75 - SDS-PAGE gel from the expression of Ecm6m1	108
Figure 76 - Ecm6m1 induced using conditions developed by Watanabe et al.	109
Figure 77 - Plasmid map of pET28a(+)-Ecm6C1.....	110
Figure 78 - SDS-PAGE gel of the fractions from Nickel affinity purification of Ecm6C1.....	111
Figure 79 - Formation of a hydroxamic acid from the amino adenylate product of the A domain.....	111
Figure 80 - Colour difference observed in the hydroxylamine assay.	112
Figure 81 - Graph showing the absorbance of phosphomolybdate-malachite green complex in samples containing Ecm1 compared to denatured Ecm1.....	113
Figure 82 - Schematic of the C domain assay using FabC, Ecm1, and Ecm6C1 with chemical probes to test substrate specificity of Ecm6C1.....	114
Figure 83 - Biosynthetic gene cluster coding for colibactin, the <i>pks</i> island.	116
Figure 84 - The elucidated steps of colibactin biosynthesis, with selected identified precolibactins, up to 2017.	118
Figure 85 - <i>In vitro</i> activity of ClbN and ClbB shows production of a prodrug motif which can be cleaved by <i>E. coli</i> expressing peptidase ClbP.....	119
Figure 86 – Intermediate 86 identified from comparative metabolomics, produced by ClbN, ClbB and a second PKS.....	120
Figure 87 - Spiro-cyclopropane containing intermediate 87 isolated from $\Delta clbP$ mutant, and dependent on <i>clbH</i> , <i>C</i> , and <i>I</i>	120
Figure 88 - Proposed structure of precolibactin A 88, identified by metabolomics, and the correct structure 89 identified by synthesis.	122
Figure 89 - Formation of aminomalonyl-ACP from seryl-ACP by dehydrogenases ClbD and ClbF.	123
Figure 90 – Precolibactin, 90, which accumulated in a $\Delta clbG\Delta clbP$ <i>E. coli</i> strain.	123

Figure 91 - Structure of precolibactin-886, 91, the first identified product of the colibactin pathway to contain an aminomalonate moiety.	125
Figure 92 - Lactone metabolite, 92, produced by <i>E. coli</i> expressing <i>clbN</i> , <i>B</i> , <i>C</i> , <i>H</i> , and <i>clbA</i>	125
Figure 93 - Biosynthesis of 1-aminocyclopropanecarboxylic acid (ACC) from SAM, catalysed by <i>ClbH</i>	126
Figure 94 - Differing cyclization patterns of precolibactins containing the <i>N</i> -Acyl-d-Asn prodrug motif present in $\Delta clbP$ strains, and those lacking the prodrug motif in strains containing <i>clbP</i>	127
Figure 95 - Synthetic precolibactin mimic, 98, capable of damaging DNA at nM concentrations.	128
Figure 96 - Putative biosynthetic route to precolibactin-886, 91, from a linear precolibactin, 99, via precolibactin-888, 100.	128
Figure 97 - DNA adduct, 101, identified from HeLa cells treated with <i>pks+</i> <i>E. coli</i> , and confirmed by <i>in vitro</i> alkylation of DNA.	129
Figure 98 - Offloaded intermediates detected in previous work using chemical probes for both PKS and NRPS systems.	130
Figure 99 - Installation of thiazole moieties by <i>ClbJ</i> and <i>ClbK</i> , along with aminomalonate incorporation, with the possibility of module-skipping in <i>ClbK</i>	131
Figure 100 - Detection of the <i>N</i> -myristoyl-D-Asn colibactin prodrug motif used to verify production levels in <i>pks+</i> <i>E. coli</i> and the $\Delta clbA$ mutant strain.	132
Figure 101 - Putative intermediate from <i>pks+</i> <i>E. coli</i> corresponding to offloading from <i>ClbK</i> , followed by dehydration after processing by <i>ClbP</i>	133
Figure 102 - Putative intermediate from <i>pks+</i> <i>E. coli</i> corresponding to offloading from <i>ClbK</i> , followed by double dehydration after processing by <i>ClbP</i>	134
Figure 103 - High resolution mass spectra of the cysteine probe from culture extract (top), and the disulfide species from culture extract (bottom).	135
Figure 104 - Aminomalonate probes, 108 – 110, previously synthesized, protected with a methyl ester bearing C4, C7, or C10 acyl chains.	136
Figure 105 – Putative dual action of aminomalonate-based chemical probes.	137

Figure 106 - UV traces at 210 nm, 254 nm, and 280 nm from the HPLC separation of the reaction mixture from the reduction of oxime, 117.	139
Figure 107 - Aminomalonate-based probes containing trifluoroethanol or butanol esters.....	140
Figure 108 - Formation of pyrazines from cyclisation and oxidation in air of α -amino- β -ketoesters in basic conditions, from Buron <i>et al.</i> ²²⁴	141
Figure 109 - Mass spectra of the probes 136 and 138, and their decarboxylated derivatives 81.	143
Figure 110 - The glycine-based chemical probe 81 and the fluoromalonate-based chemical probe 141 used to investigate colibactin biosynthesis.....	144
Figure 111 - The only putative intermediate detected in feedings conducted with the fluoromalonate probe 141 and glycine probe 81.....	145
Figure 112 - Putative offloaded intermediate and MS ² fragmentation from feeding with the glycine-based probe 81 over 72 hours.	146
Figure 113 - Precolibactin-969, 142, identified from $\Delta clbPQS$ <i>E. coli</i> strain, and the structure of proposed colibactin-645, 143, produced by the action of ClbP on 142.	147
Figure 114 – 144, a metabolite which accumulates in the ClbL expressing <i>E. coli</i> strain, but not in $\Delta clbL$ strains.....	147
Figure 115 - Action of amidase ClbL	148
Figure 116 - Structure of the genotoxic natural product colibactin, 147.	149
Figure 117 - Precolibactin-1489, 148, macrocycle derived from the cyclisation and oxidation of linear precolibactin.	150
Figure 118 - Biosynthetic pathway of colibactin accounting for all PKS, NRPS, and PKS-NRPS modules (2019).....	151
Figure 119 - The structure of glycopeptide antibiotic vancomycin, confirmed by X-ray diffraction in 1978.....	153
Figure 120 - Vancomycin binding the terminal D-alanyl-D-alanine motif of the mucopeptide through five hydrogen bonds.....	154
Figure 121 - Vancomycins inhibition of cell wall biosynthesis by preventing transglycosylation (a) and transpeptidation (b).....	155

Figure 122 - The altered binding of vancomycin to D-Ala-D-Lac in resistant strains	156
Figure 123 - The structure of glycopeptide antibiotic teicoplanin.....	156
Figure 124 - Heptapeptide derivatives from the glycosylation of glycopeptide aglycones with GtfE.....	157
Figure 125 - Functionalised vancomycin derivatives using GtfE to install an azido- sugar, followed by a click CuAAC reaction	158
Figure 126 - Generation of fluorobalhimycin through a mutasynthetic approach...	159
Figure 127 - Semisynthetic vancomycin derivatives, telavancin, 156, and oritavancin, 157.....	160
Figure 128 - Synthetic vancomycin analogue by Boger <i>et al.</i>	162
Figure 129 - The differing glycosylation patterns on the heptapeptide scaffold of the related compounds, vancomycin, chloroeremomycin and balhimycin.....	163
Figure 130 - The 67 kb gene cluster coding for the vancomycin biosynthetic enzymes and their assigned functions.	164
Figure 131 - Biosynthesis of Dhpg starting from malonyl-CoA	166
Figure 132 - Biosynthesis of Hpg beginning from either prephenate or tyrosine....	167
Figure 133 - Biosynthesis of β -hydroxytyrosine from L-tyrosine.....	167
Figure 134 - Methylation and glycosylation of the vancomycin aglycone after oxidative cross-linking and release from the NRPS.	168
Figure 135 - Sequential cross-linking of the NRPS-bound heptapeptide by P450 monooxygenases OxyB, OxyA, and OxyC.	170
Figure 136 - The replacement of the C domain of module 3 with the TE domain of module 7 causes the offloading of dipeptides.	172
Figure 137 – Proposed biosynthesis of vancomycin.....	173
Figure 138 - Vancomycin biosynthetic intermediates intercepted by chemical probes in this work.....	179
Figure 139 - Capture of leucine by the β -alanine based probe 180	182
Figure 140 - Leucine (starter unit of vancomycin biosynthesis) offloaded by a chemical probe based on β -hydroxytyrosine 178	183
Figure 141 - Offloading of leucine starter unit by the tyrosine-based probe 174....	184

Figure 142 - Offloading of a chlorinated dipeptide intermediate from the vancomycin NRPS by the Tyr probe 174	185
Figure 143 - Capture of the leucine starter unit from the vancomycin NRPS by the Gly probe 81	186
Figure 144 - Capture of a chlorinated dipeptide intermediate by the Gly probe 81	186
Figure 145 - Capture of a non-chlorinated dipeptide from the NRPS VpsA using a Hpg-based probe 178	189
Figure 146 - Capture of a chlorinated dipeptide intermediate from the vancomycin NRP pathway using Hpg probe 178	189
Figure 147 - Offloading of a chlorinated tripeptide intermediate from vancomycin biosynthesis using Hpg probe 178	190
Figure 148 - Structures of methylenomycin A and gaburedin A	194
Figure 149 - The biosynthetic gene cluster of scleric acid, with the putative proteins encoded by the genes.	194
Figure 150 - Structure of scleric acid	195
Figure 151 - Proposed structure of scleric acid, x, and its biosynthesis.	196
Figure 152 - The QncN, L, M pathway to convert xyulose-5-phosphate to a glycolicacyl-S-ACP	197
Figure 153 - Comparison of the ¹ H NMR spectra of authentic scleric acid (a), with the synthetic standard of the proposed structure (b).	198
Figure 154 - Revised biosynthetic pathway to scleric acid	199
Figure 155 - Comparison of the ¹ H NMR spectra of authentic scleric acid (top), and the synthetic standard (bottom)	201
Figure 156 - Restoration of production of scleric acid, 196, in $\Delta sclM4 \Delta sclN$ double mutant	203
Figure 157 - - Investigating scleric acid biosynthesis using a Pro-based probe.	205
Figure 158 - The reaction catalysed by nicotinamide <i>N</i> -methyltransferase	206
Figure 159 - Structures of the scleric acid derivatives synthesized to investigate the contribution of certain moieties to bioactivity.	207
Figure 160 – Intermediate, 232, offloaded by the thioamide glycine probe, 81.	211
Figure 161 - Amide isosteres: ketomethylene 233, urea 234, sulfonamide 235.	212

Figure 162 - The identical product that is formed from the offloading of either ClbI or ClbL by a chemical probe.....	217
Figure 163 - Putative intermediate, 237, offloaded by the fluoromalonate probe, 141.	218
Figure 164 - Ristomycin A, a glycopeptide produced by a gene cluster related to vancomycin	220
Figure 165 - Unchlorinated peptide offloaded using the Hpg-based probe, 178, and a tripeptide offloaded using the same chemical probe.	221
Figure 166 - Structures of kistamicin and complestatin, glycopeptide antibiotics which evolved separately to other GPAs	222

List of Tables

Table 1 - Genes involved in echinomycin biosynthesis and their assigned or putative functions.....	73
Table 2 - Amounts and concentrations of <i>S. lasaliensis</i> spores used to attempt to conjugate with <i>E. coli</i> ET12567 pUZ8002 pCRISPomyces-2 constructs.	96
Table 3 - The functions of the genes encoded by the colibactin BGC.....	116
Table 4 - Vancomycin biosynthetic intermediates captured by chemical probes, and their relative intensities.	181
Table 5 - Intermediates captured from the vancomycin biosynthetic pathway and their intensities	192
Table 6 - Table of bacterial strains used in this work.	303
Table 7 - Table of bacterial growth media used in this work. Autoclaved after addition of all components unless specified otherwise.....	305
Table 8 - Solvent gradient used to separate echinomycin extracts containing chemical probes.....	307
Table 9 - Solvent gradient used to separate colibactin extracts containing aminomalonate- and NRPS chemical probes, and vancomycin, and scleric acid extracts containing chemical probes.	308
Table 10 - Solvent gradient used to separate colibactin extracts containing the fluoromalonate chemical probe.....	308
Table 11 - Table of bacterial strains used for molecular biology and protein expression in this work.	310
Table 12 - Table of oligonucleotides (mostly PCR primers) used in this work.....	312
Table 13 - Table of plasmids used and generated in this work.....	314
Table 14 - Table of buffers and solutions used in this work. Not autoclaved unless specified.	315
Table 15 - Table of antibiotics used in this work.....	316
Table 16 - Table of PCR cycle conditions used with Q5 polymerase.	317
Table 17 - Annealing temperatures for genes cloned by PCR.	317

List of Schemes

Scheme 1 - a) Synthetic route to thioamide containing probes based on glycine, alanine, and valine. b) Failed thioamidation of Boc-protected compound, x.	82
Scheme 2 - - Synthetic route to <i>N</i> -methylated chemical probes based on glycine, alanine, and valine.....	87
Scheme 3 - a) Synthetic route to a cysteine based chemical probe. b) Attempted deprotection of <i>S</i> -Trt and <i>N</i> -Boc groups using HCl in dioxane.	90
Scheme 4 - Synthesis of glycine-based chemical probe, 81, and synthetic standard for C domains assays, 80.....	114
Scheme 5 - Synthetic route towards an aminomalonate probe protected with a pivaloyloxymethyl ester.	138
Scheme 6 - Synthetic route to aminomalonate probes protected by trifluoroethanol, 119, or butanol esters, 120, instead of the potentially reactive pivaloyloxymethyl ester.	140
Scheme 7 - Synthesis of aminomalonate probes protected by trifluoroethanol and butanol esters, containing an amide in place of the β -keto functionality.....	142
Scheme 8 - Synthesis of chemical probes for the vancomycin NRPS.....	175
Scheme 9 - Synthesis of β -hydroxytyrosine	176
Scheme 10 - Second synthetic route towards β -hydroxytyrosine proceeding through a benzylic hydroxylation	177
Scheme 11 - Synthesis of the initially proposed structure of scleric acid.....	198
Scheme 12 - Synthetic route to the synthetic standard of scleric acid	200
Scheme 13 - Synthetic route to (2-hydroxyacetyl)proline	202
Scheme 14 - Synthesis of an NRPS chemical probe based on L-proline.	204
Scheme 15 - Synthetic route to the D-scleric acid analogue.	207
Scheme 16 - Synthetic route to the L-amide derivative of scleric acid, 213.....	207
Scheme 17 - Synthetic route to the D-amide derivative of scleric acid, 214.....	208
Scheme 18 - Esterification of scleric acid with decan-1-ol.....	208
Scheme 19 - Synthetic route to the alkyne derivative, 216, by coupling with propargylamine, prepared by Rishi Chudasama.	209
Scheme 20 - Synthetic route towards the nicotinamide derivative of scleric acid...	209

Acknowledgements

Firstly, I would like to thank my supervisor, Manuela Tosin for giving me the chance to work on such an interesting and multi-disciplinary project. She has been relentlessly supportive and always encouraged me to push myself as a researcher. The guidance she has provided throughout my project has been invaluable, and I would not be where I am without her tireless work to improve my efforts.

My co-supervisor Christophe Corre and his group, particularly Fabrizio Alberti, have willingly shared their knowledge which has enabled large sections of my work, always with endless patience and a smile.

Collaborators Max Cryle and Anja Gruele at Monash University have provided the opportunity for some of the most satisfying work I've been involved in, and Alexander Brachmann at ETH Zurich has contributed a number of bacterial strains.

I am thankful to Cleidi Zampronio, Lijiang Song, and Ivan Prokes for all their help in acquiring mass spectra and NMR data of the work I have carried out.

I would like to thank the members of the Tosin group for creating such an enjoyable environment to spend the last three years and being a great sounding board for ideas. Particularly I would like to thank Rob Jenkins for being a source of synthetic chemistry advice and entertainment, and Pan Prasongpholchai for her friendship. I'm thankful to Candace Ho, Pamela Banana-Dube and Ryan Packer for their help with biological work. I would also like to thank Rachel Kerr for keeping me sane, for which I am endlessly grateful, and her provision of doughnuts. I'm also grateful to the members of CBRF for making such an easy environment and readily sharing their time and expertise.

Outside of university, I'm grateful to Kenilworth Runners, especially Emma Ford, for her friendship, conversation, and encouragement. I'd also like to thank all the boulderers of the lab for making falling off walls such a fun way to spend time.

Finally, I'd like to thank my family for all their support and encouragement without which this would not have been possible.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

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

Data used to generate Figure 145 was obtained in collaboration with Dr Anja Gruele, Monash University

Analysis of the scleric acid biosynthetic gene cluster was conducted by Dr Christophe Corre and Dr Fabrizio Alberti

TAR capture and mutagenesis of the scleric acid cluster was conducted by Dr Fabrizio Alberti

Antibacterial testing of scleric acid was conducted by Warwick Antimicrobial Screening Facility

Biological testing of scleric acid was conducted by Eli Lilly Open Innovation Drug Discovery

Antitubercular testing of scleric acid derivatives was conducted by Dr Vinayak Singh, University of Cape Town

Parts of this thesis have been published by the author:

Y. T. C. Ho, D. J. Leng, F. Ghiringhelli, I. Wilkening, D. P. Bushell, O. Kostner, E. Riva, J. Havemann, D. Passarella and M. Tosin, *Chem. Commun.*, 2017, **53**, 7088–7091.

F. Alberti, D. J. Leng, I. Wilkening, L. Song, M. Tosin and C. Corre, *Chem. Sci.*, 2019, **10**, 453–463.

Abstract

Nonribosomal peptides (NRPs) are secondary metabolites produced by nonribosomal peptide synthetases and comprise countless medicinally relevant compounds, such as the penicillin antibiotics and the anticancer agent bleomycin, as well as toxins such as microcystins. NRP biosynthesis involves the processing of enzyme-bound intermediates, which is challenging and constitutes a bottleneck in the stepwise elucidation of natural product maturation. Chemical probes mimicking natural biosynthetic extender units have been developed and used as effective tools to intercept and capture biosynthetic intermediates for identification and characterisation. This PhD work describes the use of chemical probes to investigate the NRP bio-assembly of echinomycin, colibactin, vancomycin, and scleric acid.

Echinomycin is a depsipeptide antibiotic produced by *Streptomyces lasaliensis*. Chemical probes containing a thioamide or *N*-methylated amino acid moieties inspected the substrate scope of the condensation domain, and the potential to generate unnatural peptides by NRPSs. To this end, *in vivo* editing of the gene cluster and *in vivo* reconstitution of the NRPS have also been pursued.

Colibactin is a PKS-NRPS derived genotoxin produced by *E. coli*. The biosynthetic pathway features several rare enzymatic processes, including those catalysed by novel amidase ClbL and the use of two aminomalonate extender units. Chemical probes designed to mimic aminomalonate and cysteine were synthesised and tested to explore the formation of thiazole rings and the incorporation of aminomalonate by the PKS ClbO in colibactin biosynthesis.

Vancomycin is a glycopeptide antibiotic of last resort produced by *Amycolatopsis orientalis*. A number of biosynthetic questions remain unanswered, particularly the timing of the chlorination by the halogenase VhaA. Chemical probes were used to examine timing of chlorine incorporation, with a likely timing proposed, and the substrate specificity of the pathways C domains.

Scleric acid is the product of an NRPS-containing, silent, cryptic gene cluster. Chemical synthesis has been used to unequivocally assign the structure, and the

participation of a key biosynthetic intermediate. A number of analogues have also been synthesized to develop modest antitubercular activity exhibited by scleric acid.

Overall, in this work the chemical probing of NRPSs has provided insights into the substrate tolerance of the condensation domains, offering insights which can be used in the engineering of the NRPS pathways. Intermediates captured from the vancomycin pathway support the chlorination of an enzyme-bound intermediate, and the structure of scleric acid has been identified and key parts of its biosynthesis revealed.

Abbreviations

5-FOA	5-fluoroorotic acid
A domain	Adenylation domain
<i>A. orientalis</i>	<i>Amycolatopsis orientalis</i>
A ₆₂₀	Absorbance at 620 nm
ACC	1-aminocyclopropanecarboxylic acid
AcCl	Acetyl chloride
AcOH	Acetic acid
ACP	Acyl carrier protein
AIBN	Azobisisobutyronitrile
Ala	Alanine
AM	Aminomalonate
AMP	Adenosine monophosphate
Amp	Ampicillin
Amp ^R	Ampicillin resistant
AMS	Adenosine monosulfamate
Apra	Apramycin
Apra ^R	Apramycin resistant
APS	Ammonium persulfate
ArCP	Aryl carrier protein
Asn	Asparagine
Asp	Aspartic acid
AT	Acyltransferase
ATP	Adenosine triphosphate
AVS	Adenosine vinyl sulfonamide
BAC	Bacterial artificial chromosome
BGC	Biosynthetic gene cluster
BnBr	Benzyl bromide
BnOH	Benzyl alcohol
Boc	<i>tert</i> -butyloxycarbonyl
C domain	Condensation domain
CAD	Collisionally-activated dissociation
CDA	Calcium dependent antibiotic
CDI	1,1'-Carbonyldiimidazole
Cm	Chloramphenicol
Cm ^R	Chloramphenicol resistant
CoA	Coenzyme A
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CuAAC	Copper(I)-catalysed azide-alkyne cycloaddition
Cy domain	Cyclisation domain

Cys	Cysteine
Da	Dalton
DAHP	3-deoxy-D-arabino-heptulosonate-7-phosphate
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCC	N,N'-dicyclohexylcarbodiimide
DCM	Dichloromethane
DEPBT	(3-(diethoxyphosphoryloxy)-1,2,3-benzotriazin-4(3H)-one)
DH	Dehydratase
dH ₂ O	Distilled water
Dhpg	L-3,5-dihydroxyphenylglycine
DiPaC	Direct Pathway Cloning
DIPEA	N,N-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
DSB	Double-strand break
DTT	Dithiothreitol
E domain	Epimerase domain
<i>E. coli</i>	<i>Escherichia coli</i>
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetracetic acid
ER	Enoylreductase
ESI	Electrospray ionisation
FAD	Flavin adenine dinucleotide
FTMS	Fourier Transform mass spectrometry
GlcNAc	N-acetylglucosamine
Glu	Glutamic acid
Gly	Glycine
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HCD	Higher-energy collisional dissociation
HDR	Homology directed repair
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIF-1	Hypoxia-Inducible Factor 1
His	Histidine
HIV	Human immunodeficiency virus
HMWP1	High molecular weight protein 1
HMWP2	High molecular weight protein 2
HOBt	Hydroxybenzotriazole

Hpg	L-4-hydroxyphenylglycine
HPLC	High performance liquid chromatography
HRE	Hypoxia-responsive element
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IRMPD	Infrared multiphoton dissociation
Kan	Kanamycin
Kan ^R	Kanamycin resistant
kb	kilobase
KR	Ketoreductase
KS	Ketosynthase
Lac	Lactate
LB	Luria broth
LC-MS	Liquid chromatography mass spectrometry
Leu	Leucine
LTR	Long terminal repeat
Lys	Lysine
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
MAGE	Multiplex automated genome editing
MALDI	Matrix assisted laser desorption ionisation
Met	Methionine
MIC	Minimum inhibitory concentration
MNAN	1-methylnicotinamide
mRNA	Messenger RNA
MRSA	Methicillin-resistant <i>Staphylococcus aureus</i>
MRSE	Methicillin-resistant <i>Staphylococcus epidermis</i>
MS	Mass spectrometry
MS ²	Tandem mass spectrometry
MSSA	Methicillin-sensitive <i>Staphylococcus aureus</i>
MT domain	Methyltransferase domain
MW	Molecular weight
NAc	<i>N</i> -acetyl
Nal	Nalidixic acid
NCBI	National Centre for Biotechnology Information
NDP	Nucleotide diphosphate
NHEJ	Non-homologous end joining
NMR	Nuclear magnetic resonance
NNMT	Nicotinamide <i>N</i> -methyltransferase
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetase
NTA	Nitriloacetic acid

OASIS	Orthogonal active site identification system
OD ₆₀₀	Optical density at 600 nm
OIDD	Open innovation drug discovery
Ox domain	Oxidation domain
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
Phe	Phenylalanine
P _i	Phosphate
PKS	Polyketide synthase
PPant	Phosphopantetheinyl
PP _i	Inorganic pyrophosphate
PPTase	Phosphopantetheinyl transferase
PRISM	Proteomic investigation of secondary metabolites
Q2CA	Quinoxaline-2-carboxylic acid
R domain	Reductase domain
RNA	Ribonucleic acid
RP	Reverse-phase
rpm	Revolutions per minute
<i>S. albus</i>	<i>Streptomyces albus</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
<i>S. coelicolor</i>	<i>Streptomyces coelicolor</i>
<i>S. echinatus</i>	<i>Streptomyces echinatus</i>
<i>S. lasaliensis</i>	<i>Streptomyces lasaliensis</i>
<i>S. lividans</i>	<i>Streptomyces lividans</i>
<i>S. pyogenes</i>	<i>Streptococcus pyogenes</i>
<i>S. sclerotialus</i>	<i>Streptomyces sclerotialus</i>
<i>S. venezuelae</i>	<i>Streptomyces venezuelae</i>
SAH	(S)-adenyl-L-homocysteine
SAM	(S)-adenosyl methionine
<i>scl</i>	Scleric acid BGC
SDS	Sodium dodecyl sulfate
Ser	Serine
SFM	Soya flour mannitol
sgRNA	Single guide RNA
SNAC	N-acetylcysteamine
STE	Sodium chloride-Tris-EDTA
T domain	Thiolation domain
TAR	Transformation-assisted recombination

TB	Tuberculosis
TBE	Tris-Borate-EDTA
TCEP	tris(2-carboxylethyl)phosphine
TE	Tris-EDTA
TE domain	Thioesterase domain
TEMED	Tetramethylethylenediamine
TEMPO	(2,2,6,6-Tetramethylpiperidin-1-yl)oxyl
TES	Triethylsilane
TES	2-[[1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]amino]ethanesulfonic acid
TFA	Trifluoroacetic acid
TFE	Trifluoroethanol
TH	Thioester hydrolase
THF	Tetrahydrofuran
TIPS	Triisopropylsilane
TLC	Thin layer chromatography
TMU	Tetramethylurea
Trityl	Triphenylmethyl
tRNA	Transfer RNA
TSB	Tryptic soy broth
Tyr	Tyrosine
UDP	Uridine diphosphate
UHR-TOF	Ultra high-resolution time of flight
UV	Ultraviolet light
UV-Vis	Ultraviolet-visible light
VRE	Vancomycin-resistant <i>Enterococci</i>
YIPP	Yeast inorganic pyrophosphatase
βHT	β-hydroxytyrosine
NBS	<i>N</i> -bromosuccinimide

1. Introduction

1.1 Natural products

Natural products are a broad class of small molecules derived from a biological source,¹ such as bacteria, fungi, and plants, and are often more specifically defined as secondary metabolites. These differ from products of primary metabolism in that they are not essential to the survival of the producing organism but are believed to offer a competitive advantage. This proposal is supported by the biological activity many of

these compounds have, as agents of chemical warfare, or roles such as metal transportation. The tight binding of natural products to biological targets can be demonstrated by the crystal structure of the immunosuppressant natural product rapamycin, **1**, in complex with its mammalian target FKBP12 and mTor, shown in Figure 1.²

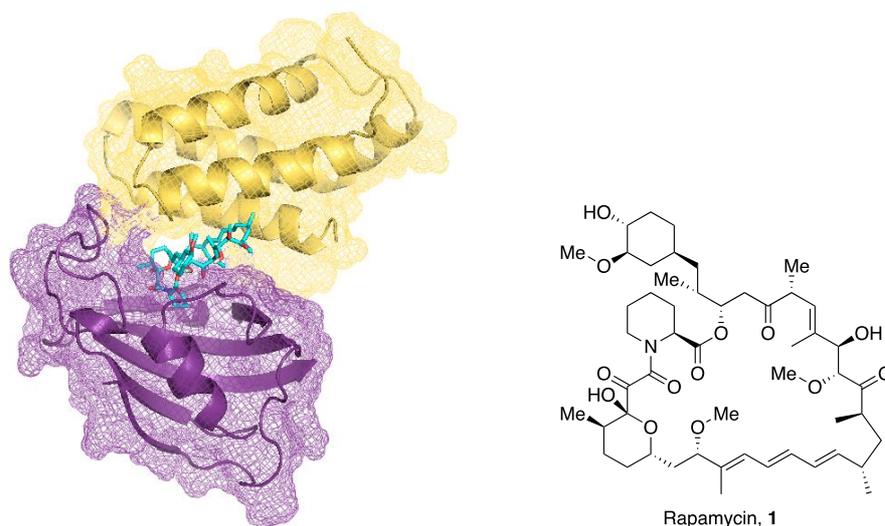


Figure 1 - Crystal structure of rapamycin bound to FKBP12 and mTor, and the chemical structure of rapamycin.

The potent action of natural products on biological targets has led to their use for human benefit for millennia, with the earliest recorded use being 2600 BC. Discovery of the fungal natural product penicillin by Alexander Fleming ushered in the era of antibiotics,³ and it is estimated that almost three-quarters of antibacterial compounds are either natural products or derivatives thereof. Furthermore, over 80% of small molecule anticancer agents developed between 1981 and 2014 were natural products or derivatives.⁴

The wide range of biological activity is made possible by the great chemical diversity of natural products illustrated in Figure 2, with alkaloids, terpenoids, polyketides and nonribosomal peptides some of the largest classes.

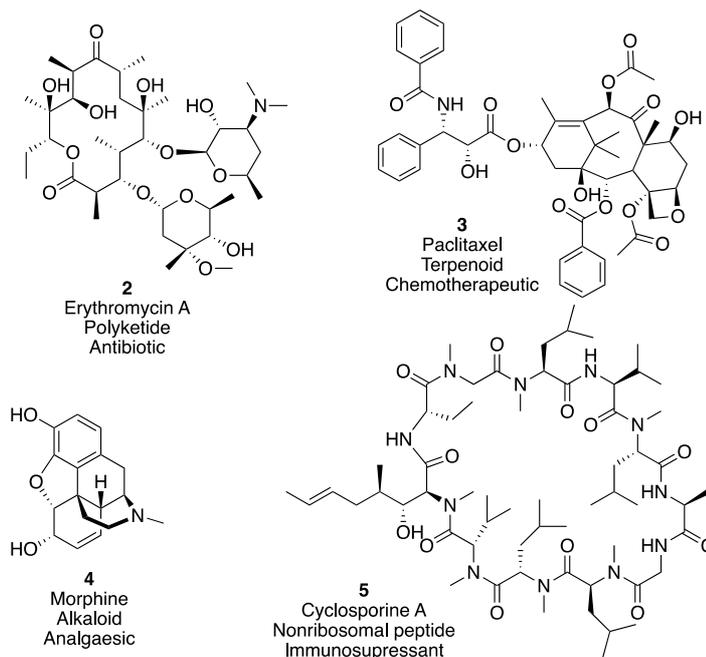


Figure 2 - Examples of the natural product families: alkaloids, terpenoids, polyketides, and nonribosomal peptides.

The utility of these compounds has led to growing demand for novel analogues, with improved specificity, pharmacokinetics, or solubility.⁵ Particularly in the context of growing antimicrobial resistance, there is a need for new antibiotics and derivatives of existing classes.⁶ For structurally complex natural products, engineering of natural product biosynthetic pathways such as nonribosomal peptide synthetases (NRPSs) are seen as a promising route to new bioactive compounds.

1.2 Nonribosomal peptides

Nonribosomal peptides (NRPs) are usually discovered from bacterial or fungal sources, however the dipeptide produced by the nonribosomal peptide synthetase Ebony in *Drosophila* demonstrates an example outside these organisms.⁷ NRPs are formed by complex, multifunctional megaenzymes, the nonribosomal peptide synthetases (NRPSs), which catalyse a large number of distinct chemical reactions to produce natural products of diverse structure and function.

The class of nonribosomal peptides were first discovered when it was observed that the biosynthesis of gramicidin S, **6**, and tyrocidine, **7**, shown in Figure 3, in extracts of *Bacillus brevis* was unaffected by the presence of the ribosome inhibitor

chloramphenicol, or RNAase. With translation not possible in this system, it was hypothesized a new mechanism must be responsible for these cyclic peptides.^{8,9}

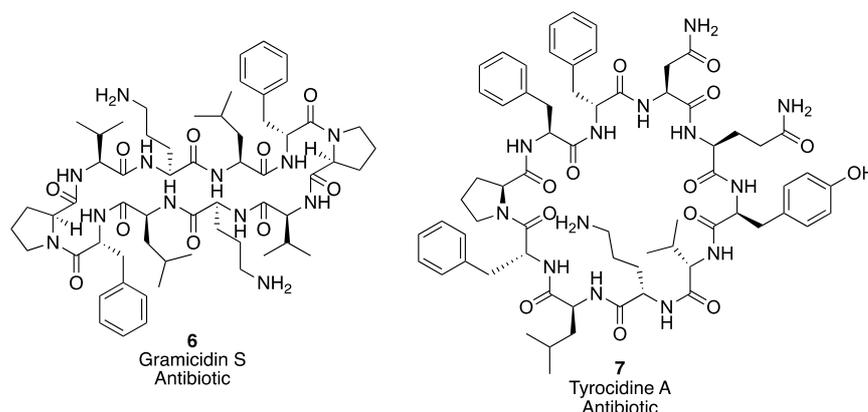


Figure 3 - Gramicidin S and Tyrocidine A, NRPs used in early investigations of biosynthesis.

Further work using partially purified enzymes of the gramicidin S synthetase showed an ATP-dependent incorporation of amino acids in a two-step process. First, the amino acid is activated to the amino-adenylate with the release of inorganic pyrophosphate (PP_i), and then transfer of the amino acid to a thiol carrier, with release of adenosine monophosphate (AMP).¹⁰ The nature of the amino acid-enzyme covalent linkage was determined through chemical degradation of the bond, and its notable stability to trichloroacetic acid.¹¹ The sequential addition of amino acids by these enzymes could also be followed through the use of radiolabelled ¹⁴C-L-Phenylalanine, and confirmed that the growing peptide chain remained enzyme bound throughout.¹² The nature of this covalent link was established by the discovery of the 4'-phosphopantethiene (PPant) cofactor on the synthetases involved in gramicidin S and tyrocidine biosynthesis by the groups of Laland and Lipmann.^{13,14} It was noted that there was around 70-75 kDa of protein per amino acid activated by the enzyme, leading to a modular hypothesis.¹⁵ This was supported by newly emerging DNA sequence data, which noted repeated protein sequences equivalent to the number of amino acid activated by the NRPS.^{16,17} This allowed the refinement of the carrier model from the proposed single PPant moiety per NRPS, to a multi-carrier model where each amino acid activated is attached by a PPant arm to the enzyme. The amino acids can then undergo amide bond formation, extending a growing peptide chain while remaining bound to the NRPS.

1.3 NRPS modules

The modular organization of NRPSs shows that each module is responsible for the incorporation of a single amino acid into the peptide chain. The order of modules often, but not always, dictates the order of incorporation of amino acids, according to the principle of colinearity.¹⁸ A minimal module responsible for the addition of one amino acid consists of three domains: an adenylation (A) domain, a thiolation / peptidyl carrier protein domain (T), and a condensation domain (C).¹⁹ Further modifications can be introduced by methyltransferase (MT), epimerization (E), and cyclisation (Cy) domains, amongst others. The peptide chain is then normally released by the action of a thioesterase domain (TE) in the final module.

1.3.1 Adenylation domains

As mentioned, adenylation domains catalyse two chemical reactions. Firstly, the activation of the carboxylic acid functionality of the amino acid, or other substrate, to form the aminoacyl-adenylate with the release of PP_i, followed by nucleophilic attack of the thiol of the PPant arm of the T domain to yield a thioester with the associated release of AMP, shown in Figure 4.

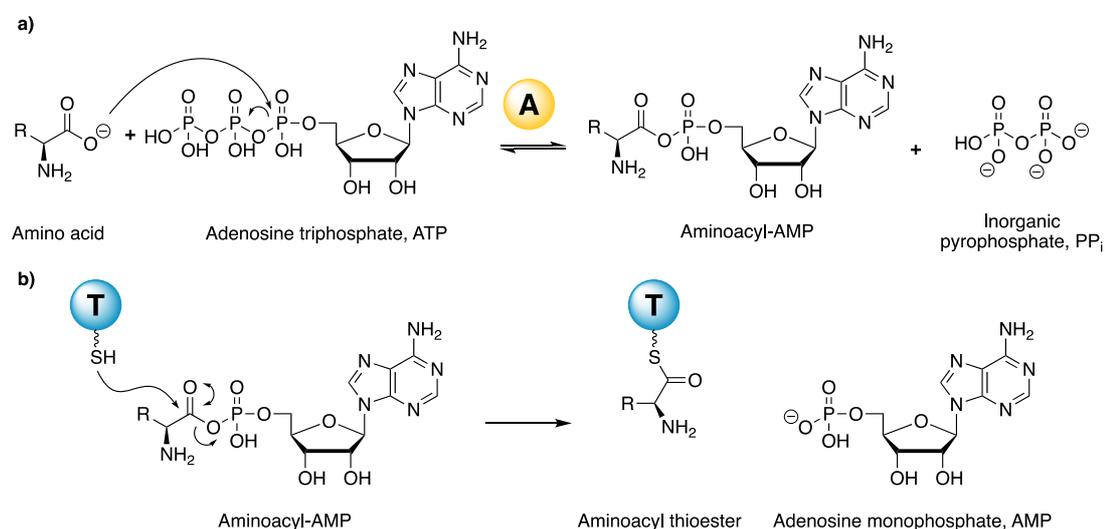


Figure 4 - a) The adenylation of an amino acid by reaction with ATP to form a reactive aminoacyl-AMP.

b) The reaction of the aminoacyl-AMP with the 4'-PPant thioester of the T domain to form an aminoacyl thioester.

A large amount of the variety of NRPs is drawn from the A domains ability to activate a massive range of substrates, with over 500 currently known.²⁰ In comparison, the ribosome uses a similar ATP-dependent activation of amino acids, but is able to utilise just the 20 proteinogenic amino acids with tRNA.²¹ Each A domain is also specific for the activation of its substrate, acting as the primary gatekeeper in the synthesis of NRPs.²² However, some natural flexibility is inherent in a number of NRPSs, possibly due to the lower cost of error compared to ribosomal peptide synthesis.^{23,24}

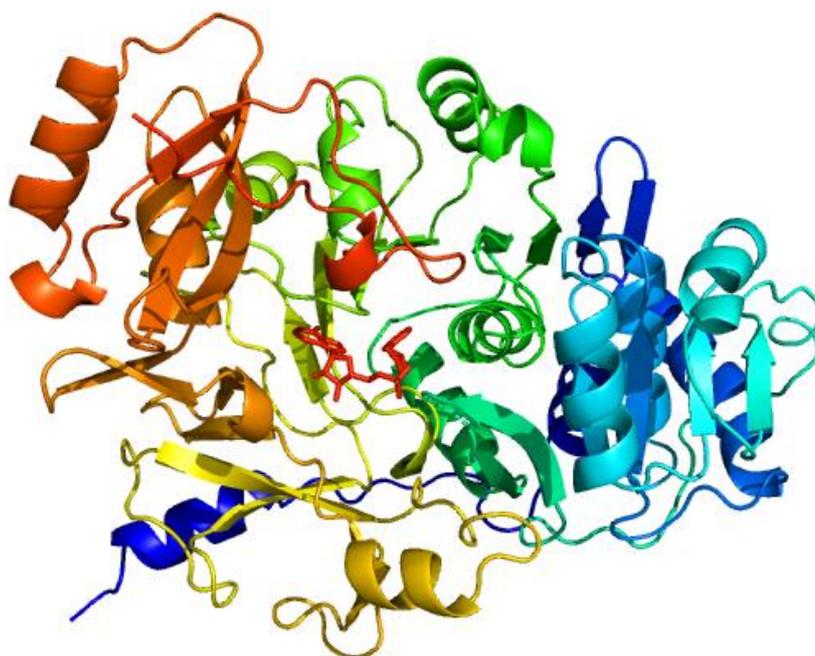


Figure 5 - Structure of PheA (GrsA) co-crystallized with phenylalanine and AMP.²⁵

The crystal structure of the gramicidin S synthetase GrsA, also known as PheA, in complex with the substrate phenylalanine and AMP, shown in Figure 5, gave valuable insight into the determinants of substrate selectivity by the A domain.²⁵ These include the conserved residues Asp235, which hydrogen bonds with the α -amino group, and Lys517, which stabilizes the α -carboxylate through electrostatic interactions. The crystal structure of DhbE, an A domain from bacillibactin biosynthesis which activates 2,3-dihydroxybenzoate, extended this to aryl acid activating domains. In these enzymes, the conserved aspartic acid residue is no longer present, allowing discrimination between aryl and amino acids.²⁶

The high similarity of NRPS A domains allowed conserved sequences in the substrate binding pocket to be related to the amino acids activated by the domains. This allowed the identification of 10 key residues which determine the shape and charge of binding pocket, and which could be used to predict A domain specificity.^{27,28} This has since been extended to use in bioinformatic tools which are able to predict A domain specificity in uncharacterized putative NRPS biosynthetic gene clusters (BGCs) using just genome sequence data.^{29,30}

Structural biology data has also been used to reveal important aspects of A domain function. The A domain consists of a large N-terminal subdomain and a smaller C-terminal subdomain. For the two reactions catalysed by the A domain, it has been shown that the C-terminal subdomain and the T domain associated need to move to facilitate the interaction, with a large rotation of the C-terminal subdomain to present different faces of the protein to the active site that carries out both reactions.³¹ This is proposed to guide the PPant arm of the carrier protein into the active site.³² The impact of protein-protein interactions on the A domain has also been demonstrated by the recent use of multi-domain constructs by the groups of DiVentura and Dittmann. In microcystin biosynthesis it was shown that using a C-A-T construct greatly enhanced the specificity of activation of amino acids by the NRPS compared to the A-T construct.³³ Extending this to modules, it was shown that module 10 of the NRPS responsible for tyrocidine could activate leucine when expressed as a single A-T module, but when combined with the downstream module 9, as an A-T-C-A-T construct, leucine could only be activated in the presence of ornithine, the substrate of module 9. It is suggested some A domains may not repeatedly activate their substrate in the absence of condensation domain activity.³⁴

1.3.2 Thiolation domain

The role of the thiolation domain was a key idea of the multiple-carrier model of nonribosomal peptide biosynthesis. The identification of the pantetheine arm by which peptidyl chains are bound to the enzyme was analogous to the known role of acyl carrier proteins (ACPs) in fatty acid biosynthesis. Like ACPs, NRPS T domains are post-translationally modified by a 4'-phosphopantetheine on a conserved seryl

residue within the (L/I)GxDS(L/I) motif.³⁵ Without this modification the carrier protein, and thus the NRPS, is inactive.

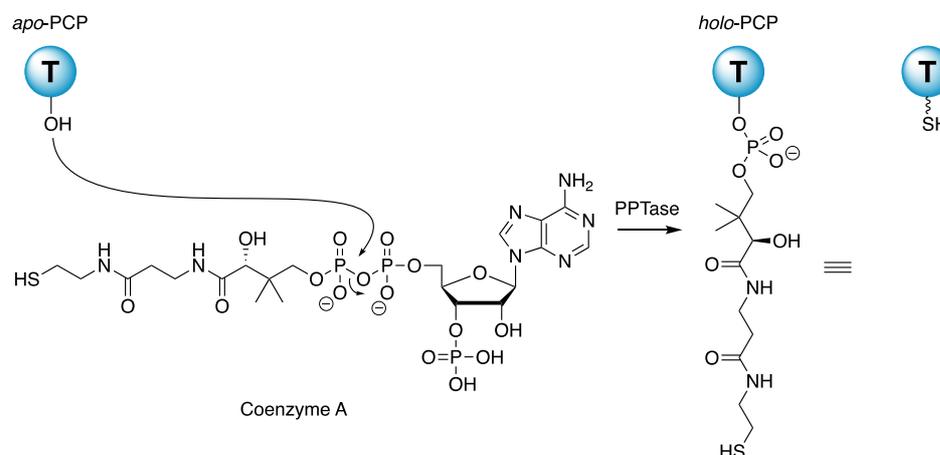


Figure 6 - Loading of the active site serine of thiolation (T) domains using coenzyme A by a PPTase to give holo-ACP, with the 4'-PPant cofactor represented by a wavy line (left).

This modification is catalysed by a phosphopantetheinyl transferase (PPTase) which uses coenzyme A as a substrate, joining the pantetheine arm to the active site serine through a phosphodiester linkage, as shown in Figure 6; this converts the carrier protein from an *apo* to a *holo* state. PPTases can possess strict substrate specificity, such as the *Escherichia coli* protein holo-acyl carrier protein synthase (ACPS) which will phosphopantetheinylate acyl carrier proteins, but not aryl or peptidyl carrier proteins.³⁶

However, despite possessing the same 4 helical bundle as ACPs, T domains have little sequence homology outside the cofactor binding site.³⁷ The length and alignment of helices differ from those observed in acyl carrier proteins, and can undergo conformational change to present different residues of the helices to up or downstream enzymes for specific protein-protein interactions.³⁸ This provides a rationale for the specific interaction of PPTases with the *apo* form of the carrier protein. This conformational switch moves the PPant cofactor 100 Å across the face of the T domain, and move the PPants thiol group around 16 Å.³⁹ Combined with the 20 Å length of the phosphopantetheinyl arm, this range of motion allows the T domain to shuttle substrates between all the domains required to produce an NRP while keeping the substrates enzyme-bound.

1.3.3 Condensation domain

The condensation domain of an NRPS is responsible for the formation of amide bonds between two substrates bound to adjacent T domains, with the concomitant release of water. They are present in every NRPS module that catalyses chain extension and are thus part of a minimal module.

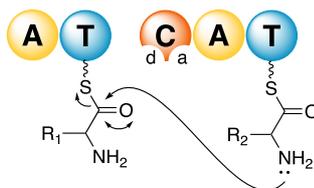


Figure 7 - Amide bond formation catalysed by the condensation domain with an amino acid at the acceptor site (a), and the growing peptide chain at the donor site (d).

In spite of this crucial role, the exact mechanism by which the condensation domain acts is still not well understood. Comparison of the primary sequences of NRPSs allowed the identification of a 350 amino acid region responsible for catalysing elongation, and of a conserved HHxxxDG motif.⁴⁰ It was noted that there was an almost invariant distance between this motif and the attachment site of the PPant cofactor. Direct evidence for this proposal came when mutation of the second histidine in the motif in the NRPS GrsB abolished the peptide bond forming activity.⁴¹ It was later shown that by mutation of 10 strictly conserved residues in GrsB that R62 and D151 are also essential for activity, with histidine suggested as a general base, and aspartate and arginine forming a stabilising salt bridge. R67, H146, and W202 led to insoluble proteins, and were deemed to be important to the overall protein structure.⁴² However mutations of histidine in C domains within the NRPS which produces the siderophore vibriobactin showed that VibH and VibF-C2 suffered only modest reductions in activity, suggesting general base catalysis may not be a universal mechanism.^{43,44} Crystal structures of C domains aided greatly in understanding the catalytic mechanism, and while no substrates could be co-crystallized, a tunnel could be observed which was of sufficient length and diameter to accommodate two substrates. The tunnel emerged on opposite sides of the C domain, allowing access to

both the pantetheinyl bound amino acids. Modelling also showed the nucleophilic α -amino group 3 Å from the catalytic histidine.^{44,45} The Schmeing group used site-directed mutagenesis to insert a cysteinyl residue into the tunnel between the acceptor site and the histidine motif, and alkylated this Cys using an alkyl bromide to tether a reaction-competent substrate mimic to the enzyme.⁴⁶ This allowed the obtainment of a crystal structure showing the conserved His156 accepting a hydrogen bond from the α -amino group of the substrate mimic. It was proposed that this orients the amino group for nucleophilic attack and catalysis takes place primarily through substrate positioning, in a mechanism analogous to the ribosome.

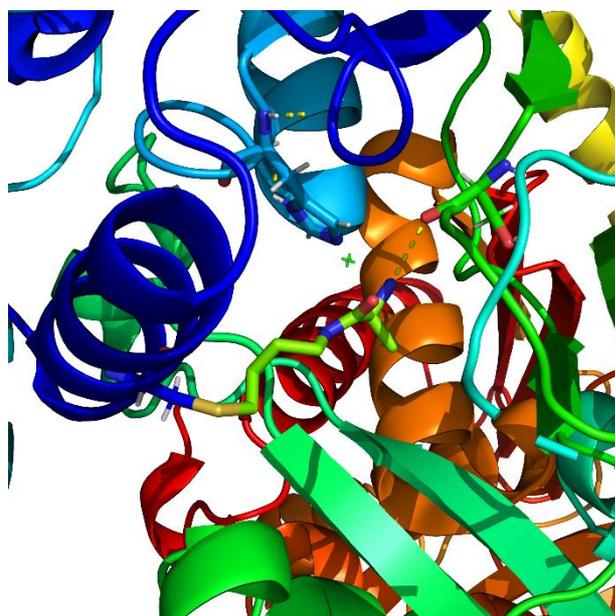


Figure 8 - Crystal structure of CDA-C1(E17C) with covalent substrate mimic.⁴⁶

In addition to the crucial bond-forming reaction, C domains can also act as secondary gatekeepers to ensure the fidelity of produced NRPs. Condensation domains possess two substrate binding sites: an acceptor site which binds the pantetheine bound amino acid which will accept the growing peptide chain, and a donor site which binds the chain that will be extended and passed downstream. Studies using non-cognate substrates in the form of either enzyme-bound amino acids, or aminoacyl-*N*-acetylcysteamine thioesters (aminoacyl-SNACs), showed that the sites of C domains appear to be selective, exerting a preference for the stereochemistry of the upstream and downstream amino acids. Acceptor sites have been shown to also

have preference for the side chain of the amino acid.^{47,48} Binding of the acceptor amino acid to an enantioselective binding pocket in the acceptor site of the C domain until reaction with an upstream peptide chain has been proposed as the mechanism by which NRPSs control directionality of peptide synthesis, and prevent mis-initiation of peptide synthesis by downstream modules.⁴⁹ The crystal structure of a substrate mimic at the acceptor site of CDA-C1 of calcium-dependent antibiotic NRPS allowed rational mutagenesis to alter acceptor site specificity to a degree, but also suggest that a specificity code similar to that of A domains may not be possible.⁴⁶

C domains can thus be grouped based on the stereochemical preference of the donor and acceptor sites into ^DC_L and ^LC_L clades. ^DC_L domains catalyse amide bond formation between a chain with a C-terminal D-configured residue and an L-amino acyl thioester, while ^LC_L catalyses condensation between two L-amino acids. Phylogenetic analysis also reveals a third C domain subtype, the starter C domain. This occurs exclusively as the first C domain of an NRPS and catalyses the reaction between a β-hydroxy carboxylic acid and an amino acid.⁵⁰

The stereochemical specificity of C domains is not however absolute. Notably it has been shown that the C domain of module 7 in teicoplanin biosynthesis will accept the incorrect D-amino acid in place of the natural L-tyrosine substrate. The phylogenetic analysis of Rausch *et. al.* showed that this C domain clusters with ^DC_L domains, leading to the hypothesis that this ^DC_L domain has evolved towards acceptance of an L-configured C-terminus, without losing activity towards D-configured peptides.⁵¹

1.3.4 Cyclisation domain

Cyclisation (Cy) domains are related to condensation domains, forming part of the same enzyme superfamily. They too catalyse the formation of an amide bond between acceptor and donor substrate, but additionally they catalyse the cyclodehydration of the thiol or hydroxyl moieties of cysteine, serine or threonine.⁵² The condensation proceeds in an analogous manner to that of the condensation domain, followed by a two-step cyclodehydration, with attack of the donor carbonyl oxygen by the nucleophilic hydroxyl or thiol of the acceptor amino acid. The formed

hydroxyl group is then eliminated in a dehydration reaction, forming the carbon-nitrogen double bond as shown in figure 9.⁵³ This forms the thiazoline, oxazoline or methyloxazoline heterocycle functionalities or their reduced forms that are relatively common in NRPs, and important for bioactivity in a number of natural products such as bleomycin.⁵⁴ Following the action of Cy domains, the heterocycles can then be further modified by oxidation (Ox) or reduction (R) domains either embedded in the assembly line or acting in *trans*.⁵⁵

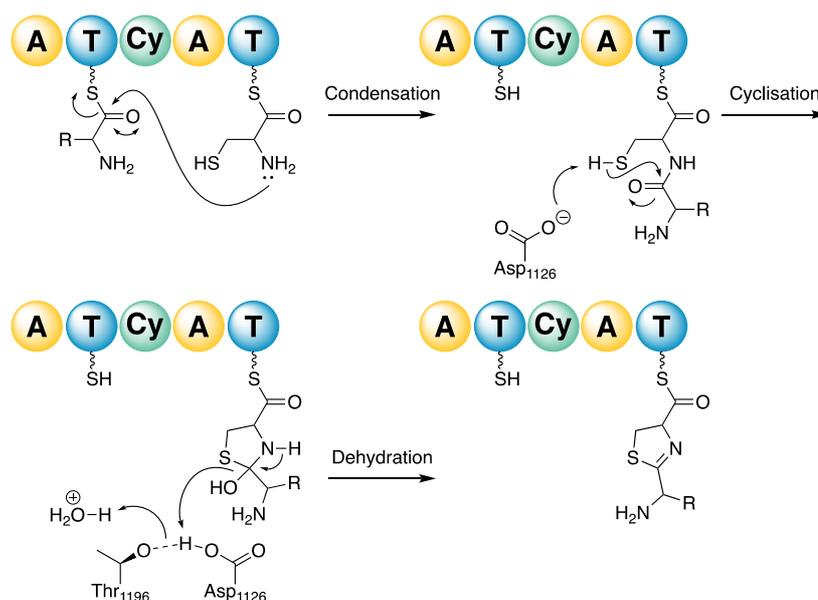


Figure 9 - Representative thiazole formation catalysed by the cyclisation domain: Condensation, nucleophilic attack of the side chain residue to form a heterocycle, and dehydration to form a thiazoline.

While mutagenesis studies showed that the condensation and cyclodehydration steps were independent and must involve some residues with a single function, the process by which these reactions occurred was unknown. The mechanism is different from that of the condensation domain in that they lack the conserved HHxxxDG motif, but instead have a DxxxxD motif.⁵⁶ Mutations to these aspartic acid residues show they are crucial to condensation and cyclisation. Analysis of vibriobactin NRPS VibF revealed the second Cy domain catalysed only amide formation, and comparison of primary sequence with other T Cy domains revealed nine amino acids which were conserved across Cy domains, but not VibH Cy2. Introducing these mutations at these amino acids into the NRPS which produces bacitracin, a thiazole containing natural

product, revealed that N900A and S984A mutants had condensation activity, but could not catalyse cyclisation.

The first crystal structure of a Cy domain, that of bacillamide synthetase, BmdB-Cy2, as well as mutagenesis of residues in proximity to the active site, allowed the determination of the roles of the conserved aspartate residues.⁵⁷ This revealed the aspartates as structurally, but not catalytically, important and that Cy domains likely rely on substrate positioning for catalysis. Additionally, conserved Cy motifs PVVFTS and SQTPQVxLD, which are spatially near the active site, contain essential threonine and aspartate residues. The aspartate is proposed to remove a proton from the acceptor amino acids nucleophilic side chain, and the threonine may donate a proton to form the hydroxyl-thiazolidine, before protonation of the hydroxyl by the aspartic acid allows the loss of water.

1.3.5 Epimerisation domain

Epimerisation (E) domains, like cyclisation domains, are also part of the condensation domain superfamily. These domains are responsible for converting L-amino acids to non-proteinogenic D-amino acids, and thereby introduce a large amount of structural diversity. Incorporation of D-amino acids allows NRPs to access additional conformation space, and also confer resistance to proteases.⁵⁸ The gramicidin S synthetase, GrsA, was one of the first examples of this epimerase activity, and by production of truncated constructs the domain responsible for epimerisation could be identified.⁵⁹ E domains are inserted after the T domain in the upstream module, and before the downstream C domain, giving a C-A-T-E-C-A-T domain architecture. E domains in starter units act upon the aminoacyl-thioester, but downstream E domains utilise the peptidyl-thioester as the substrate.⁶⁰

The reaction occurs *via* a deprotonation mechanism, acting up the thioesterified amino acid or peptidyl chain, likely with a resonance-stabilized enolate intermediate as shown in Figure 10.

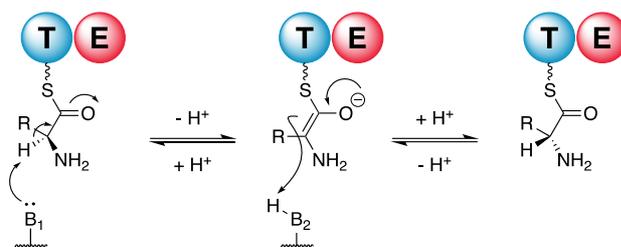


Figure 10 - General mechanism of epimerisation to convert L-amino acids to D-amino acids.

E domains share a similar motif to C domains, with a conserved HHxxxD, but the glycine after aspartic acid is less conserved. Mutagenesis experiments in the NRPS GrsA showed that the second His residue and the Asp, along with Y976, are essential for epimerase activity. Conversion of L-Phe to D-Phe on GrsA reached equilibrium at 1:1.9, suggesting a slight stabilisation of the D-epimer in the active site.⁶¹ The crystal structure of the E domain of TycA shows the channel found in C domains is blocked on the acceptor side in E domains, only allowing interaction with the donor carrier protein.⁶² Based on this model, it has also been suggested that the conserved Glu882 may form a catalytic dyad with His743, but the exact mechanism of deprotonation in epimerisation domains remains unclear.

Additionally, a dual epimerisation/condensation (E/C) domain has been reported in the biosynthesis of lipopeptide arthrofactin. The natural product contains 6 D-amino acids, but none of the NRPS modules contain typical E domains. The A domains in these modules activate L-amino acids which are epimerised and then condensed by the dual E/C domain. These domains contain a HHxxxDG motif as in C domains, but an additional HH(I/L)xxxxDG motif towards the N-terminus of the domain.⁶³

1.3.6 X domain

The other non-essential NRPS domain related to the condensation superfamily is the X domain. This domain is specific to the NRPSs of the glycopeptide antibiotic family of natural products. These all possess heavily cross-linked structures, rigidified by aryl-ether bonds installed by cytochrome P450 monooxygenases. The X domains, present in the final module of the NRPSs, lack the HHxxxDG motif of C domains, and also lack a substrate binding tunnel. The function of these degraded C domains was

unknown until they were recently found to be responsible for the recruitment of the P450 enzymes to act on the NRPS bound peptide substrate.⁶⁴

1.3.7 Methyltransferase domain

Methyltransferase (MT) domains in NRPSs are typically responsible for the *N*-methylation of amino acids in the assembly line. The domains of around 420 amino acids are positioned after the adenylation domain in the module architecture, and act after amino acid activation and loading onto the T domain. They use (*S*)-adenosyl methionine (SAM) as a co-substrate to donate a methyl group to the α -amino group of the substrate.⁶⁵ This methylation, much like the use of D-amino acids, can lead to strong conformational effects which are essential for natural product bioactivity, as in the case of cyclosporin.⁶⁶ MT domains can also act in *trans* to the NRPS, as is the case in chloroeremomycin biosynthesis, where the N-terminal D-leucine is methylated at the heptapeptide stage once released from the NRPS.

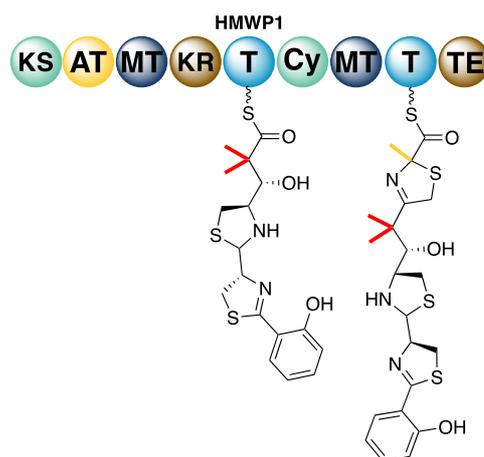


Figure 11 - Methylation in yersiniabactin biosynthesis by PKS-NRPS HWMP1. Methylation shown in red is catalysed by the 1st MT in the PKS region, and the methylation highlighted in gold is catalysed by the C-MT in the NRPS region, acting on the thiazinoyl-S-T thioester.

Methylation of carbon centres is also possible through the action of C-methyltransferase domains. These do not act on aminoacyl-thioesters, but can act on precursors to non-proteinogenic amino acids, such as the biosynthesis of β -methyl glutamate in CDA biosynthesis.⁶⁷ Another example of a C-MT domain is in the synthesis of the siderophore yersiniabactin. Here the C-MT domain is thought to act

on the thiazolinoyl-thioester bound to the T domain, rather than the enzyme-bound cysteine.

In yersiniabactin biosynthesis, the first predicted MT domain, based upon the protein sequence, is non-functional and does not contribute to the final product. It is an example of an interrupted A domain, where an auxiliary domain is inserted between core sequences of the A domain, resulting in a dual function.⁶⁸ However, in NRPS HMWP2, the SAM binding site contains a glycine to arginine mutation, rendering it non-functional. Other natural products containing MT domains in interrupted A domains include pyochelin and thiocoraline.

1.3.8 Thioesterase domain

Once extension of the peptide chain with its appropriate tailoring has been completed, the chain must be released from the enzyme. In most bacterial systems, this is accomplished by a dedicated thioesterase (TE) domain at the C-terminus of the NRPS.⁶⁹

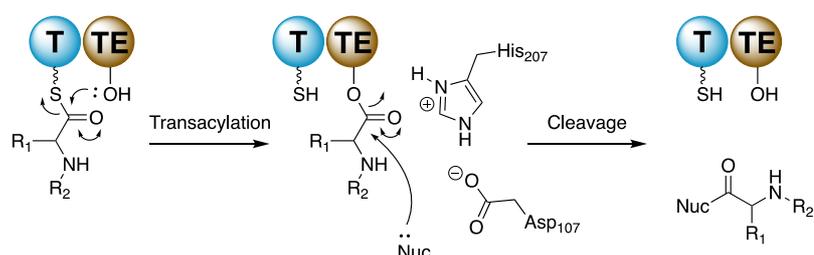


Figure 12 - Chain release of an NRPS product by the catalytic triad of a thioesterase domain and a nucleophile.

These domains use a conserved Ser-His-Asp catalytic triad to affect this chain release as shown in Figure 12.⁷⁰ The peptidyl chain is transferred from the PPant moiety of the T domain to the serine of the catalytic triad in the TE domain. The His, which is stabilized by the aspartate residue, removes the hydroxyl proton of the Ser to allow nucleophilic attack of the thioester. Once the peptidyl chain is attached to the NRPS by a new ester linkage, this can then be attacked by an internal or external nucleophile. An external nucleophile, such as water, results in the hydrolysis of a linear peptide chain, as is the case for the ACV peptide in penicillin biosynthesis.⁷¹

Alternatively, internal nucleophiles can result in the formation of cyclic peptides or depsipeptides by intramolecular amide or ester formation. Examples of these mechanisms can be seen in the biosynthesis of the macrolactone echinomycin,⁷² where the hydroxyl of a serine acts as the nucleophile, and the cyclic peptide gramicidin S, which uses the N-terminus of the NRP to attack the ester linkage as shown in Figure 13.⁷³

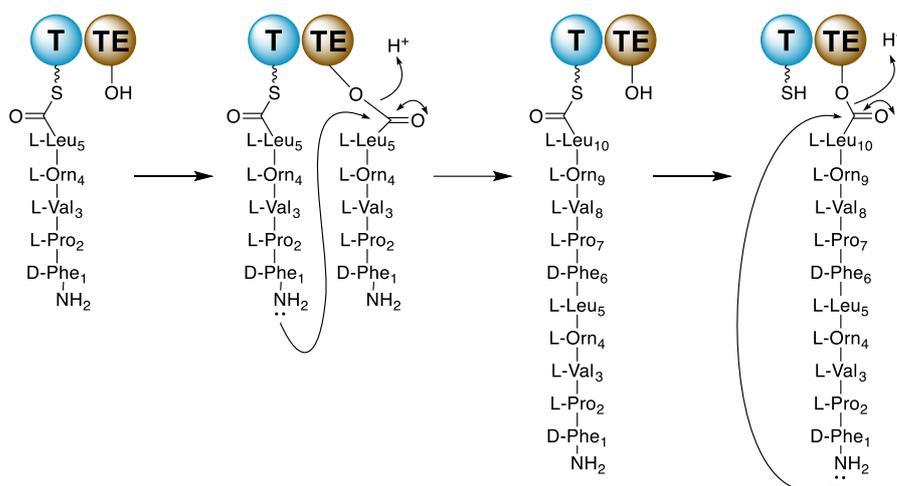


Figure 13 - Homodimerisation and cyclisation by the formation of a peptide bond in the TE release of gramicidin S.

In these cases, the TE domains act to conduct both the dimerisation and macrocyclisation processes. A peptide chain is moved to the TE domain, allowing the NRPS to attach a second peptide chain to the T domain immediately upstream of the TE domain. The dimerisation process can then occur, with the nucleophilic residue of the TE-bound chain attacking the peptidyl-PPant thioester to form a dimer, before the nucleophilic residue of the previously T domain-bound chain attacking the TE ester bond to complete the macrocyclization and release from the NRPS. Thioesterase domains have been shown to have a reasonably relaxed substrate specificity, tolerating a range of chain length and diversity in sequence outside of specific key residues.⁷⁴ It has been proposed the only essential residues are those responsible for a preorganization of the linear peptide which ensures the correct orientation for ring closure.⁷⁵

Some BGCs also encode for Type II TEs. These are responsible for the hydrolysis of mis-primed thiol groups of the 4'-PPant cofactor. As a large portion of

cellular CoA exists as acetyl-CoA, many PPTases can use this as a substrate. However this can lead to a blocked NRPS as acetyl-CoA is not a competent substrate. Type II TEs can hydrolyse these acetyl group to recover the activity of the system.⁷⁶ It is also proposed that while hydrolysis of aminoacyl intermediates is slower, the lack of downstream processing of incorrectly loaded amino acids would result in a longer occupancy, allowing type II TEs to cleave these as well.⁷⁷

1.3.9 Post Extension Tailoring

Once the peptide chain is complete, containing a selection of unique building blocks, D-amino acids, and methylation, it can be additionally altered through the action of tailoring enzymes. These can include oxidative cross-linking, glycosylation, acylation, and halogenation amongst others. A number of these modifications can be illustrated in the biosynthesis of the glycopeptide antibiotic teicoplanin.

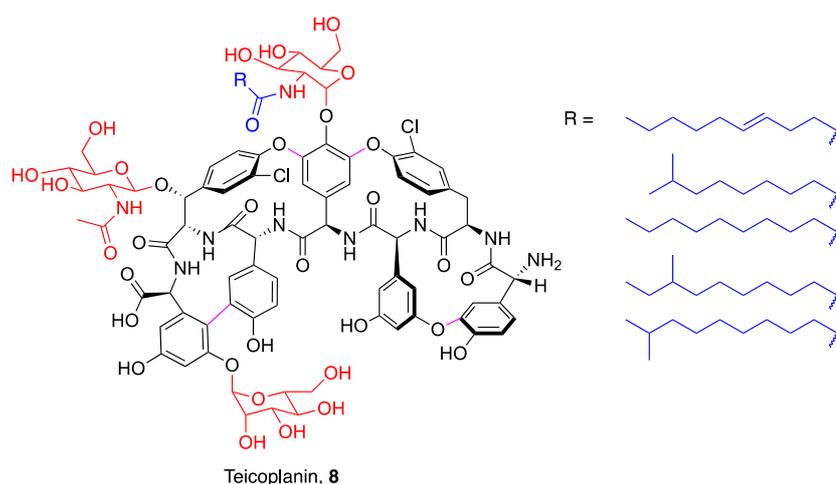


Figure 14 - Teicoplanin with post extension tailoring with *N*-acetylglucosamine, mannose, acyl chain, and oxidative cross-linking.

Teicoplanin has four oxidative cross-links installed by four P450 monooxygenases acting on the T domain bound peptide.^{78,79} These are then hydrolysed by a TE domain which is selective for the cross-linked heptapeptide.⁸⁰ This aglycone is then decorated by 2 *N*-acetylglucosamine (GlcNAc) sugars, one of which is deacylated and then acylated, and a mannose to produce mature teicoplanin.⁸¹

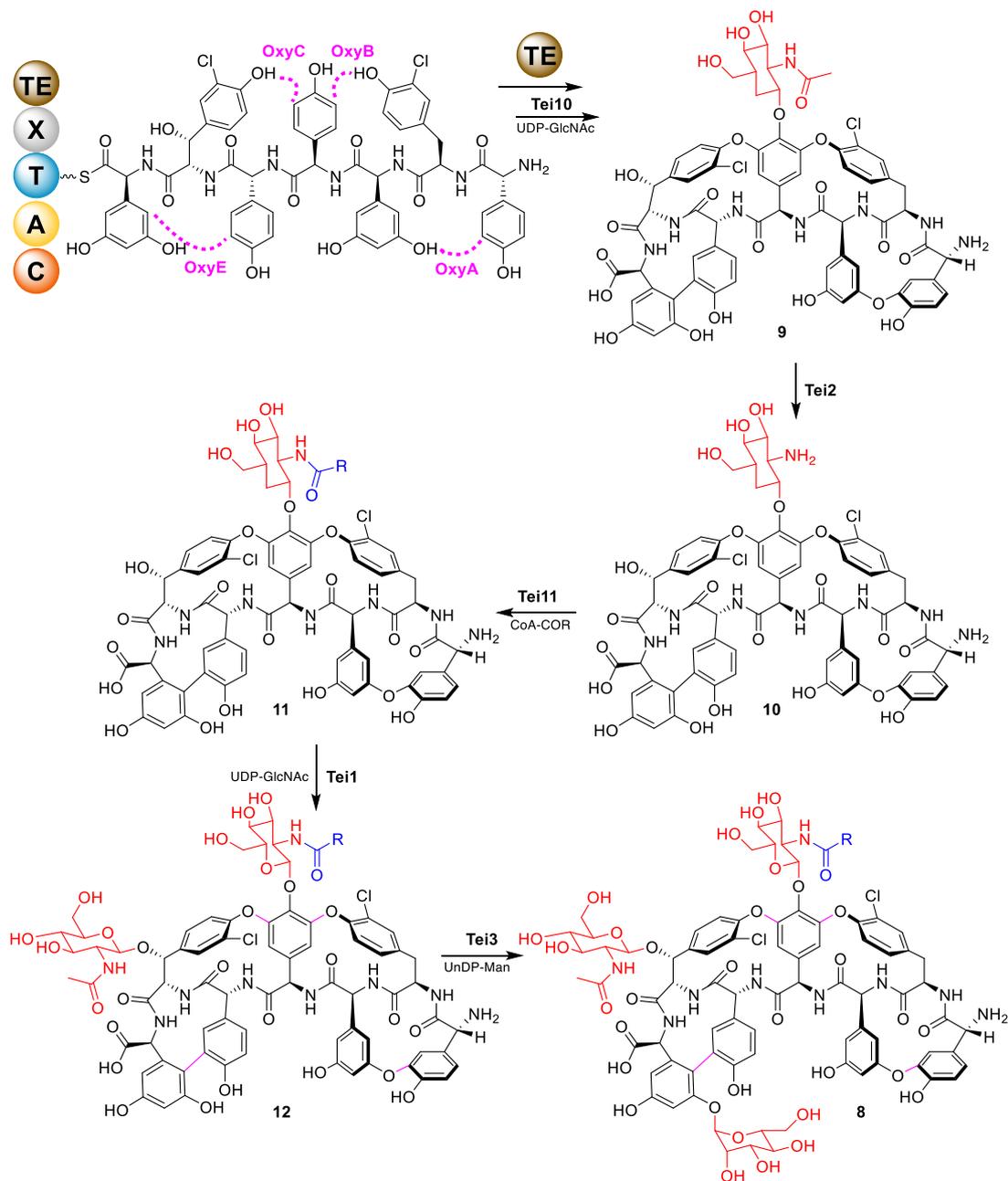


Figure 15 - Post extension tailoring steps to produce teicoplanin, including 4 cross-links, 3 glycosylations and acylation.

1.4 Generation of structural diversity by NRPSs

As seen by the number of modules that exist to provide tailoring functions in nonribosomal peptide synthetases, a huge range of chemically diverse structures are possible. In addition to the structures enabled by the use of non-proteinogenic amino

acids, epimerisation, methylation, cyclisation and macrocyclisation, further variety can be found in systems where whole modules diverge from a linear arrangement. Marahiel proposed a grouping system for NRPSs with varying module organisation, which will be used to explore these biosynthetic pathways.⁸²

1.4.1 Class A – Linear NRPSs

A classic example of a linear NRPS are the megaenzymes responsible for the biosynthesis of the heptapeptide antibiotic of last resort, vancomycin. Vancomycin has a rigidified 3-D structure which binds the D-Ala-D-Ala motif of the *N*-Acetylmuramic acid and *N*-Acetylglucosamine peptide terminus of peptidoglycan, shown in Figure 16.⁸³ This prevents the binding of penicillin-binding proteins, and subsequent transpeptidation. In this way it inhibits cell wall biosynthesis and leads to cell death in Gram-positive bacteria.

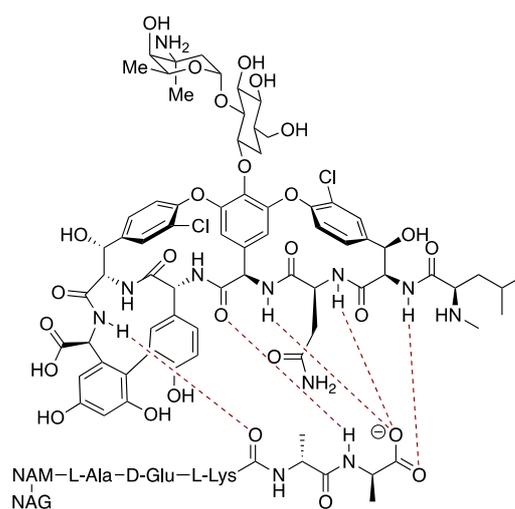


Figure 16 - Vancomycin binding the D-Ala-D-Ala motif of peptidoglycan through five hydrogen bonds.

This NRP is constructed by a 7 module NRPS which possesses a standard C-A-T domain organisation and follows the principle of colinearity. That is, the number and order of modules and the substrate specificity of each module correspond with the identity and order of the amino acids in the final NRP.⁸⁴ The linear peptide contains epimerised amino acids as predicted by the presence of epimerisation domain, and the non-proteinogenic amino acids 4-hydroxyphenylglycine, 3,5-dihydroxyphenylglycine, and β -hydroxytyrosine.

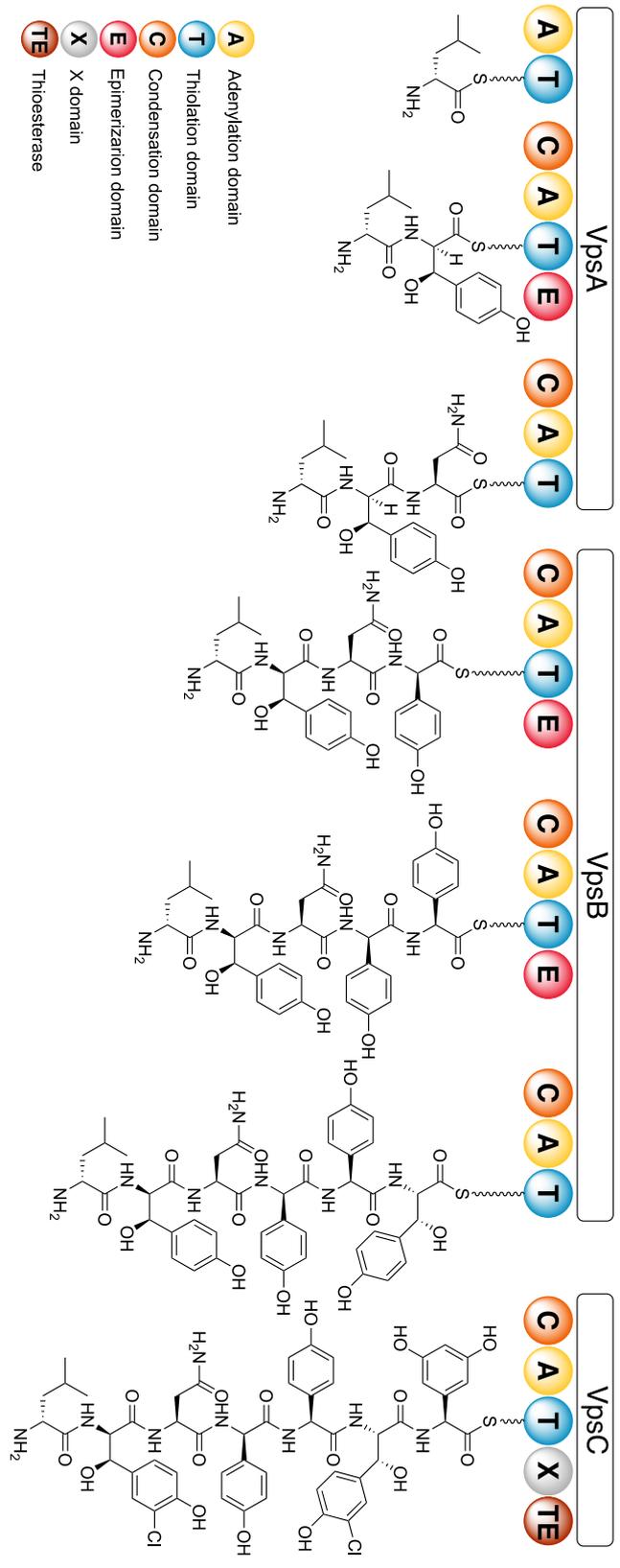


Figure 17 - Biosynthesis of class I NRP vancomycin heptapeptide by 3 NRPS.

These aromatic amino acids are crucial for tailoring oxidative cross-linking, which is in turn essential to create the correct conformation to bind D-Ala-D-Ala. After this cross-linking, the heptapeptide is released from the assembly line by a terminal TE domain.⁸⁵

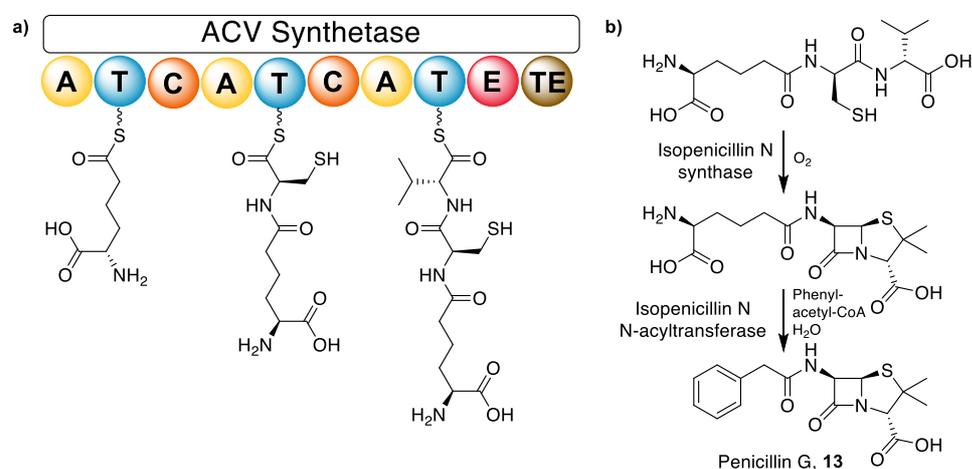


Figure 18 - a) Colinear biosynthesis of ACV peptide by ACV synthetase. b) Further processing of ACV peptide to generate penicillin G.

Another example of this linear processing is the ACV peptide precursor to the penicillins, β -lactam antibiotics. This consists of L- δ -(α -aminoadipoyl)-L-Cysteinyll-D-valine. This was discovered to be produced by a single enzyme in *Aspergillus nidulans*.⁸⁶ As understanding of NRPSs was gained, the module and domain architecture of the ACV synthetase could be deciphered and found to be as would be expected from the product, activating L-aminoadipate, L-cysteine, and L-valine, with an epimerisation domain to produce the terminal D-valine.⁸⁷

1.4.2 Class B – Iterative NRPSs

Iterative NRPSs use modules more than once to produce peptides that are larger than would be expected from examination of the NRPS which produces them. Echinomycin, a depsipeptide antitumour antibiotic produced by bacteria including *Streptomyces echinatus* and *Streptomyces lasaliensis*, is an example.⁸⁸ The two quinoxaline moieties of its pseudo-dimeric structure are used to bind DNA, which makes it a potent inhibitor of RNA synthesis.⁸⁹ Its 36 kb gene cluster was identified, containing 2 NRPS genes, coding for 4 modules.⁹⁰ 8 modules would have been

expected based on the dimeric structure of echinomycin. The FabC used to load the tryptophan-derived quinoxaline starter unit is recruited from fatty acid biosynthesis.

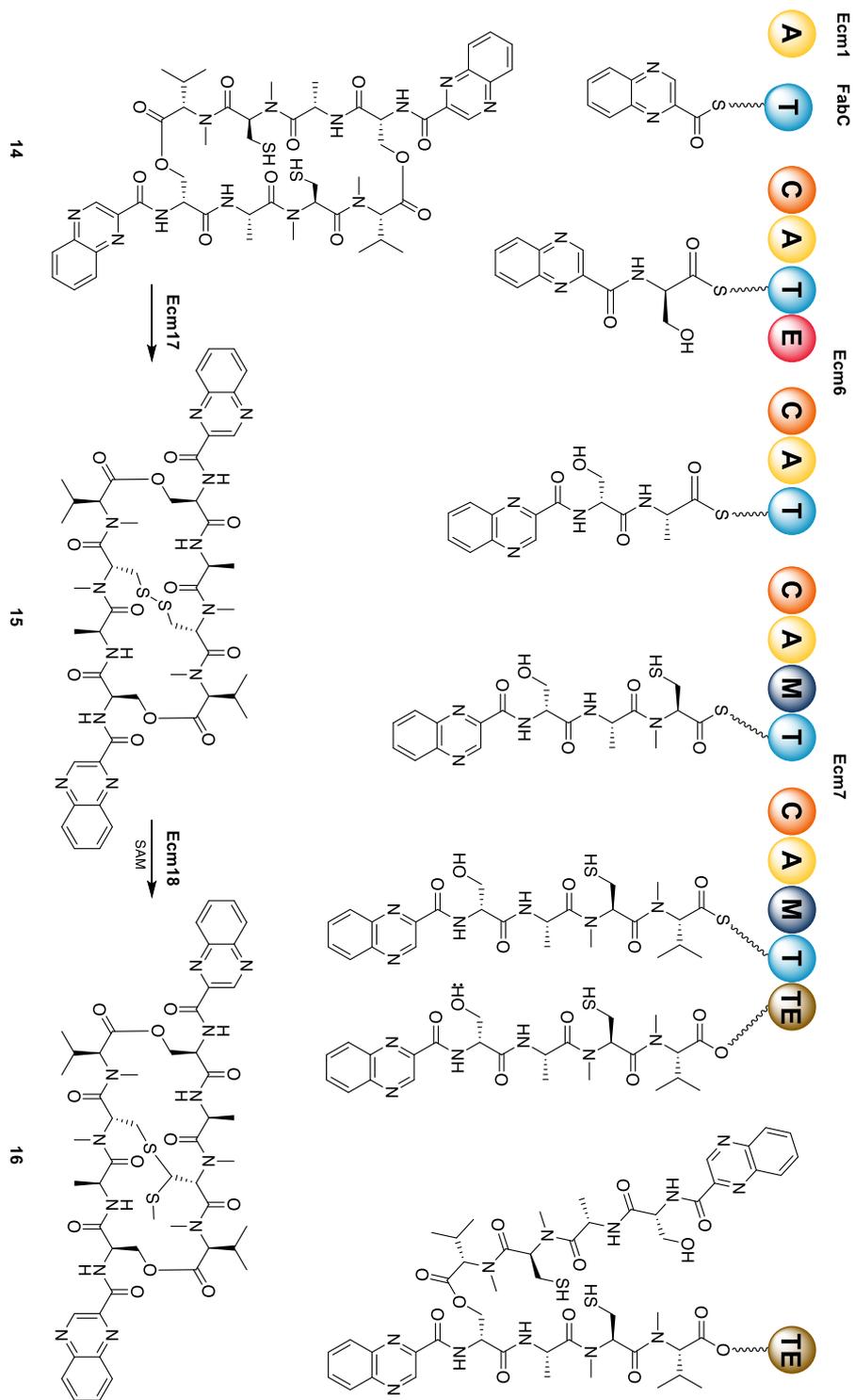


Figure 19 - Echinomycin biosynthesis by an iterative class II NRPS using each module twice, followed by dimerisation and release as a depsipeptide. Ecm 17 forms a disulfide bond, which is methylated and rearranged to a thioacetal by Ecm18.

Based on the domain organisation, a biosynthetic hypothesis could be presented, shown in Figure 19. Each module is used twice to form two tetrapeptides with an epimerised serine and *N*-methylated cysteine and valine. Further study of the thioesterase domain of the final module showed that it was capable of catalysing the peptide homodimerisation and cyclorelease.⁹¹ Additionally, it should be noted that while serine, cysteine and valine residues could be predicted based on the specificity code of the A domains, the incorporation of alanine could not be predicted with glycine suggested.

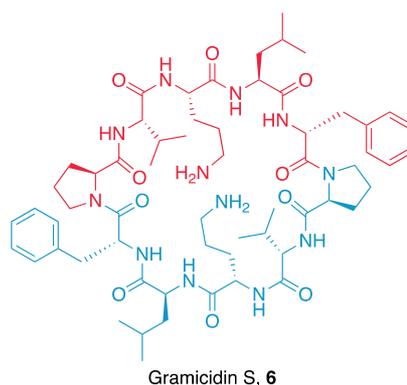


Figure 20 - Gramicidin S showing the dimeric structure formed from two pentapeptide chains by an iterative NRPS.

Gramicidin S, **6**, is a cyclic peptide antibiotic discovered in 1942, which has been extensively studied as one of the earliest known NRPs.⁹² The structure, shown in Figure 20, was confirmed by synthesis before its biosynthesis was studied in detail.⁹³ The cyclic structure is formed by the head-to-tail macrocyclisation of a decapeptide. It has been shown that in a similar manner to echinomycin, the thioesterase can catalyse the dimerisation and cyclisation of two linear pentapeptide chains. As a result, each module is used twice in the biosynthesis of each molecule of gramicidin S.⁷³

1.4.3 Class C – Nonlinear NRPSs

In nonlinear NRPSs, the domain architecture can differ dramatically from the usual C-A-T structure found in most NRPSs. Possession of an uncommon arrangement of these core domains is the characteristic trait of these nonlinear systems, which diverge greatly from colinearity.

The NRPS producing yersiniabactin, a siderophore from *Yersinia pestis*, is a clear example. This combines both NRPS and polyketide synthase (PKS) genes both within and across enzymes.⁹⁴ While in conventional NRP biosynthesis, a single adenylation domain loads a single T domain, in yersiniabactin biosynthesis one A domain on HMWP2 is responsible for loading three T domains with cysteine, shown in Figure 21. One T domain is located within the same module, another is in the next module, and the third is located on HMWP1, a second NRPS entirely.⁵²

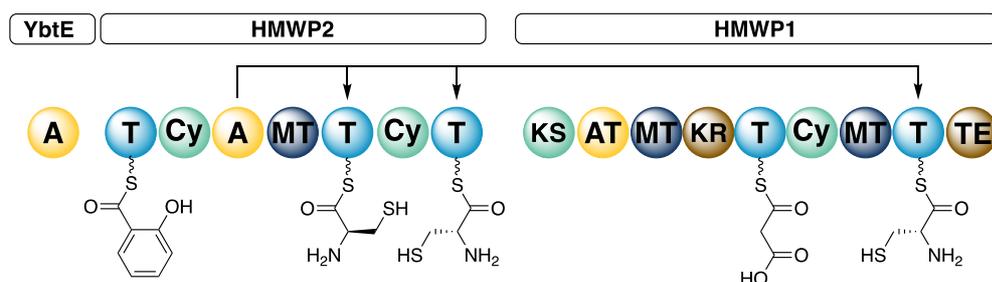


Figure 21 - The domain organisation of the yersiniabactin PKS-NRPS, showing an unusual A domain which loads three T domains.

A further example of rare domain organisation is found in the biosynthetic pathway of the genotoxin colibactin. While this pathway possesses a range of unusual features, which will be discussed in subsequent chapters, the most relevant is the NRPS ClbH. This has a A-C-A-T organization, where the A₂ domain loads the T domain of the same module. However, the A₁ domain activates and loads L-serine to a standalone acyl carrier protein, ClbE.⁹⁵ This is especially interesting given the structural differences noted in previous sections between peptidyl and acyl carrier proteins.

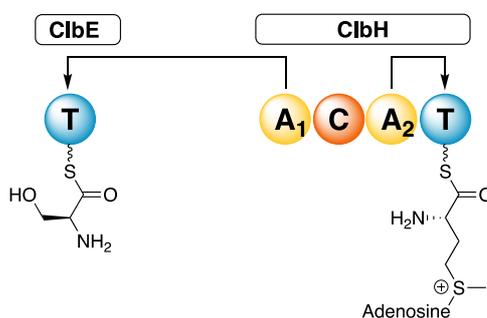


Figure 22 - Unusual domain structure in NRPS ClbH which contains two A domains. A₂ loads SAM onto the modules T domain, while A₁ is used to load L-serine onto standalone ACP ClbE.

1.5 Approaches to study NRP biosynthesis

Given the massive importance of nonribosomal peptides to humanity in light of their biological activity, there is substantial interest in the mechanism by which they are biosynthesized. It is hoped that a more detailed understanding of the essential structural and mechanistic elements of NRPSs could lead to rational engineering of these pathways to produce novel derivatives. Techniques that have been used to dissect and study NRPSs will be discussed below.

1.5.1 Isotopic labelling

One of the earliest approaches to study NRPS enzymes was the use of isotopically labelled precursors. Notably, Laland and co-workers used radioisotopically labelled ^{14}C L-Phe to characterize the order of incorporation of amino acids in the partially purified gramicidin synthetase enzymes.¹² Investigation into amino acid activation by ACV synthetase exploited $[4\text{-}^2\text{H}_6, ^{18}\text{O}_2]$ -valine to follow the loss of oxygen during the formation of the covalent enzyme species.⁹⁶ The use of stable isotopes is preferable compared to the radiochemistry of earlier work. These isotopes can be combined with high-resolution mass spectrometry to great effect, such as finding the molecular formula using an isotopologue. In a reverse-approach, non-labelled amino acids were fed to cultures growing in ^{13}C media, to give a characteristic mass shift for each amino acid incorporated in the peptides. This is a particularly appealing strategy if the labelled precursors are prohibitively expensive. These approaches can be combined with tandem mass spectrometry (MS^2) to fragment the peptide, and assign the position of each amino acid in the structure.⁹⁷

Labelling experiments can also be extended to determine the absolute configuration of amino acids, and to differentiate between isobaric moieties, such as leucine, isoleucine, and *N*-methylvaline using appropriately fed cultures. Incorporating deuterium, ^2H , at the C_α of an amino acid allows the activity of E domains to be examined. With epimerisation, the ^2H would be exchanged with solvent and lost, giving a mass shift of 1 Da. However, this does require the use of transaminase-deficient knockout strains, otherwise the labelled amino acids will be converted to α -keto acids, with the loss of the label.⁹⁸ Isotopically labelled precursors

can also be used in a less targeted manner. When combined with molecular networking techniques, the mass shift undergone by certain metabolites can be used to identify inherent substrate flexibility in NRPSs, and find structurally related natural products.⁹⁹ The use of auxotrophic *E. coli* strains as heterologous hosts with ¹³C amino acids can also be used to identify unknown metabolites from a pathway. Notably in colibactin biosynthesis L-[U-¹³C]-methionine, known to be required for the biosynthesis of an aminocyclopropane moiety by a SAM-dependent mechanism,¹⁰⁰ was shown to be incorporated twice in late-stage metabolites of the pathway, forming a key piece of evidence in the biosynthetic hypothesis.¹⁰¹

1.5.2 *In vitro* enzyme characterisation

Most detailed mechanistic studies of NRPSs have come from *in vitro* experimentation and characterisation of the functions of domains and key residues within. The move from cell-free extracts of producing organisms to over-expressed and purified protein from *E. coli* has been key to this. Work to sequence the gene clusters encoding NRPS was essential, allowing the ability to delete or mutate specific domains and modules selectively to investigate function. This was used to great effect in the dissection of the structure of GrsA. Truncation of the protein at the C-terminus by 291-aa revealed a loss of epimerase activity, while a further C-terminal deletion of the 4'-PPant binding site prevented thioester formation, providing clear evidence for the modular nature of NRPS.⁵⁹

A second vital innovation for the *in vitro* study of NRPS was the discovery of the promiscuous phosphopantetheinyl transferase (PPTase) Sfp from *Bacillus subtilis*. Unlike previous PPTases which had demonstrated high carrier protein specificity, Sfp was found to be able to both load a wide range of carrier proteins, but also use a variety of CoA derivatives to modify these proteins. This allowed the use of custom acyl-S-phosphopantetheinyl moieties, allowing examination of the effects of non-cognate substrates,⁴⁷ and interrogation of mechanistic details.¹⁰²

Given C domains require at least one T domain bound substrate, Sfp opened the door to a greater understanding of these domains. The substrate specificity of the donor and acceptor sites of EntF, as discussed earlier, could now be examined using

synthetic CoAs, or later synthetically more accessible SNAC substrate mimics.⁴⁸ Extending this methodology to peptidyl-CoA compounds loaded *via* Sfp has allowed the interrogation of TE domains which acts as cyclases, such as fengycin cyclase, which does not accept peptidyl-SNACs as substrates.¹⁰³ Peptidyl-CoAs have also been extensively used to understand the action of accessory tailoring enzymes which act upon the NRPS-bound peptide, such as the P450s in glycopeptide biosynthesis.¹⁰⁴

Adenylation domains have also been thoroughly investigated using *in vitro* methods, with perhaps the broadest range of tools available to study their activity. Due to their gatekeeping function it is important to understand when engineering NRPSs if poor product formation is due to poor substrate activation by the A domain, or a lack of acceptance further downstream. With two reactions catalysed by A domains, it is perhaps preferable to study the initial activation of an amino acid with ATP, as this does not rely on loading a T domain or an equivalent step.

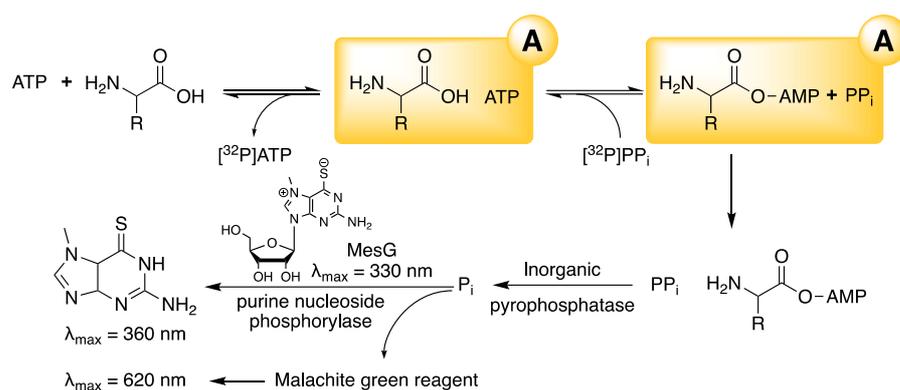


Figure 23 - Common methods for *in vitro* study of A domain activity. Showing the measurement of [³²P]ATP to measure the reversible breakdown of the aminoacyl-AMP, and the use of phosphate (P_i) in malachite green or MesG based assays.

This approach is the foundation of the oldest A domain activity assay. Formation of an aminoacyl-AMP and PP_i from the amino acid and ATP is a reversible process. Using radiolabelled ³²PP_i, ³²ATP can be generated and quantified.¹⁰⁵ The back reaction to ATP correlates with the second reaction of loading onto a T domain. Given the specialised handling and measurement of radioisotopes, alternative colourimetric approaches have been developed. These rely on the cleavage of the PP_i produced to inorganic phosphate, P_i, by a coupled enzyme reaction. An assay developed to use malachite green relies on a chromogenic malachite green-ammonium molybdate

complex, which forms a green complex which absorbs at 620 nm in the presence of P_i .^{106,107} An alternative approach uses a purine nucleoside phosphorylase and P_i to breakdown chromogenic 7-methylthioguanosine, MesG, producing a shift in λ_{max} from 360 nm to 330 nm.¹⁰⁸ An additional consideration is that some enzymes do not readily release PP_i . In these systems the use of hydroxylamine as a nucleophile to form a hydroxamic acid from the acyl-AMP was demonstrated.¹⁰⁹ These hydroxamic acids then form a chromogenic complex upon the addition of Fe(III), however this method can lack sensitivity and suffer from an unstable iron complex. This use of hydroxylamine as a proxy nucleophile for the T domain has subsequently been adopted by the MesG-based assay.¹¹⁰

1.5.3 *In vivo* studies

In vivo experiments have also been used to gather information on NRPS. While offering a less controlled system than individual enzymes *in vitro*, working with whole organisms can give insights that are not possible in other systems, but require careful interpretation. For example, when mutating conserved residues in the C domain of EntF to identify those which were essential, it was found that mutations which resulted in a 200-fold drop in activity *in vitro* had a relatively limited effect *in vivo*. It was proposed that the amide bond formation catalysed was not rate-limiting, and thus enterobactin production was only affected by the most severe mutations.¹¹¹

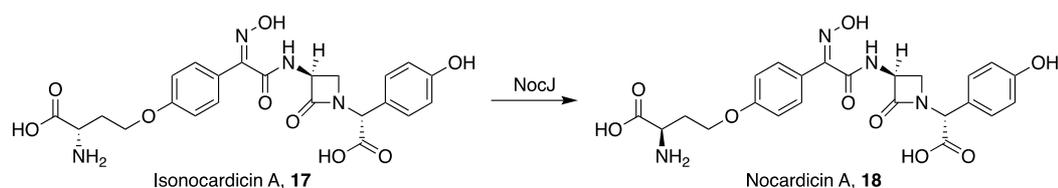


Figure 24 - Epimerisation of isonocardicin A, 17, to produce nocardicin A, 18, catalysed by NocJ.

Strains with knockouts in key biosynthetic genes can also be used to accumulate precursors, and thereby shed light on the timing of transformations. In the biosynthesis of nocardicin A, a β -lactam antibiotic, the gene *nocJ* was knocked out by insertional mutagenesis. This led to the accumulation of a precursor, isonocardicin A, while nocardicin production could be restored with complementation of *nocJ*. Combined with *in vitro* assays, this confirmed the role of NocJ as an epimerase which

acts at the final step in biosynthesis.¹¹² This approach is particularly illuminating in the study of tailoring P450 enzymes in glycopeptide antibiotic biosynthesis.

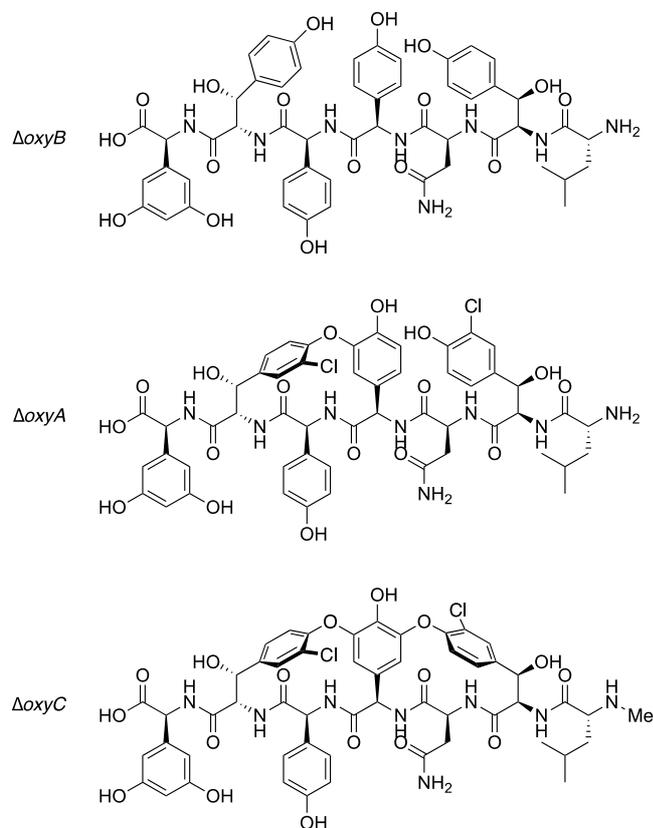


Figure 25 - Example of heptapeptides isolated from P450 monooxygenase knockout mutant strains of *Amycolatopsis balhimycina*.

The vancomycin-type glycopeptide balhimycin has three P450s, each responsible for a single aryl-ether or aryl-aryl crosslink. Knockouts of the gene *oxyA*, *oxyB*, and *oxyC* showed accumulation of metabolites with differing levels of cyclization, shown in Figure 25. *OxyB* mutant products had no crosslinks, whereas products from *oxyA* and *oxyC* mutants contained a single and two crosslinks respectively; these experiments revealed both the reactions catalysed by the P450s and suggested their order.¹¹³

However, not all organisms are amenable to the genetic manipulation necessary for *in vivo* studies. Some are not suited to large-scale growth or are genetically intractable, making heterologous expression necessary.¹¹⁴ Mutations can then be introduced when the BGC is in a host organism with well characterised genetic

tools, such as *E. coli* or *S. coelicolor*.¹¹⁵ Manipulation of key biosynthetic genes in a heterologous host was used to identify the role of multiple methyltransferases in bottromycin biosynthesis, a ribosomally produced and post-translationally modified peptide (RiPP). In the producer *Streptomyces* sp. BC16019 genetic manipulation was limited to single-crossover mutations with relatively low efficiency. This limited the number of mutations possible and could lead to polar effects on the expression of downstream genes. The BGC was therefore moved to *S. coelicolor* and methyltransferase knockouts introduced, allowing the reactions catalysed by each enzyme to be characterised.¹¹⁶

Traditionally gene cluster have been isolated from genomic DNA using cosmids or bacterial artificial chromosomes (BACs) to generate libraries with large inserts. However, these approaches involving library construction and screening are often labour-intensive and time-consuming. Recently the wealth of genomic sequence data has allowed the targeted capture of specific clusters, using technologies such as Direct Pathway Cloning (DiPaC)¹¹⁷ or transformation-assisted recombination (TAR).¹¹⁸ TAR has been used to capture the gene cluster for the lipopeptide antibiotic taromycin A.

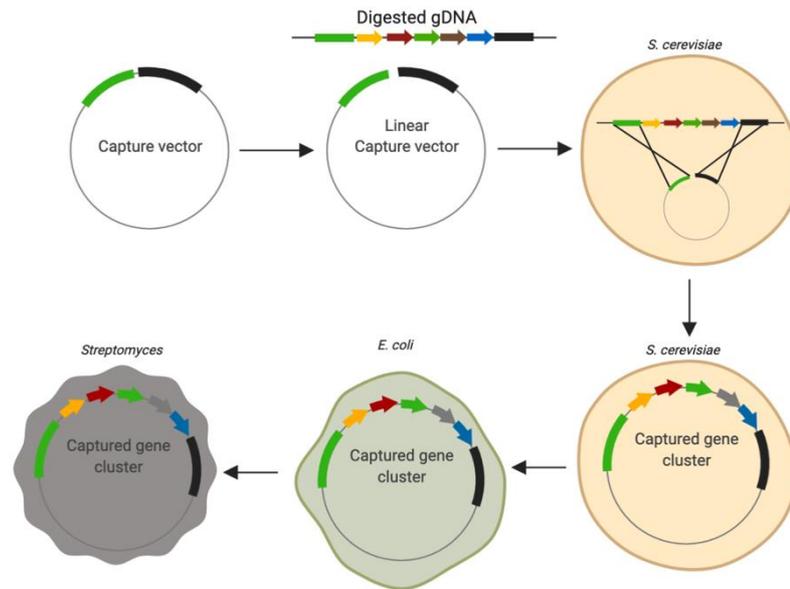


Figure 26 - TAR Workflow. The capture vector and genomic DNA are linearised by digestion and introduced to *Saccharomyces cerevisiae* where homologous recombination occurs. The vector containing the captured gene cluster can then be transformed into *E. coli* and conjugated in a *Streptomyces* host.

This technique relies on a capture vector containing 60 base pair areas of homology with the ends of the desired gene cluster. The linearised vector and digested genomic DNA can then be introduced to *Saccharomyces cerevisiae* spheroplasts and utilise yeasts inherent homologous recombination activity to introduce the cluster into the vector, re-circularising it. The vector can then be transformed into *E. coli* and then introduced into a *Streptomyces* host by conjugation, and integrated into the chromosome. This approach has been developed to include a counter-selection marker on the capture vector. The enzyme URA3 converts 5-fluoroorotic acid to toxic anti-metabolite 5-fluorouracil, to select against false-positives resulting from non-homologous end joining (NHEJ).¹¹⁹

The repertoire of molecular biology tools to edit genomes has also made significant progress recently. In particular, the advent of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas9 tools offer considerable advantages compared to traditional methods. A standard method to introduce mutations relied on homologous recombination of introduced DNA and the bacterial chromosome.

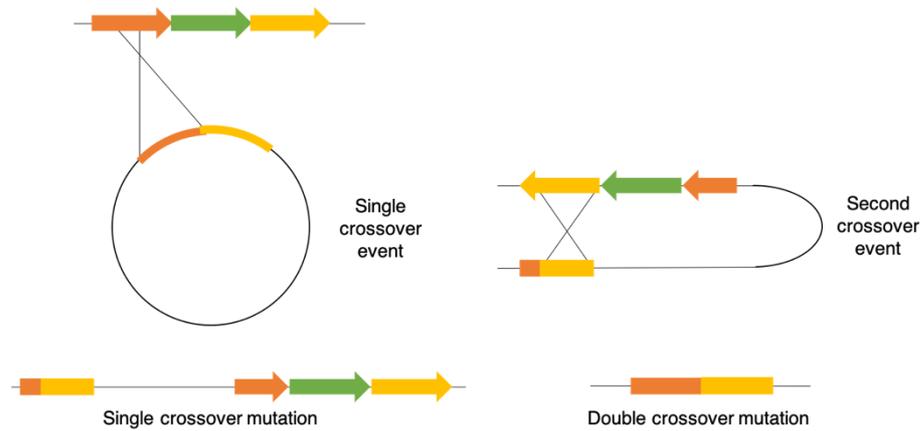


Figure 27 - Single and double crossover mutations, showing the large insert left behind in a single crossover compared to the cleaner double crossover mutation.

This could occur as either single or double-crossover events.¹²⁰ Single crossover events inserts the entire delivery plasmid into crossover site, which can lead to polar effects, and antibiotic selection must be used to maintain it. This limits the number of mutations to the number of effective antibiotics. A double crossover is more stable and has less impact on the edited DNA region, but relies on a much less common second crossover, which must be screened for by a counter-selection marker. Replica plating can then determine which colonies have undergone a second crossover, and which have simply reversed the first crossover.

The CRISPR-Cas9 system was originally demonstrated in *Streptococcus pyogenes* but has since been adapted for a wide variety of eukaryotes and prokaryotes including humans, plants, yeast, *E. coli*, and *Streptomyces*.¹²¹ The CRISPR-associated protein Cas9, an endonuclease, cleaves targeted double strand DNA. This targeting relies on two RNA strands, CRISPR RNA (crRNA), and trans-acting crRNA (tracrRNA), and this RNA duplex directs DNA cleavage. Identification of the target sequence requires base pairing to a 20 bp protospacer sequence on the crRNA, as well as protospacer-adjacent motif (PAM) in the target DNA sequence. This PAM is NGG for *S. pyogenes*, where N is any nucleotide.¹²² These RNAs can be joined, giving a single-guide-RNA (sgRNA). The double-strand break then occurs 3 bp upstream from the PAM motif.¹²¹ Editing of the crRNA allows the targeting of any DNA sequence which contains the PAM motif. Once cleaved DNA is then repaired by either homology-directed repair (HDR), which can be manipulated to introduce mutations

by providing homologous DNA, or non-homologous end joining, which can introduce frame-shifts and disrupt the genes function.¹²³

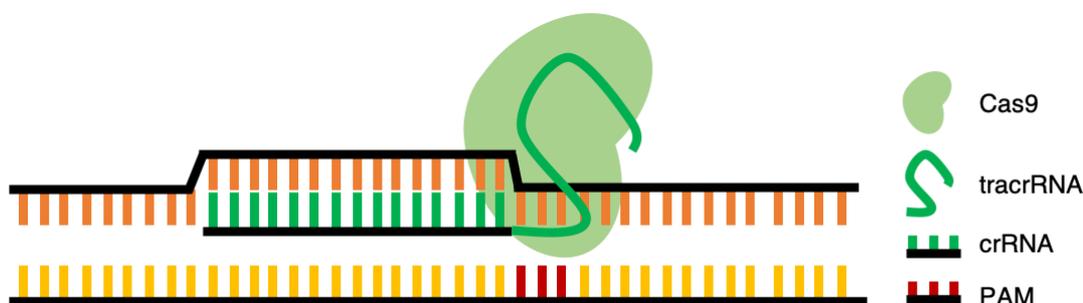


Figure 28 - Targeting of the Cas9 nuclease by a single guide RNA (sgRNA) consisting of CRISPR RNA (crRNA) and trans-acting crRNA (tracrRNA) neighbouring the protospacer adjacent motif (PAM) in the target DNA sequence.

A number of CRISPR/Cas9 systems have been developed for the editing of Streptomyces, allowing for a quick generation of mutants without markers in the target site.¹²⁴ The pCRISPomyces systems developed by Zhao and coworkers utilises a 20 bp insert coding for the sgRNA introduced by Golden Gate assembly, and homology arms introduced by Gibson assembly to rapidly create a construct for genome engineering.^{125,126} These elements are combined with the *S. pyogenes cas9* codon-optimized for Streptomyces to give a system which has been used for rapid genome editing in *S. lividans*.¹²⁷

1.5.4 Mass spectrometry

The development of soft ionisation techniques such as matrix-assisted-laser-desorption ionisation (MALDI), and electrospray ionisation (ESI) has allowed the production of intact ions of large biomolecules.¹²⁸ Top down mass spectrometry, where proteolysis or tandem mass spectrometric fragmentation (MS^n) is limited to give large fragments is a preferred method for the study of natural product biosynthesis.¹²⁹ Early work used a limited trypsin digest of the PKS-NRPS HMWP1 from yersiniabactin biosynthesis, followed by ESI Fourier Transform mass spectrometry (FTMS) to detect the ACP bearing different thioester bound intermediates.¹³⁰ This technique is suited to the detection of pathway intermediates due to the soft ionisation of ESI and compatibility with MS^2 , and the high-resolution afforded by FTMS

allowing the detection of small mass changes with high accuracy.¹³¹ Investigation of small proteins, such as carrier proteins, can also be conducted with the high accuracy and used to analyse chemical modifications to the bound acyl intermediate. In the oxidation of prolyl-S-T domain to pyrrolyl-S-T domain in clorobiocin biosynthesis, two consecutive 2 Da losses could be observed using a rapid quench, and allowed the identification of the intermediate species, shown in Figure 29.¹³²

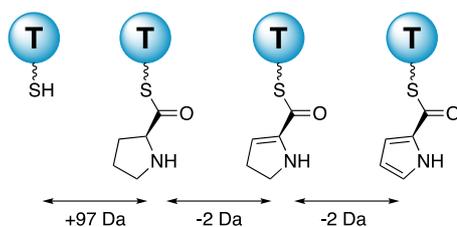


Figure 29 - T domain-bound intermediates in the biosynthesis of pyrrolyl in clorobiocin, detected by ESI-FTMS.

Using MS² in the form of either infrared multiphoton dissociation (IRMPD) or collisionally activated dissociation (CAD) led to the elimination of the phosphopantetheinyl moiety of PKS/NRPS carrier proteins, shown in Figure 30. This reduces the size of the ions 100-fold, to less than 1000 Da. These smaller ions can be detected at sufficient resolution to assign an empirical formula to the acyl species.¹³³ This was extended to lower resolution instruments using MS³ to generate a characteristic fragmentation pattern of the 4'-PPant moiety.¹³⁴

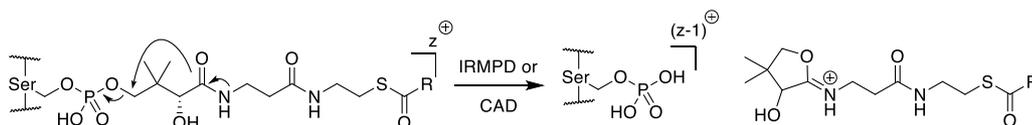


Figure 30 - Proposed mechanism for the IRMPD or CAD induced ejection of the pantetheine moiety of NRPS.

This technique has been used to characterize intermediates bound to the PKS in the biosynthesis of lovastatin, with 5 PPant intermediates ejected from their carrier proteins.¹³⁵

The pantetheine ejection strategy has also been adopted in multiple proteomics approaches, such as PrISM and OASIS. PrISM (Proteomic Investigation of Secondary Metabolites) identifies high MW proteins (>200 kDa), which are likely to be megasynthases, on SDS-PAGE gels. These are then subjected to an in-gel trypsin

digest, and pantetheine ejection. Loss of the characteristic PPant moiety marks the enzyme as a potential PKS or NRPS, and the MS/MS data gathered can be used to link the proteins back to gene sequences.¹³⁶ OASIS (Orthogonal Active Site Identification System) uses pantetheine based probes bearing either alkyne or biotin moieties to enrich proteomic samples for modular synthetases, in efforts to enhance natural product discovery.¹³⁷

1.5.5 Chemical probes

In addition to the tools previously mentioned, several chemical probes have also been developed to interrogate nonribosomal peptide biosynthesis. Indeed, the aminoacyl-SNACs used by Ehmann *et al.* can be considered as probes for the C domain, which bypass the selectivity constraints of the A domain. Further chemical probes for the C domain which act as reaction-competent substrates have since been created by the Schmeing group. These rely on site-directed mutagenesis to introduce a reaction cysteine residue into the pantetheine binding channel of the C domain, which can be alkylated by the chemical probe. A linker of varying length is used to mimic delivery of the substrate by the T domain. This approach allowed the recent crystallisation of a C domain with a competent substrate for the first time, allowing examination of the residues important for acceptor site specificity.⁴⁶

Mechanism-based probes have also been developed to investigate A domain activation, and protein-protein interactions between A domains and their respective T domains. These probes mimic the salicylyl-AMP produced by A domain MbtA in mycobactin biosynthesis, which would be loaded onto the aryl carrier protein (ArCP) domain. They contain a vinylsulfonamide Michael acceptor, which reacts with the thiol of the 4'-PPant to form a covalent bond, shown in Figure 31. The probes were shown to require the A domain for reactivity with the ArCP, and the reaction was specific to the cognate ArCP.¹³⁸ In their ability to crosslink proteins, these are similar to the pantetheine derivatives used to cross ACP and ketosynthase (KS) domains in PKS biosynthesis.¹³⁹

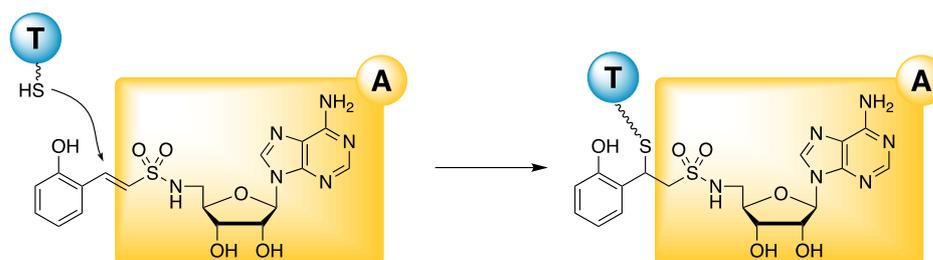


Figure 31 - Vinylsulfonamide based probes for A domains, which stabilise A-T domain interactions.

The ability of probes to stabilise certain protein conformations has also been exploited. A domains have two distinct conformations for each of the reactions catalysed, which can be studied using probes as substrate mimics. Replacing the AMP of an aminoacyl-adenylate with a non-hydrolysable adenosine monosulfamate (AMS) increases the number of interactions with the substrate pocket, leading to potent binding and can trap the A domain in the amino acid activating conformation.¹⁴⁰ Using a Michael acceptor, as shown in figure 31, in the form of an adenosine vinyl sulfonamide (AVS) can be used to trap the thioester forming conformation. This strategy was successfully used to solve the structure of an A-T didomain in an uncharacterised pathway from *Pseudomonas aeruginosa*.³²

Proteomics probes have also taken advantage of the tight binding provided by the AMS functionality. Modification of the 2' oxygen of ribose allows the attachment of a long PEG linker to avoid steric clashes with avidin when a biotin handle is attached. The demonstrate high selectivity for specific A domains with no cross-reactivity, and provide a versatile probe to investigate the proteome.¹⁴¹ This approach was extended by using photoreactive benzophenone or a 'clickable' alkyne in place of biotin, shown in Figure 32, where orthogonal reactivity could be used to simultaneously label multiple A domains in an NRPS allowing functional characterisation.¹⁴²

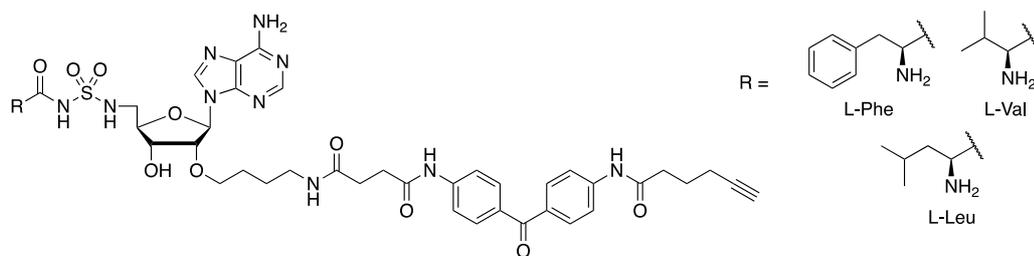


Figure 32 - Benzophenone and alkyne functionalized proteomic probes developed by Kakeya and co-workers.

TE domains have also been targeted by mechanism-based probes, as shown by the use of an α -chloroacetamide derivatized CoA to study the protein structure of a T-TE didomain of NRPS EntF. The chloroacetamide can react with the catalytic Ser residue to form an epoxide, which can be hydrolysed to form a stable amide analog of the thioester.¹⁴³ Auxiliary enzymes such as the P450 monooxygenase responsible for β -hydroxy amino acid formation in skyllamycin biosynthesis have also been studied using modified CoA compounds. By exploiting the stable coordination of nitrogen-containing heterocycles to Fe(III), complexes of the T domains and the P450 could be crystallized to give insight into the residues important for protein-protein interactions.¹⁴⁴

Chemical probes for the identification of natural products from complex mixtures based on reactive moieties have also been developed. These probes contain a reactive moiety such as a nitroso, thiol, or azide, combined with a UV chromophore and a halogen for MS tagging, as shown in Figure 33.

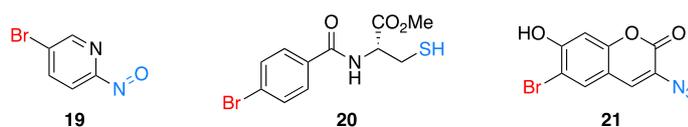


Figure 33 – Probes for the functionalization of natural products, with the halogen MS tags highlighted in red, and the reactive functionality in blue. Aromatic structures provide a distinctive UV signal.

Electrophilic functionalities such as epoxides and α,β -unsaturated ketones, and β -lactams can react with thiols to form easily detected adducts.¹⁴⁵ Extended unsaturated systems undergo nitroso-Diels-Alder cyclisations, allowing identification by the distinctive chromophore and bromine isotopic pattern.¹⁴⁶ This strategy was used to particular effect in understanding the biosynthetic pathway of vatiamides A-F,

which are produced by a highly unusual PKS/NRPS, where the PKS module interacts with 3 NRPS modules to produce an array of products.¹⁴⁷

Many of these chemical probes are limited to *in vitro* systems using purified enzymes, and those that are not are mainly directed to the study of adenylation domains. Chemical probes for the chain extension steps for a megaenzyme assembly line *in vivo* have been developed by the Tosin group for the use in PKS biosynthesis.

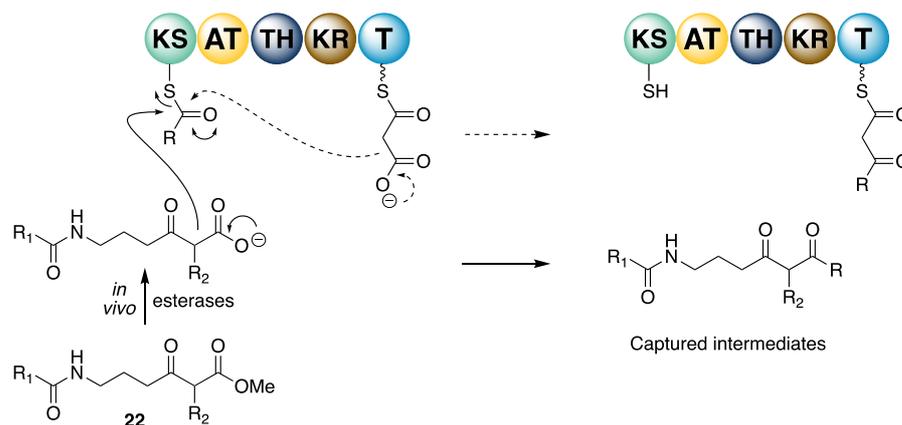


Figure 34 – Mechanism of intermediate capture. PKS probes mimic the enzyme-bound malonyl extender unit and compete to undergo decarboxylate Claisen condensation. Captured intermediates are no longer enzyme bound and can be studied by mass spectrometry.

Chain extension in these systems uses a decarboxylative Claisen condensation, typically using a malonyl group to extend by two carbon units. Chemical probes which mimic the 4'-Ppant bound malonyl group were developed, containing a CH₂ in place of the sulfur of the thioester. These were shown to be capable of intercepting chain extension in a range of systems by competing with the extender unit as shown in Figure 34, offloading the growing chain for characterisation by mass spectrometry.^{148,149}

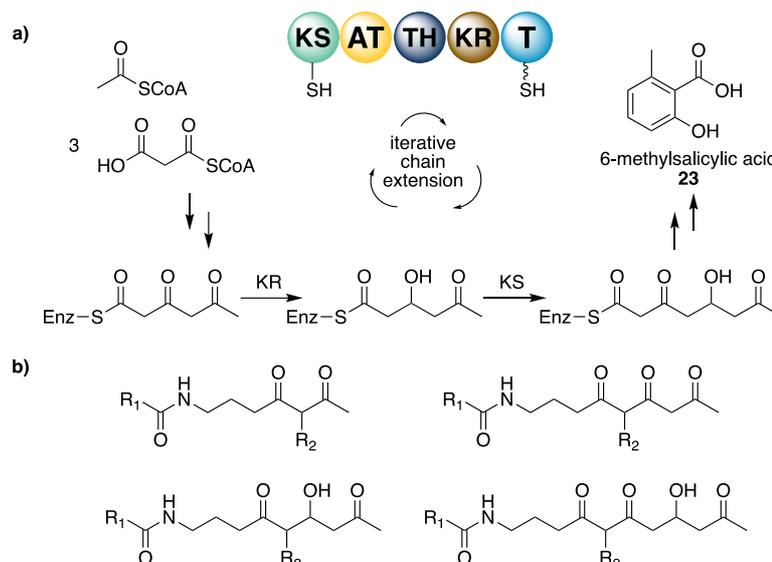


Figure 35 – a) The biosynthesis of 6-methylsalicylic acid by an iterative type I PKS. B) Offloaded intermediates which support the proposed timing of reduction in the biosynthetic pathway.

This includes highly iterative systems such the type I PKS of 6-methylsalicylic acid, **23**, biosynthesis, providing direct evidence for the timing of dehydration and cyclisation as shown in Figure 35.¹⁵⁰ Further modifications of the probe have incorporated azide moieties allowing the derivatisation of captured intermediates by copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) chemistry or a Staudinger reaction.¹⁵¹ Acetoxymethyl esters, **24**, have also be used in place of methyl esters to give greater concentrations of active probe *in vivo*, allowing the characterisation of more offloaded species.¹⁵² These represent a novel method to characterise putative biosynthetic intermediates and to gain insight into biosynthetic process, such as the timing of on assembly line processing.

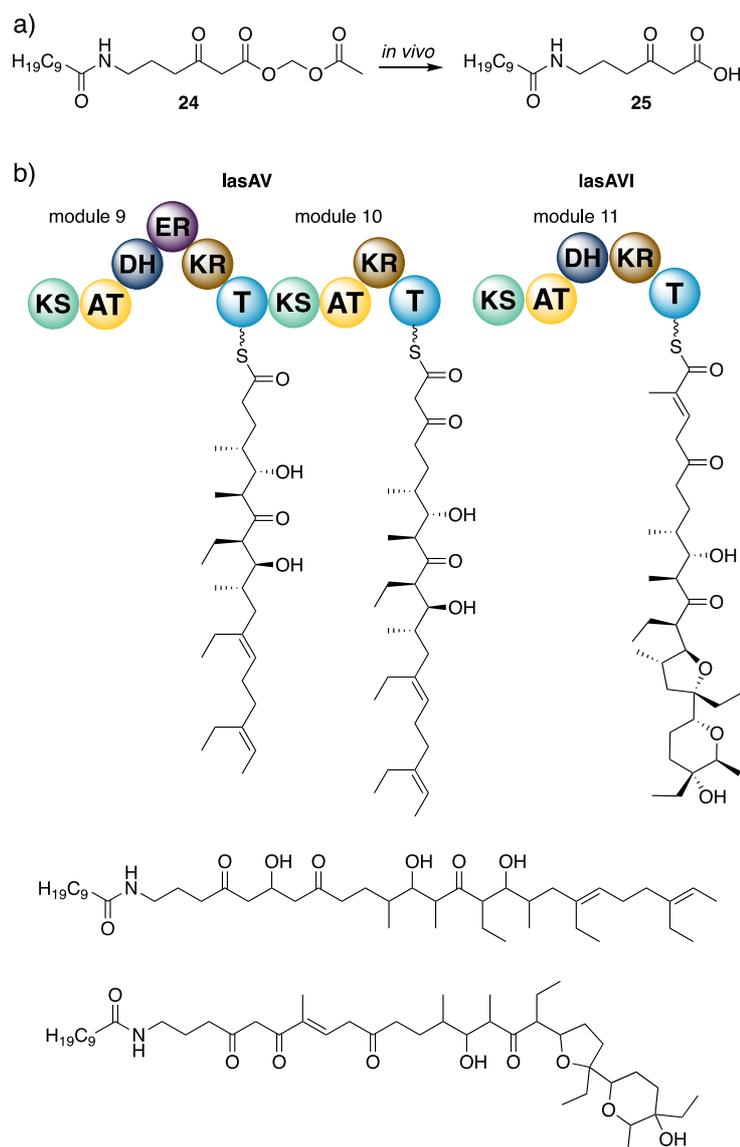


Figure 36 – Chain termination probes for PKS biosynthesis. *In vivo* hydrolysis of acetoxyethyl ester precursors generate the activate β -ketoacid probe 25. b) Offloading of late-stage biosynthetic intermediates from the lasalocid A synthase.

This approach has been very recently adapted to the study of nonribosomal peptide synthetases, which I have contributed to.¹⁵³ It has been shown that chemical probes designed to mimic the PPant-bound aminoacyl species which are intermediates in amide bond formation are capable of offloading biosynthetic intermediates from NRPSs. The chemical probes differ from the aminoacyl thioesters in the replacement of the thioester with an amide bond, making them resistant to hydrolysis and stable in culture extracts (Figure 37).

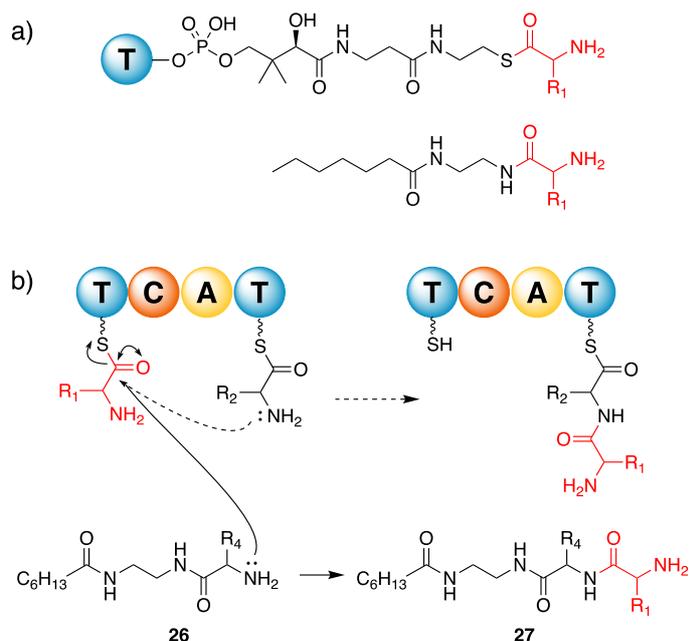


Figure 37 - a) Chemical probes for NRPSs designed to mimic PPant-bound aminoacyl thioesters. b) Chemical probes compete with the acceptor amino acids of the pathway to offload NRP intermediates which can be analysed by LC-MSⁿ.

The utility of these probes was demonstrated in the offloading of intermediates from the echinomycin biosynthetic pathway in *Streptomyces lasaliensis*. Varying the acyl chain length between C₄, C₇, and C₁₀ showed that while more hydrophobic C₁₀ chains were more effective in penetrating the cell, C₄ chains displayed lower toxicity to the strain, with a C₇ acyl chain being an effective compromise. A range of amino acids could be used, with probes bearing glycine or β-alanine particularly effective at capturing intermediates. This is possibly due to the lack of steric interaction deriving from the absence of a side chain, although a diverse range of side chains were able to capture intermediates from non-cognate modules.

1.6 Project Aims

The aim of the work described in this thesis was to prepare novel chemical probes for the investigation of nonribosomal peptide and hybrid polyketide-nonribosomal peptide biosynthetic pathways. The echinomycin system, for which the efficacy of chemical probes has already been briefly described above, will be used to examine the effect of structural changes on the original probes. While the

condensation domain demonstrates a substrate selectivity, discriminating between L- and D-amino acids, the requirement of the other functional elements such as the carbonyl group for amide bond formation has not been examined.

The NRPS and PKS-NRPS systems of vancomycin and colibactin will be interrogated with chemical probes to investigate the timing of biosynthetic events which occur on the assembly line. In vancomycin biosynthesis the timing of chlorination during assembly remains unelucidated, while in colibactin the timing of thiazoline oxidation in the generation of the complete colibactin has not been established. Additionally, necessary biosynthetic enzymes ClbO, a PKS, and ClbL, a putative amidase, had no assigned function. Investigation of the function of these enzymes by chemical probes will also be carried out.

In our model system of echinomycin biosynthesis, the engineering of the pathway *in vivo* will also be conducted with the aim to generate peptides chemoenzymatically through the reaction of chemical probes with the NRPSs and enrich offloading of biosynthetic intermediates by the inactivation of downstream modules. Reconstitution of NRPS modules *in vitro* will also be conducted to examine the formation of amide bonds between chemical probes and donor amino acids catalysed by the condensation domain. The substrate promiscuity observed *in vivo* will be examined to explore the rational engineering of NRPSs to generate novel NRPs.

The natural product scleric acid is produced by a previously silent cryptic gene cluster. The structure will be elucidated using chemical synthesis to produce synthetic standards of the compound and key biosynthetic intermediates. The NRPS module involved will also be examined using chemical probes.

2. Investigation of echinomycin biosynthesis

Echinomycin, **16**, shown in Figure 38, is a nonribosomally synthesised bis-intercalator depsipeptide. It was initially discovered in 1957 from the soil bacterium *Streptomyces echinatus*.⁸⁸ It has since been found to be produced by 4 other *Streptomyces* species, including *Streptomyces lasaliensis*.¹⁵⁴ A member of the quinomycin antibiotics, echinomycin consists of two quinoxaline chromophores attached to a cyclic pseudo-dimeric core, and it has been found to possess potent antitumour, antibacterial and antiviral activity.

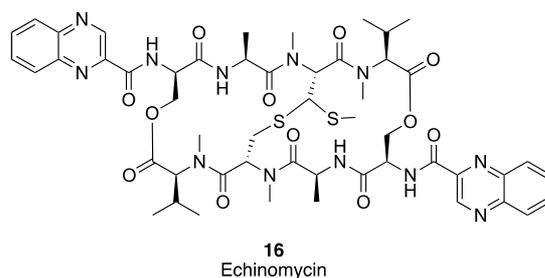


Figure 38 - Structure of the nonribosomal peptide echinomycin.

2.1 Echinomycin bioactivity

2.1.1 Antitumour activity

Echinomycin binds strongly to the minor groove of DNA, with the two quinoxaline moieties intercalating between base pairs, shown in Figure 39. It does not bind RNA and interacts weakly with single-stranded DNA. This binding causes the DNA helix to lengthen and unwind.¹⁵⁵ It can prevent transcription by binding to template DNA, inhibiting DNA-directed RNA synthesis.⁸⁹ This DNA binding activity is a significant contributor to antitumour activity, as it blocks chromatin condensation and DNA replication, disturbing the cell cycle and resulting in cell death.¹⁵⁶ DNA footprinting using methidiumpropyl-EDTA•Fe(II) has shown echinomycin binds DNA selectively around a 5'-ACGT-3' sequence, with the quinoxaline rings intercalating either side of the CG sequence.¹⁵⁷ The alanine residue in the backbone participates as the amide NH is hydrogen-bonded to the guanine N3, while the alanine carbonyl receives a hydrogen bond from the guanine N2.¹⁵⁸

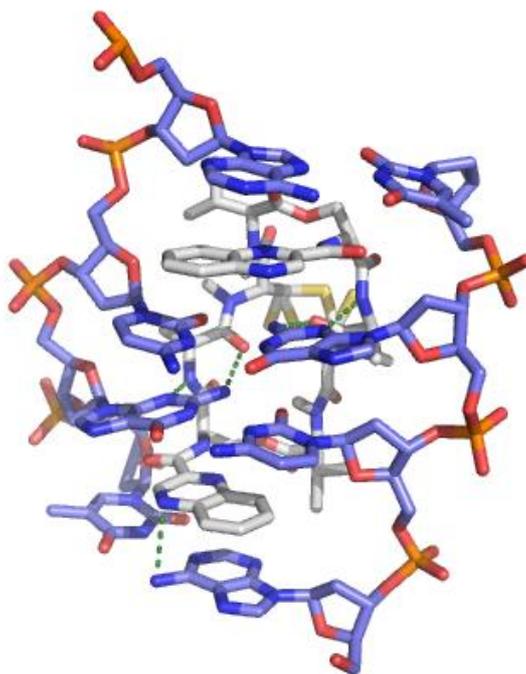


Figure 39 - Echinomycin binding DNA by intercalation of the quinoxaline rings, with hydrogen bonds between guanine bases and alanine nitrogen and carbonyl moieties shown in green.

Additionally, echinomycin has been shown to bind to DNA and thereby preventing interaction with transcription factors, giving a targeted inhibitor of certain downstream processes. Hypoxia-Inducible Factor-1 (HIF-1) is a transcription factor which controls genes involved in glycolysis, angiogenesis, migration and invasion, which are implicated in tumour progression and metastasis.¹⁵⁹ HIF binds to a hypoxia-responsive element (HRE) in DNA, which contains a core sequence of 5'R(A/G)CGTG-3'. Echinomycin can also bind to this HRE, inhibiting the binding of HIF-1, and may allow the control of HIF-1-dependent gene expression.¹⁶⁰

Twenty one DNA bis-intercalators have been discovered from Actinomycetes.¹⁶¹ The base pair specificity of these depends on the amino acid sequence and the central cross-bridge. For example, echinomycin binds a 5'-CG-3' sequence, while triostin A shown in Figure 40, which has an identical amino acid sequence but a central disulfide bridge compared to a shorter thioacetal, selectively binds 5'-AT-3'.

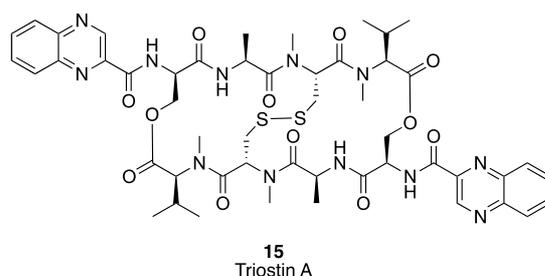


Figure 40 - Structure of triostin A , 15, containing a disulfide cross-bridge rather than the thioacetal of echinomycin.

2.1.2 Antibacterial activity

Echinomycin has also been investigated as an antibacterial agent. Tested against a range of clinical isolates of *Staphylococcus aureus*, and methicillin-sensitive (MSSA) and methicillin-resistant (MRSA), echinomycin was more effective than vancomycin, the current drug of choice.¹⁶² The MIC₉₀ and MBC₉₀ of echinomycin were lower than those of vancomycin, in spite of its hydrophobicity and short *in vivo* half-life, which limited anti-tumour clinical trials.¹⁶³ It has also shown great efficacy against clinical isolates of vancomycin-resistant enterococci (VRE), with an MIC₉₀ of 0.125 µg/ml.¹⁶⁴ Activity is also reported against the human pathogenic strains *Bacillus anthracis*, *Listeria monocytogenes*, *Streptococcus pneumoniae*, as well as the Gram-negative *Shigella dysenteriae*.¹⁶⁵ The method of antibacterial action is not currently known, but is thought to be linked to the DNA intercalating properties of echinomycin.

2.1.3 Antiviral activity

Echinomycin has also shown antiviral activity. Particularly, it has been shown to inhibit Tat transactivation, which is essential for HIV replication. Tat activates transcription of viral mRNA from the HIV long terminal repeat (LTR) and stabilises elongated transcripts.¹⁶⁶ Echinomycin has also been shown to inhibit the synthesis of viral proteins including viral haemagglutinin and neuraminidase by the influenza virus at low concentrations. At higher concentrations, the synthesis of all viral proteins was inhibited, while cellular protein synthesis was unaffected.¹⁶⁷

2.2 Echinomycin biosynthesis

Echinomycin is constructed by an iterative NRPS, utilising an unusual starter unit and further post-assembly modification to install the central cross-bridge. The BGC (Figure 41) in *Streptomyces lasaliensis* is found on the 520 kb giant linear plasmid pKSL. The functions of the genes are shown in Table 1. While some echinomycin BGCs are chromosomally encoded, this has been suggested as a factor in the spread of the compound between diverse microorganisms.¹⁵⁴

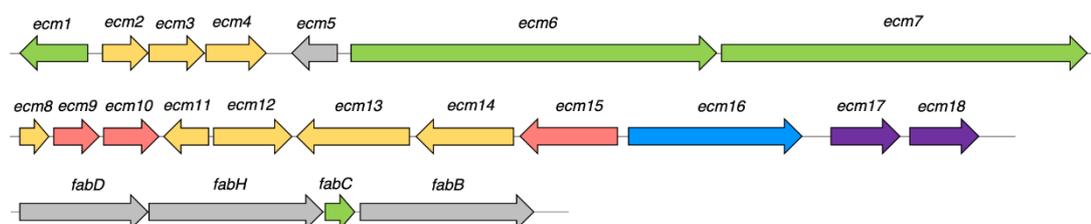


Figure 41 - The echinomycin gene cluster (*ecm*) and the standalone acyl carrier protein (ACP) from primary metabolism, *fabC*.

Gene	Function	Gene	Function	Gene	Function
<i>Ecm1</i>	Standalone A domain	<i>Ecm8</i>	MbtH-like protein	<i>Ecm15</i>	Transcriptional regulator
<i>Ecm2</i>	Type II thioesterase	<i>Ecm9</i>	Putative DNA-binding response regulator	<i>Ecm16</i>	Excinuclease ATPase
<i>Ecm3</i>	Putative dehydrogenase	<i>Ecm10</i>	TetR-like transcriptional regulator	<i>Ecm17</i>	Thioredoxin reductase
<i>Ecm4</i>	FAD-dependent oxidoreductase	<i>Ecm11</i>	Tryptophan 2,3-dioxygenase	<i>Ecm18</i>	SAM-dependent methyltransferase
<i>Ecm5</i>	Putative transposase	<i>Ecm12</i>	Cytochrome P450 dioxygenase	<i>FabC</i>	Fatty acid synthase ACP
<i>Ecm6</i>	NRPS	<i>Ecm13</i>	NRPS		
<i>Ecm7</i>	NRPS	<i>Ecm14</i>	Putative hydrolase		

Table 1 - Genes involved in echinomycin biosynthesis and their assigned or putative functions.

2.2.1 Quinoxaline-2-carboxylic acid biosynthesis

The quinoxaline chromophore essential for bioactivity had been found to be derived from L-tryptophan using ^{15}N NMR to identify the incorporation of labelled amino acids. Feeding with ^{15}N -L-tryptophan showed that the two nitrogen heteroatoms of the quinoxaline derived from the indole and amino groups, while using unlabelled tryptophan did not enrich the ^{15}N NMR signals of the quinoxaline.¹⁶⁸ Following the sequencing of the echinomycin BGC, a biosynthetic pathway to quinoxaline-2-carboxylic acid (Q2CA) could be proposed based on homology to genes of known function, shown in Figure 42.¹⁶⁹

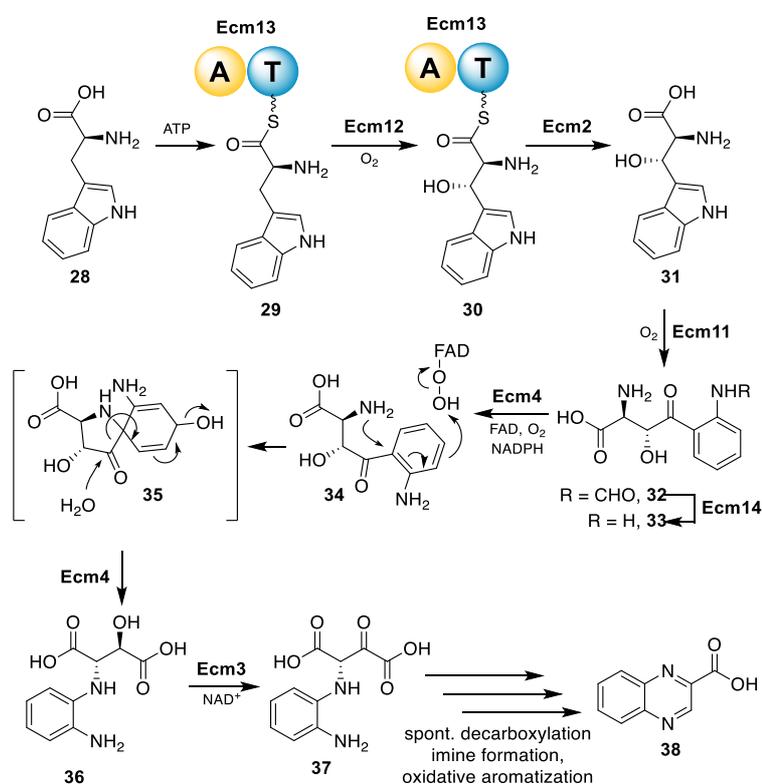


Figure 42 - Proposed biosynthesis of quinoxaline-2-carboxylic acid. The ring opening catalysed by Ecm14 is the latest step to be experimentally characterised.

It was proposed that L-tryptophan, **28**, is first loaded onto a thiolation domain of the A-T didomain of Ecm13. It is then hydroxylated in the β -position, as seen in the biosynthesis of nikkomycin.¹⁷⁰ The β -hydroxytryptophan, **30**, is then hydrolysed from the T domain by the thioesterase Ecm2, and the indole ring is opened by a tryptophan 2,3-dioxygenase Ecm11. This is similar to the action of tryptophan-2,3-dioxygenase in the bacterial anthranilate pathway.¹⁷¹ These initial steps have been demonstrated *in vitro*, while the stereochemistry of the β -hydroxytryptophan had already been determined.^{172,173} The product of the ring opening, 6'-3-hydroxykynurenine, **33**, has been shown to be incorporated into echinomycin by the feeding of a deuterated analogue, confirming its role as an intermediate in the biosynthesis.¹⁷⁴ 6'-3-hydroxykynurenine is also an intermediate in the biosynthesis of 3-hydroxyquinaldic acid, which is used in the related quinomycin-type antibiotic SW-163D, **39**, shown in Figure 43.

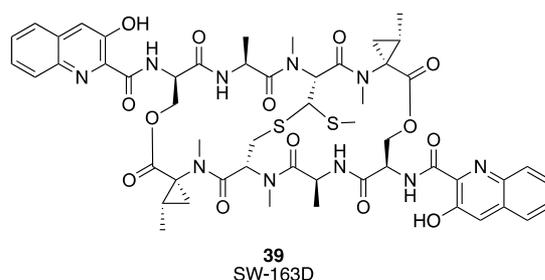


Figure 43 - Structure of quinomycin-type antibiotic SW163-D , containing a 3-hydroxyquinaldic acid chromophore.

However, the final steps of the biosynthesis, including the unprecedented oxidative cyclization and hydrolysis catalysed by Ecm4, have not been confirmed *in vitro*.

Aromatic acids are usually incorporated into NRPS by loading onto an aryl carrier protein (ArCP). In the echinomycin BGC there is no dedicated ArCP, but using cell extracts of the producer *S. echinatus* it was found that the ACP, FabC, from fatty acid biosynthesis is co-opted to carry the quinoxaline-2-carboxylic acid. It is loaded by a free-standing A domain, Ecm1, which will load only the cognate ACP or homologues. FabC has 98% identity to the fatty acid synthase (FAS) ACP of *S.*

coelicolor, which is also capable of binding Q2CA, while the ArCP from actinomycin biosynthesis is not.¹⁷⁵

2.2.2 Nonribosomal peptide synthetase

The echinomycin BGC contains two nonribosomal peptide synthetase genes. From the genetic sequence, the module and domain organisation can be predicted. Ecm6 is a bimodular NRPS, with a C-A-T-E-C-A-T domain structure, and Ecm7 is also bimodular with a C-A-M-T-C-A-M-T-TE domain structure. As the four NRPS modules are sufficient for the production of octapeptide echinomycin, it can be seen that the modules are used iteratively, each turning over twice per echinomycin molecule, as shown in Figure 44. Bioinformatic prediction also allowed the identification of the A domain specificity of modules 1, 3, and 4, which activate L-serine, L-cysteine, and L-valine respectively. The epimerisation domain of module 1 is responsible for the conversion of L-serine to D-serine, while the MT domains of module 3 and 4 methylate cysteine and valine. The C-terminal TE domain catalyses the homodimerisation and cyclo-release of the depsipeptide from the NRPS.⁹¹ This has been demonstrated *in vitro* using SNAC-tetrapeptides, which form the octapeptide dimer and the dilactone. Using an octapeptide-SNAC containing an ester linkage with the TE domain resulted in the formation of the dilactone.

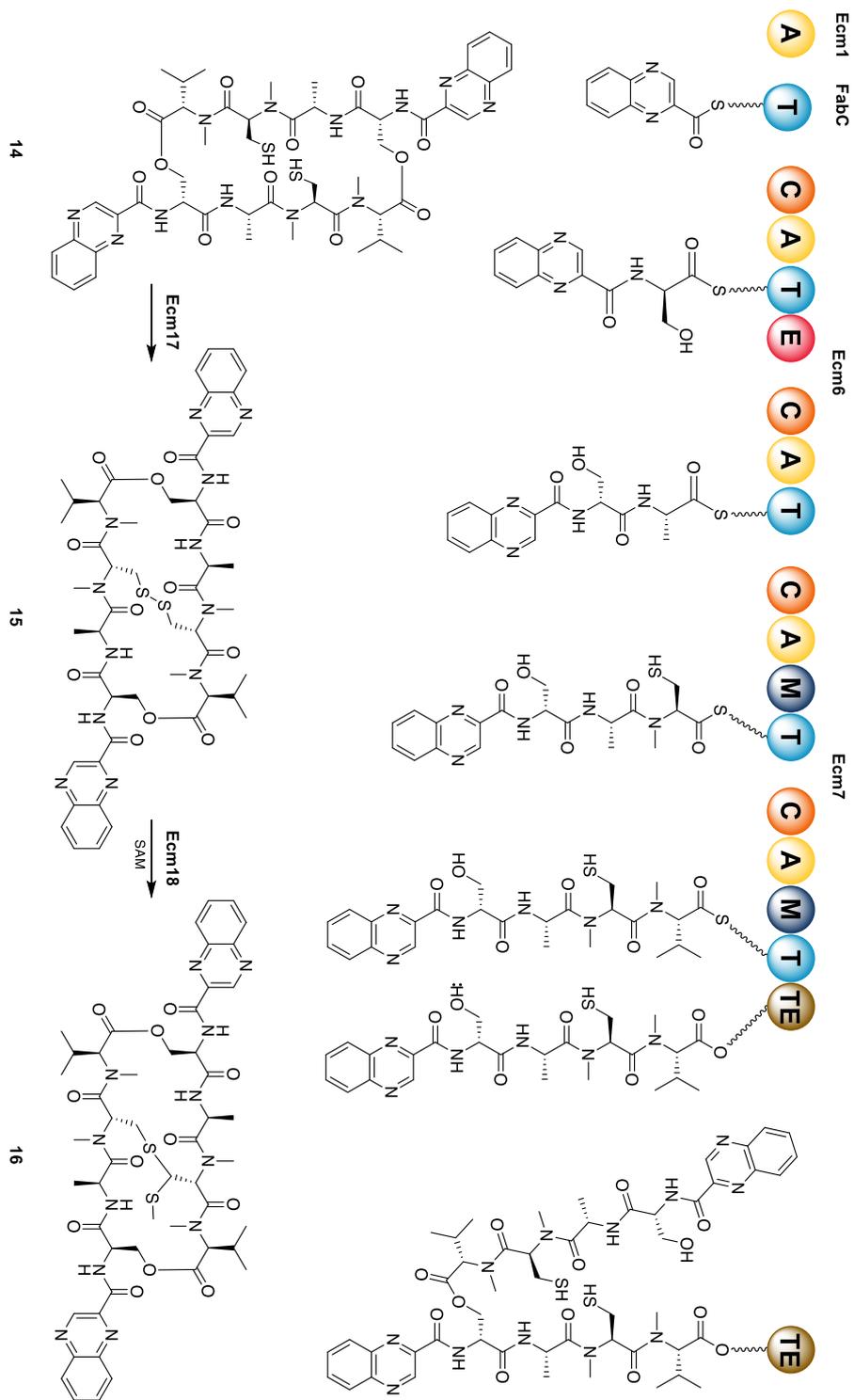


Figure 44 - Biosynthesis of echinomycin by an iterative NRPS , followed by post-assembly tailoring to install the central cross-bridge.

2.2.3 Post-assembly line tailoring

After cyclo-release of the depsipeptide from the NRPS, further tailoring enzymes are required to install the thioacetal cross-bridge. The disulfide bridge construction from the cysteine residues is catalysed by Ecm17, which has been shown to oxidise the dithiol of triostin A, **15**, *in vitro*. It is capable of oxidising the cysteines in the reducing cytoplasm. The disulfide species was hypothesized as the precursor to the thioacetal functionality, as *S. echinatus* protoplasts can convert these species to the thioacetal equivalent by methylation.¹⁷⁶ Ecm18 has high homology to a SAM-dependent methyltransferase, and purified Ecm18 incubated with triostin A and SAM can catalyse the conversion to a thioacetal *in vitro*. This is believed to proceed through a sulfonium ylide, with deprotonation followed by methyl transfer, illustrated in Figure 45.

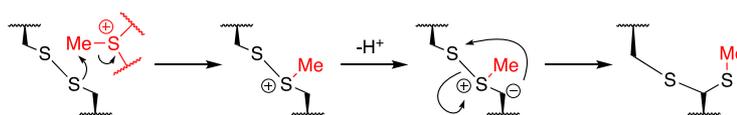


Figure 45 - Proposed mechanism of the conversion of the disulfide to the thioacetal of echinomycin catalysed by Ecm18, using SAM (shown in red).

2.3 Chemically probing echinomycin biosynthesis

Previous work in the Tosin group led to the development of small molecule chemical probes, shown in Figure 46, to intercept and offload intermediates in the biosynthesis of NRPs. These probes were used on echinomycin as a model system and showed the ability to offload peptide intermediates from the pathway to yield probe-intermediate conjugates as an unnatural peptide, as shown in Figure 47.

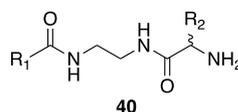


Figure 46 - Generic structure of a chemical probe for NRPS systems, in which the length of the acyl chain can be varied, and a range of amino acids used.

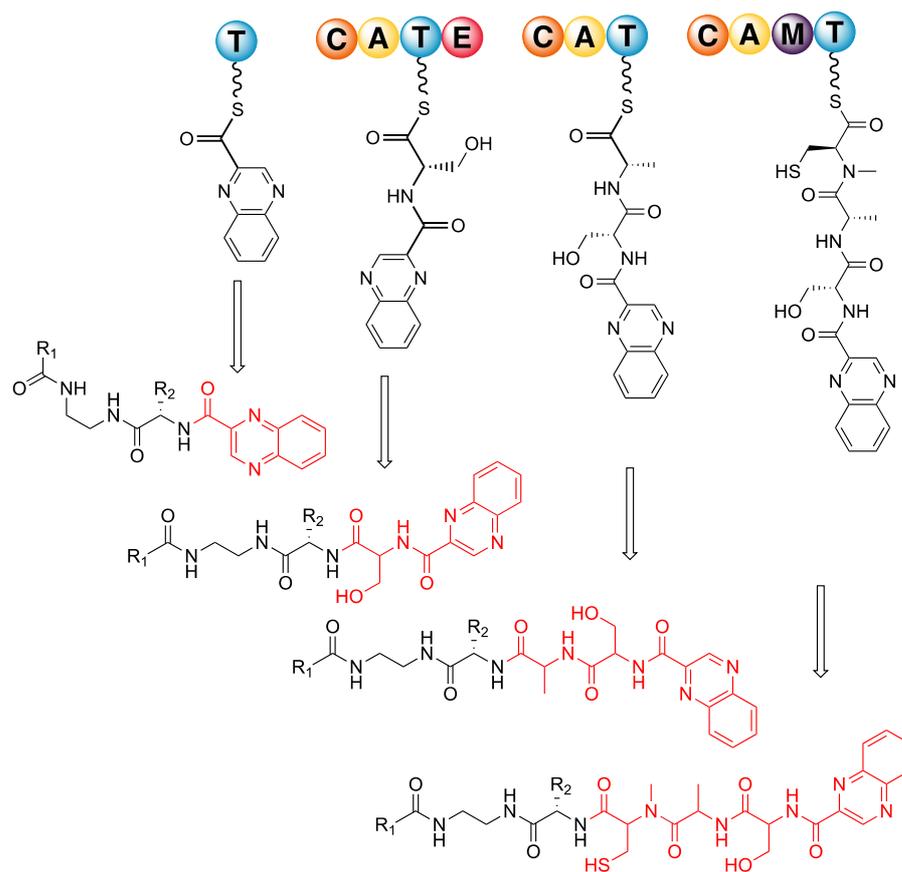


Figure 47 - Offloading of peptide intermediates of the echinomycin NRPSs Ecm6 and Ecm7 using chemical probes.

It was observed that a range of peptide intermediates could be offloaded by probes based on non-cognate amino acids, as well as the cognate amino acid of the module.¹⁵³ This suggested that the condensation domains may have a degree of inherent substrate flexibility, which may be possible to exploit to enzymatically synthesise unnatural peptides.

2.3.1 Thioamide-based chemical probes

The chemical probes are based upon a mimic of the phosphopantetheinyl arm of the T domain loaded with an aminoacyl unit. While structurally simpler than the PPant arm, the amide bond present in the probes would limit the diversity of peptides that could be synthesized using C domains as catalysts with these mimics.

To explore the tolerance of C domains at the acceptor site, it was decided to replace the amide bond of the pantetheine portion of the probe with a thioamide.

A thioamide functionality was chosen as they are considered bioisosteric replacements for amide bonds.¹⁷⁷ However, the greater size of sulfur compared to oxygen leads to a longer bond (1.65 Å to 1.20 Å). There is also a greater rotational barrier around the C-N bond.¹⁷⁸ These properties have been exploited in the design of peptidomimetics, where increased conformational rigidity can be desirable, while thioamides may also offer increased resistance to proteolysis.¹⁷⁹ Also, thioamides are a moiety that also participate in hydrogen bonding; they are weaker hydrogen bond acceptors and stronger hydrogen bond donors.

Thioamides are also a rare functional group in natural products. Recent work has expanded the number of thioamide structures to eleven, however this is a tiny fraction of natural products.¹⁸⁰ In known examples, the thioamide moiety provides functionality not offered by the analogous amide bond, such as in closthioamide, a potent antibiotic against MRSA and VRE, whereas the amide analogue has no antibiotic activity, both shown in Figure 48. The orthogonal chemistry that can be performed on thioamides in the presence of amides can also allow its use as a chemical handle, as exploited in the synthesis of an amidine analogue of vancomycin.¹⁸¹

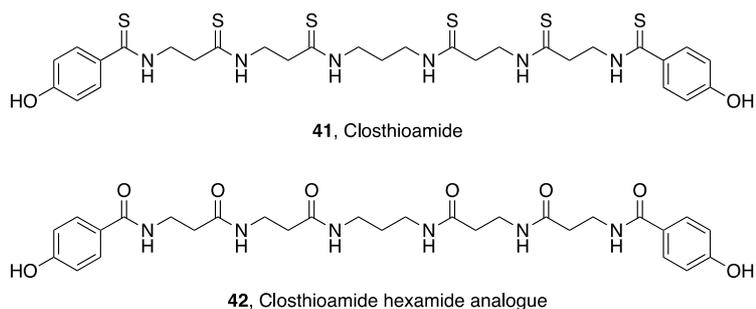
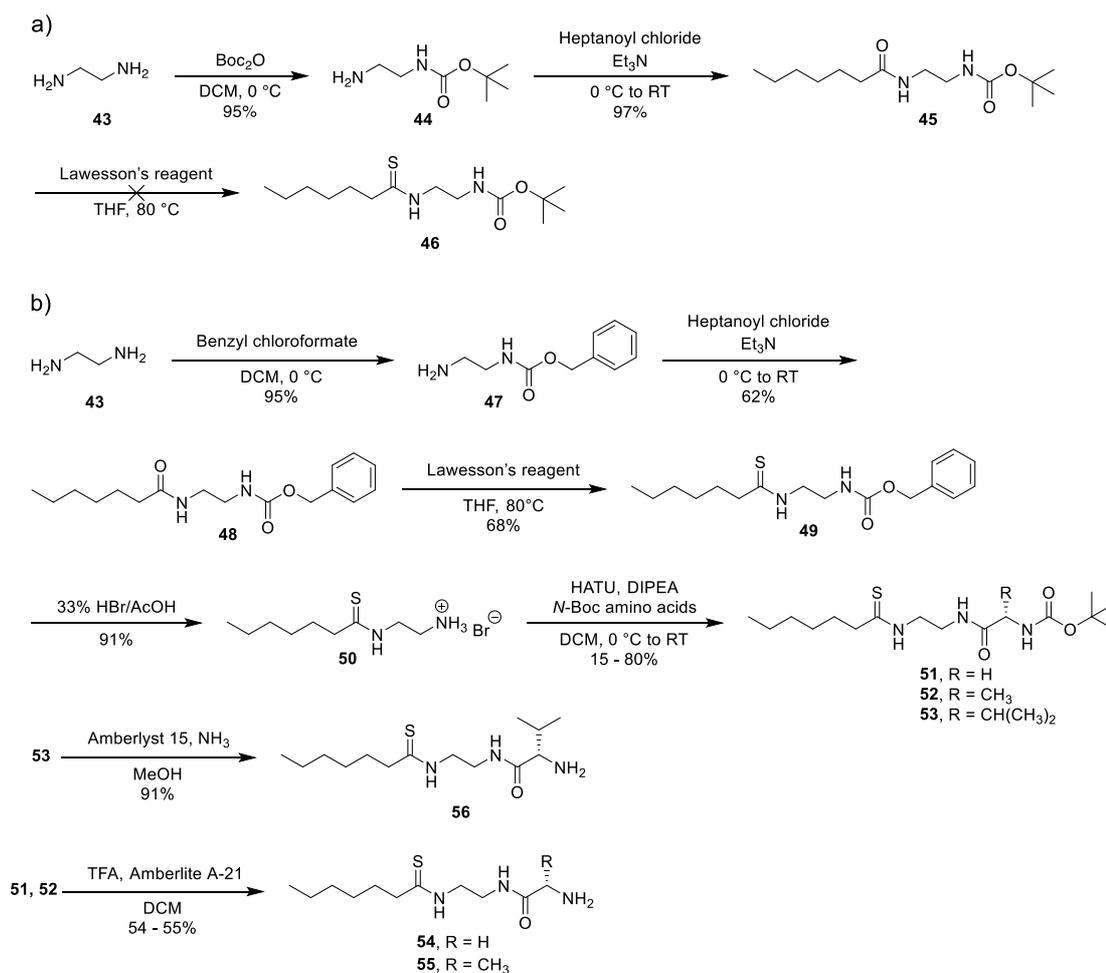


Figure 48 - Structure of thioamide-containing natural product closthioamide, 41, and its biologically inactive hexamide analogue, 42.

The synthesis of the thioamide probes was proposed to be accomplished by a similar method to the probes previously prepared, introducing the sulfur using Lawesson's reagent.¹⁸² As shown in Scheme 1, this was initially approached using *N*-Boc monoprotection of ethylene diamine, followed by acylation of the second amine with heptanoyl chloride. However, during the reflux of **45** with Lawesson's reagent, the Boc protection was cleaved, and none of the desired product was formed, shown in Scheme 1b. The use of heptanoyl chloride to afford the C₇ *N*-acyl chain was chosen based on previous work which showed that a longer *N*-decanoyl chain gave greater cellular uptake, but increased toxicity, while the opposite effect was observed for a shorter *N*-butyryl chain. *N*-heptanoyl chains offer a good compromise between these two extremes giving good cellular uptake and limited toxicity.¹⁵³



Scheme 1 - a) Synthetic route to thioamide containing probes based on glycine, alanine, and valine. b) Failed thioamidation of Boc-protected compound, 45.

Given the lability of the *N*-Boc protection, the Cbz protecting group was therefore employed. This should survive the extended elevated temperatures of the reaction with Lawesson's reagent carried out to convert the amide to a thioamide. This afforded the protected thioamide **49** in good yield. Deprotection of the Cbz-protected amine was not possible due to the poisoning of the Pd catalyst by sulfur and preventing the reaction. To circumvent this, HBr/ AcOH was used to remove the carboxylbenzyl group, to give the bromide salt **50** of the free amine. This could then be coupled to different Boc-protected amino acids using HATU. Coupling using HATU produces tetramethylurea (TMU) as a by-product, which is disadvantageous for use in biological experiments, as TMU is potentially toxic to bacterial cells.¹⁸³ This TMU impurity coeluted with the protected thioamide products under a range of

chromatographic conditions. Additionally, the standard method to remove Boc protection employs trifluoroacetic acid (TFA). This would produce the trifluoroacetate salt of the probe, which can be considerably more toxic than other counter-ions.¹⁸⁴ To resolve these issues, it was decided to remove the Boc group using the strongly acidic ion-exchange resin Amberlyst 15.¹⁸⁵ Here the sulfonic acid of the resin removes the Boc group, and sequesters the protonated amine through ionic interactions with the sulfonate. This can then be washed to remove impurities such as TMU, before elution using ammonia to give the free base of the thioamide valine probe **56**. However, this deprotection method did not scale well, so deprotection proceeded with TFA, and basic resin Amberlyst A-21 was used to give the free base of probes **54** and **55**. This strategy was used to synthesise probes containing glycine, alanine, and valine residues. In our previous work, glycine-based probes had been shown to offload a range of different intermediates of NRP biosynthesis, potentially due to the lack of steric bulk and chirality, allowing it to diffuse into the active site. Echinomycin modules 2 and 4 incorporate alanine and valine respectively, so would presumably interact well with these modules.

Probes **54** – **56** were supplemented to solid and liquid cultures of echinomycin-producer *S. lasaliensis* ACP12(S970A) to investigate potential NRP intermediate capture. This mutant strain cannot produce the polyether natural product lasalocid A, reducing the background of the organic extracts. Seed cultures were grown in M79 media at 30 °C for 3 days, and then used to inoculate MYM agar plates containing 0.2 – 2.0 mM of thioamide probes, **54** – **56**. These plates were then incubated for 5 days at 30 °C, before extraction using ethyl acetate (EtOAc). The organic extracts were concentrated and redissolved in methanol (MeOH), then run on an Orbitrap Fusion high resolution mass spectrometer. Strains grown in identical conditions, but in the absence of the probes were used as control experiments. The resulting data were inspected, searching for the masses of the putative intermediates of NRP biosynthesis intercepted by the thioamide probes, and MS² fragmentation of these parent ions used to support the structure.

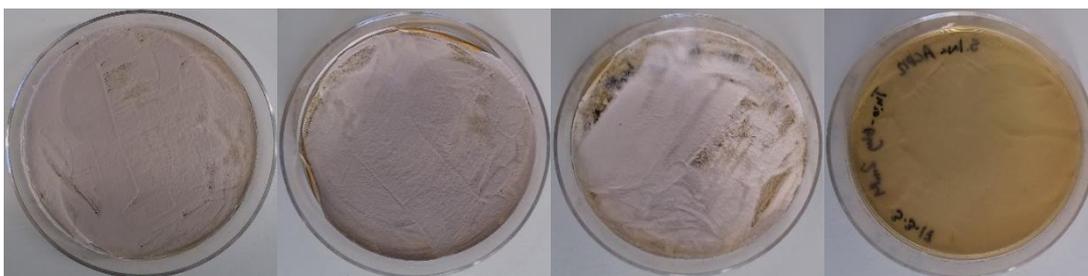


Figure 49 - Feeding of thioamide-glycine probe **54** to *S. lasaliensis* S970A at 0.2 mM (centre-left), 1.0 mM (centre-right), and 2.0 mM (right) concentrations, alongside a control plate (left).

The glycine-based thioamide probe displayed a concentration dependent toxicity towards the *S. lasaliensis*, with slightly reduced growth compared to the control plates at 0.2 mM. At 2 mM concentration there was essentially no growth of the strain, as seen in Figure 49. The reasons for this toxicity are unknown, but already demonstrate the significant effects of the bioisosteric substitution of the amide. The capture of NRP intermediates by the probes was hindered by this, as much less probe was available in growing *S. lasaliensis* cultures. Despite this, the feeding of probe **54** at 0.2 mM concentration did show the offloading of the quinoxaline-2-carboxylic acid starter unit, seen in Figure 50. Fragmentation of the ion produced daughter ions consistent with the offloading of the quinoxaline moiety by the probe containing a thioamide. Later stage intermediates could not be detected at 0.2 mM, and no intermediates were detected at higher concentrations.

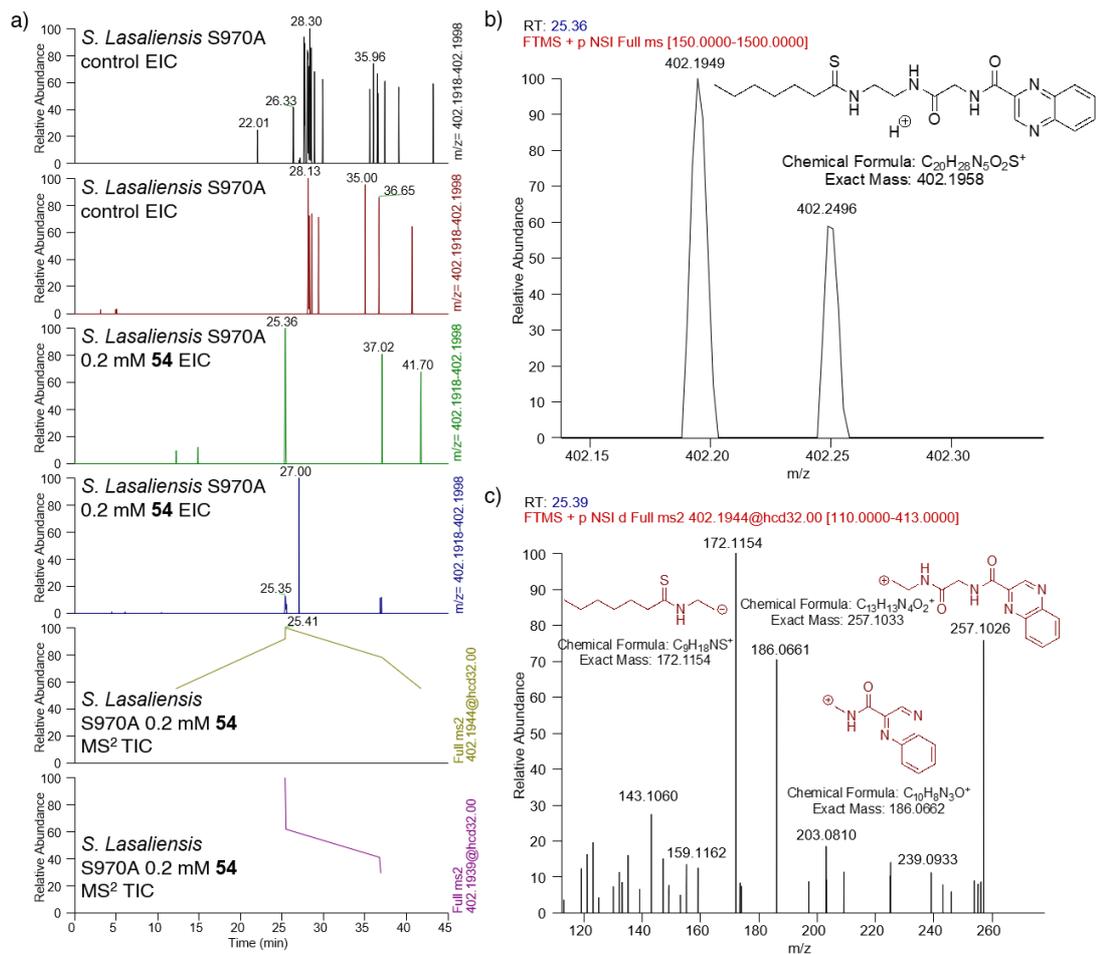


Figure 50 - Detection of the offloading of the quinoxaline-2-carboxylic acid starter unit from echinomycin NRPS using thioamide-glycine probe 54. Mass is detected in LC-HRMS at a retention time of 25.4 mins (a), with high resolution mass (b) and fragments from higher-energy collision dissociation (HCD) to support the structure (c).



Figure 51 - Feeding of thioamide-alanine probe 56 to *S. lasaliensis* S970A at 0.2 mM (centre-left), 1.0 mM (centre-right), and 2.0 mM (right) concentrations, alongside a control plate (left).

The alanine-based probe **55** was even more toxic to *S. lasaliensis* than the glycine-based probe, seen in Figure 51. It showed a small amount of growth over 5 days at 0.2 mM, and no notable growth at 1.0 or 2.0 mM concentrations. This poor growth clearly reduces the number of viable cells which could be expressing NRPSs, and as such, no putative intermediates could be detected using probe **55**.



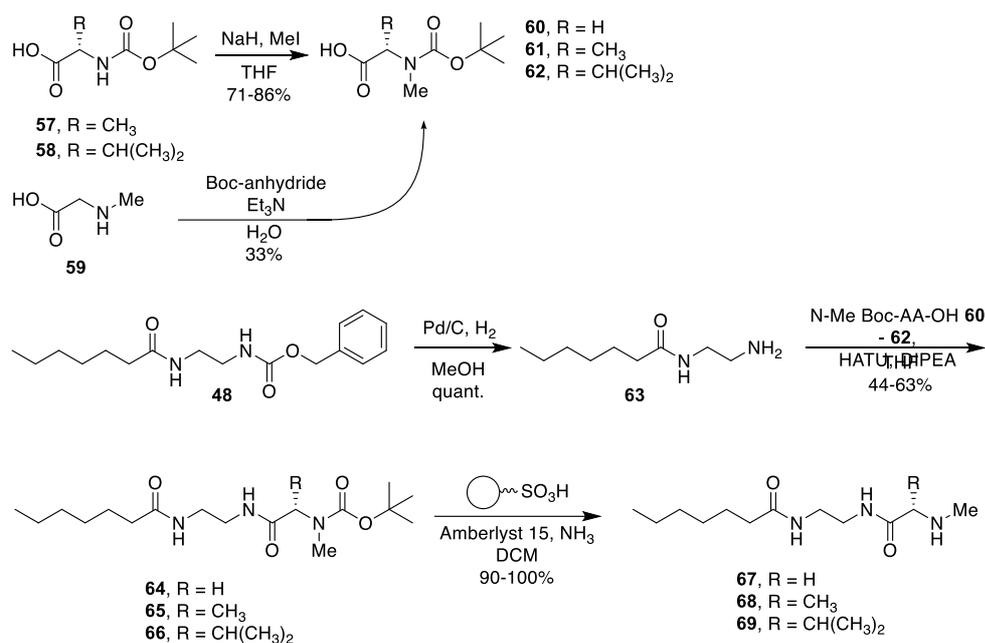
Figure 52 - Feeding of thioamide-valine probe 56 to *S. lasaliensis* S970A at 0.2 mM (centre-left), 1.0 mM (centre-right), and 2.0 mM (right) concentrations, alongside a control plate (left).

Valine-based thioamide probe **56** was also highly toxic to *S. lasaliensis*, almost completely inhibiting growth at a 0.2 mM concentration, shown in Figure 52. Again, no intermediate species captured by the probe could be observed. Given the low abundance of the quinoxaline intermediate captured by **54** at 0.2 mM concentration, it was decided not to lower the concentration of the probes further. Any increase in growth would be a trade off against decreased abundance of the probe and any related species.

2.3.2 *N*-methylated chemical probes

Another set of novel probes were also designed to test the substrate tolerance of the acceptor position of the condensation domains of the echinomycin NRPS,

containing *N*-methylated amino acids. *N*-methylation is catalysed by MT domains in NRPS, and often confers beneficial properties to the natural products, such as enforcing bioactive conformations on the compounds. Echinomycin contains four *N*-methylated residues, with both cysteines and both valines being modified by the MT domains in modules 3 and 4 respectively. With the exception of a small number of fungal NRPSs, A domains exclusively active unmethylated amino acids, which are methylated when bound as aminoacyl thioesters on the NRPS.¹⁸⁶ The methylation state of the amino acid can then affect the peptide bond formation by the condensation domain, possibly due to steric factors affecting substrate positioning in the active site rather than any inherent selectivity.¹⁸⁷ The natural use of an *N*-methyl amino acid in the condensation domains of modules 3 and 4 suggest these domains may tolerate *N*-methylated probes. The condensation domain of module 1 showed a wide substrate tolerance to non-cognate and unnatural amino acids in previous work, so methylated probes would allow further exploration of the substrate tolerance of this domain.



Scheme 2 - Synthetic route to *N*-methylated chemical probes based on glycine, alanine, and valine.

The *N*-methylated probes were synthesized according to Scheme 2. A previous intermediate **48** was deprotected by hydrogenolysis and the free amine, **63**, coupled to *N*-methylated amino acids. These were prepared by methylation of Boc-amino acids using sodium hydride and methyl iodide in the case of L-alanine and L-valine. To

prepare *N*-Me-Gly, **60**, the secondary amine of commercially available sarcosine was protected using Boc anhydride. These Boc-protected probes **64** – **66** were then deprotected using Amberlyst 15 resin as described for the thioamide probes.



Figure 53 - Feedings of *S. lasaliensis* S970A with *N*-methylated Gly probe 67 at concentrations of 0.2 mM (centre-left), 1.0 mM (centre-right), and 2.0 mM (right), alongside a control plate (left).

As seen in Figure 53, *N*-Me-glycine probe **67** displayed no toxic effect on the growth of *S. lasaliensis* at the tested concentrations of 0.2 mM, 1mM, and 2.0 mM. No NRP intermediates were captured by the probe though. This could be due the poor positioning in the condensation domain active site as a result of methylation, preventing peptide bond formation. Alternatively the steric bulk of the methyl group may reduce the rate of the condensation reaction, as is observed in synthetic coupling procedures.¹⁸⁸



Figure 54 - Feedings of *S. lasaliensis* S970A with *N*-methylated Ala probe 68 at 0.2 mM (centre-left), 1.0 mM (centre-right), and 2.0 mM (right), alongside a control plate (left).

Feeding the *N*-Me-alanine probe **68**, shown in Figure 54, also resulted in a small decrease in the growth of *S. lasaliensis* at all concentrations from 0.2 mM to 2.0 mM compared to control plates. Again, analysis of organic extracts of the cultures did not reveal any intercepted intermediates when analysed.



Figure 55 - Feedings of *S. lasaliensis* S970A with *N*-methylated Val probe **69 at 0.2 mM (centre-left), 1.0 mM (centre-right), and 2.0 mM (right), alongside a control plate (left).**

The *N*-Me-valine probe **69**, shown in Figure 55, also had a small detrimental effect on bacterial growth at all concentrations. While the intact probe could be detected in the organic extracts by mass spectrometry, no intermediates could be observed. This is surprising given previous described effectiveness of a valine-based probe to offload NRP intermediates.¹⁵³ *N*-methyl-L-valine is the cognate acceptor amino acid in module 4 of the NRPS Ecm7, which has previously been offloaded from, and would be expected to yield intermediates in this case. Given the positioning of the *N*-Me-valine probe should mimic the cognate substrate, it may suggest that a reduced rate of reaction is the reason for the lack of intermediates observed with *N*-methylated probes.

2.3.3 Cysteine-based chemical probe

Earlier work on chemical probes for NRP biosynthesis in the group has synthesized and tested three of the four cognate amino acids of echinomycin biosynthesis, using probes based upon L-serine, L-alanine, and L-valine. However, the synthesis of a probe based upon the module 3 cognate substrate, cysteine, had been problematic. Deprotection of the *S*-trityl and *N*-Boc of **70**, previously prepared by Candace Ho, a Masters student in the Tosin group, using TFA and triethylsilane (TES) resulted the in formation of a minority of the desired monomer, **71**, in 10% yield, shown in Figure 56, and mostly in the unwanted disulfide species **72**.

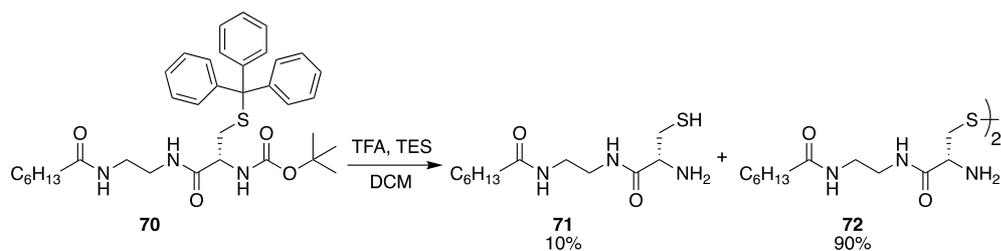
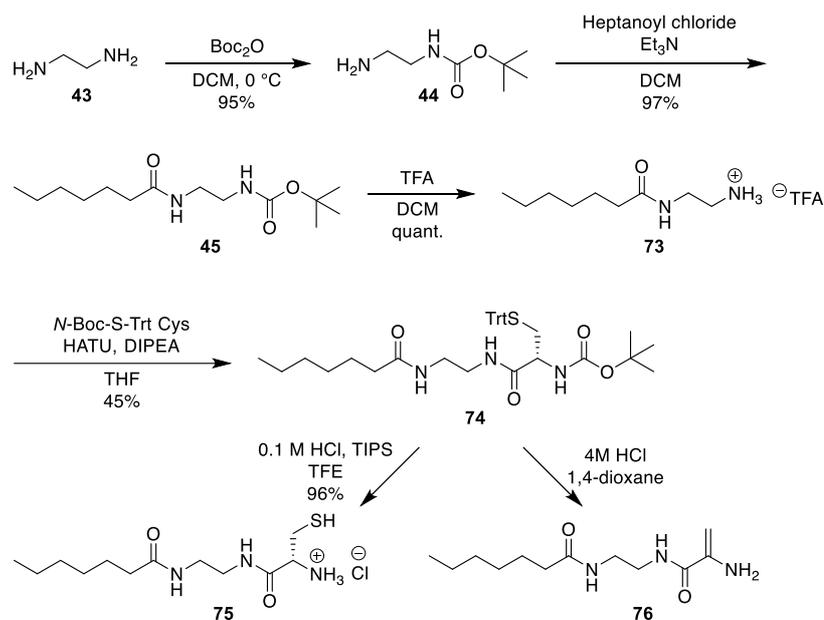


Figure 56 - Previously performed deprotection of a cysteine based chemical probe using TFA.

It was unclear whether this disulfide species would have trouble entering the bacteria due to containing multiple amines, which would be protonated at physiological pH. Additionally, this probe may not be reduced intracellularly to the desired monomer which would act as a probe for NRPS biosynthesis. A new synthetic procedure was therefore needed to produce a monomeric cysteine-based probe in high yield. Subsequently, we undertook the preparation of an *N*-heptanoyl cysteine probe according to Scheme 3.



Scheme 3 - a) Synthetic route to a cysteine based chemical probe. b) Attempted deprotection of *S*-Trt and *N*-Boc groups using HCl in dioxane.

Common intermediate **73** was prepared by monoprotection of ethylene diamine, acylation with heptanoyl chloride, and deprotection using trifluoroacetic acid. *S*-trityl-*N*-Boc cysteine was coupled using HATU to afford protected probe **74**.

Attempts were made to deprotect **74** using Amberlyst 15 resin, followed by elution with ammonia. Under these basic conditions, the disulfide was rapidly formed and was the only product isolated. While the use of reducing agents dithiothreitol (DTT) and tris(2-carboxyethyl)phosphine (TCEP) to form the monomer **75** was explored, the separation of these reducing agents and the probe proved difficult, and left the possibility of oxidation to the disulfide after the removal of DTT or TCEP.

As the oxidation of cysteine residues to disulfides is reduced at lower pHs, the basic workup to remove the TFA anion from a deprotected cysteine probe was undesirable. To be able to use the probe in its acidic form, a biocompatible anion such as acetate or chloride was needed. To this end, the simultaneous deprotection of the *S*-trityl and *N*-Boc groups by HCl was investigated. As shown in Scheme 3, the removal of the both protecting groups could be achieved using 4M HCl in dioxane, but elimination of H₂S was observed before the complete removal of the *S*-trityl group, shown in Scheme 3, **76**. Milder conditions were needed to minimize the elimination of the thiol group, while more effectively cleaving the trityl group. The use of mildly acidic trifluoroethanol (TFE), has been reported to enhance the removal of acid-labile protecting groups, enabling the concentration of HCl to be greatly reduced.¹⁸⁹ The use of 0.1 M HCl in TFE, with triisopropylsilane (TIPS) to scavenge the trityl cation allowed the direct synthesis of the hydrochloride salt of the cysteine-based probe **75**, shown in Scheme 3. The lack of the disulfide species was confirmed by ¹³C NMR, where the carbon adjacent to the thiol should have a shift of around 28 ppm and a carbon adjacent to a disulfide would have a shift around 43 ppm.

The cysteine-based probe, **75**, was fed to *S. lasaliensis* S970A at concentrations of 1.0 mM and 2.0 mM, pictured in Figure 57. However, no intercepted intermediate species could be observed. This could potentially be due to a number of factors. The reduced growth due to the toxicity of the probes to the strain will reduce the number of cells which can produce echinomycin, while any retardation of cell growth could affect the onset of production of secondary metabolites. Additionally, the effective concentration of probe in the cell will be further reduced by the formation of the disulfide species *in vivo*, which can be observed in the mass spectra of EtOAc extracts of the samples. Furthermore, any intermediates which are offloaded by the

probe bearing a free thiol could potentially undergo oxidation afterwards, forming mixed disulfides and reducing the amount of any offloaded intermediates available to be detected.

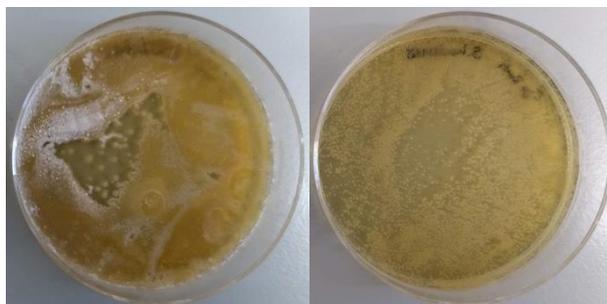


Figure 57 - Feedings of *S. lasaliensis* S970A with Cys probe 75 at 1.0 mM (left), and 2.0 mM (right).

The cysteine probe has also been used to examine the biosynthesis of colibactin, in chapter 3.

2.4 Genetic manipulation of *S. lasaliensis* with CRISPR-Cas9

Given the previously shown ability of chemical probes to interact with a range of condensation domains in the echinomycin NRPSs, it was hoped that this apparent substrate flexibility could ultimately be exploited to generate novel analogues of echinomycin. In general, C domains of the echinomycin assembly line appear to tolerate a wider range of acceptor amino acids than those activated by the adenylation domain of the same module. With a model of the specificity conferring residues of the adenylation domain established, a rational mutagenesis strategy could be used to reprogram A domains to activate a non-cognate amino acid which could be utilised by the C domain, as has been reported in the engineering of calcium dependent antibiotic (CDA).¹⁹⁰

Towards this aim, it was decided to use CRISPR-Cas9 to edit the BGC of echinomycin in *S. lasaliensis*. Initial work focused on the in-frame deletion of the catalytic HHxxxDG motif of the C domains of modules 1 and 2 of *Ecm6*. These would create a strain incapable of producing echinomycin and which would serve as useful controls to show that the probes are offloading intermediates using the C domain. A CRISPR-Cas9 strategy was chosen due to the relatively quick process compared to a traditional double-crossover mutation. The double-crossover approach requires a

single-crossover after conjugation, and then potentially multiple rounds of culture in non-selective conditions to allow the second crossover, before screening to determine which colonies have undergone a second crossover event or simply reversed the original crossover. In contrast, the exconjugants from the CRISPR-Cas9 approach should contain the desired mutation with a high efficiency, without the need for extensive re-plating and counter-selection.

The CRISPR-Cas9 system chosen for use was the pCRISPomyces-2 system.¹²⁷ This plasmid system encodes a codon-optimised *cas9* gene from *S. pyogenes*, which otherwise contains a number of TTA codons which can be transcription-level regulators in *Streptomyces* species.¹⁹¹ The 20 bp protospacer containing a 3'-protospacer adjacent motif of NGG can be introduced into the synthetic guide RNA (sgRNA) sequence using Golden Gate assembly. The pCRISPomyces-2 plasmid can also incorporate homology arms to promote homology-directed repair (HDR) of the Cas9 induced double-strand break, rather than non-homologous end joining (NHEJ).

While the full genomic sequence of *S. lasaliensis* was not available, the gene cluster of echinomycin has been previously published.¹⁹² From this, protospacers could be designed to target the condensation domains of modules 1 and 2 of *Ecm6*. These contained 15 bp sequences (12 bp + PAM) which were unique amongst the cluster, to minimise off-target effects of the Cas9.

Condensation domain 1

gRNA P1: 5' - ACGCGGACGCCGAACGCGTCGGTG-3'

gRNA P2: 5' - AAACCACCGACGCGTTCGGCGTCC-3'

Condensation domain 2

gRNA P1: 5' - ACGCGCCCATCACGTGCTGCTCGA-3'

gRNA P2: 5- AAACTCGAGCAGCACGTGATGGGC-3'

Figure 58 - Oligonucleotide sequences used to introduce the gRNA construct into pCRISPomyces2 for each condensation domain of *Ecm6*.

The two oligonucleotides which composed each protospacer, shown in Figure 58, were annealed. The protospacers were introduced to the pCRISPomyces-2 plasmid via Golden Gate assembly using the BpiI sites present, as in Figure 59.

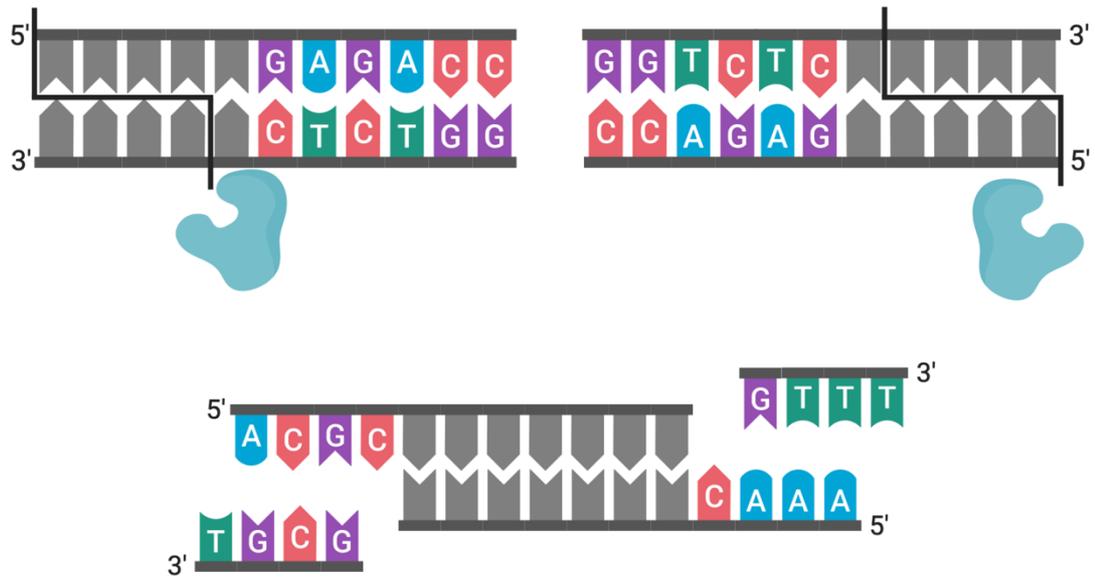


Figure 59 - Golden Gate assembly process - Type II restriction enzymes can cleave outside their recognition site, which can be exploited to excise the restriction site from the vector, preventing digestion of a ligated insert. The resulting overhang can be used to anneal an insert with complementary overhangs using DNA ligase in a single process.

The resulting plasmid was then transformed into *E. coli* DH5a using electroporation and selected for using apramycin. The resulting colonies were screened by restriction digest using BpiI, the restriction digest of which should have been excised during Golden Gate assembly, and by sequencing as shown in Figure 60.

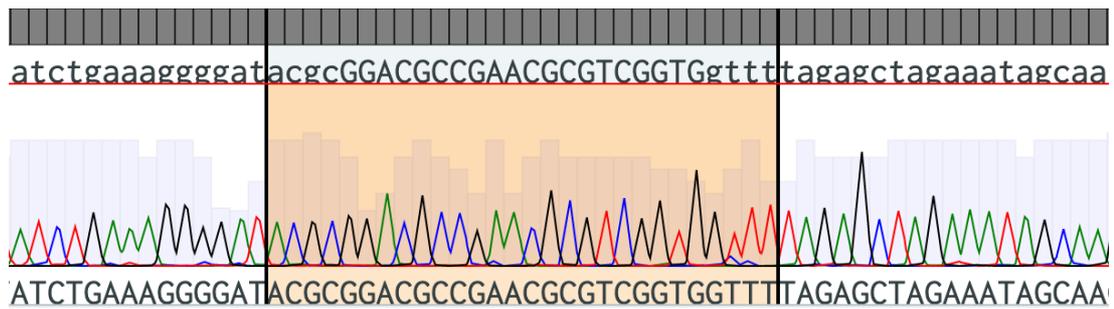


Figure 60 - Sequencing confirms the correct insertion of the protospacer DNA for condensation domain 1 using Golden Gate assembly.

The correct plasmids were linearised using XbaI and dephosphorylated using Shrimp Alkaline Phosphatase to prevent re-ligation during Gibson assembly. Two 800 bp homology arms were cloned from genomic DNA for each construct, to enable selective deletion of the 7 bp HHxxxDG motif using HDR. Gibson assembly, illustrated in Figure 61, was then used to introduce the homology arms to the pCRISPomyces-2 plasmids. This mixture was then used to transform chemically-competent *E. coli* TOP10, with transformants selected for using apramycin. The plasmids were then isolated from transformants, and the success of the Gibson assembly confirmed by sequencing and digestion with KpnI. These showed the plasmids to be as desired, with pCRISPomyces-2-domain2 containing a single silent point mutation.

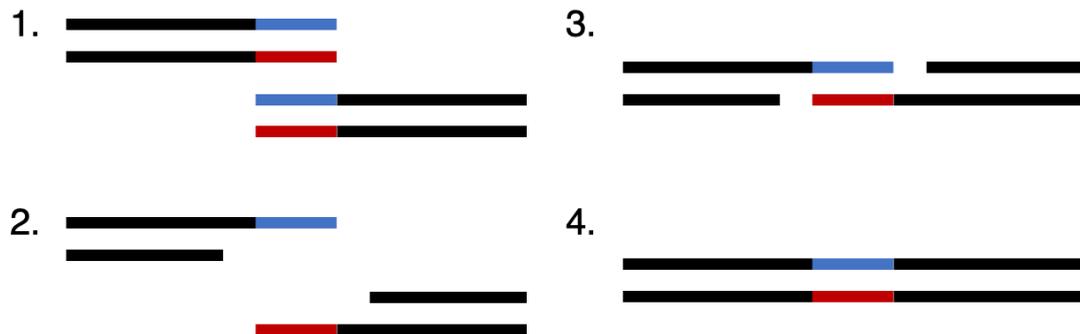


Figure 61 - Gibson assembly process: 1) Double-stranded DNA with overlapping ends are used. 2) 5'-exonuclease chews back the 5' ends of the DNA. 3) The complementary 3' ends can anneal together, and DNA polymerase extends the 3' ends. 4) DNA ligase can seal the nicks in the double-stranded DNA.

The plasmids were then transformed into the strain *E. coli* ET12567/pUZ8002 for conjugation, and selected for using apramycin, chloramphenicol, and kanamycin. *S. lasaliensis* was grown on SFM agar plates for 5 days, until sporulating. The spores

were harvested using distilled water, and conjugation was attempted. Despite repeated attempts to introduce the pCRISPomyces-2 constructs into *S. lasaliensis* using varying amounts and ratios of *E. coli* and *Streptomyces* Table 2, no exconjugants were produced.

Amount of spores / μL	Amount of 2xYT / μL	Spore concentration $\mu\text{L}/\text{mL}$
10	500	19
20	500	38
25	500	48
50	500	90
20	300	63
40	300	118
40	200	167
80	200	286

Table 2 - Amounts and concentrations of *S. lasaliensis* spores used to attempt to conjugate with *E. coli* ET12567 pUZ8002 pCRISPomyces-2 constructs.

While the pCRISPomyces-2 system has shown success in the genome editing of *S. albus* and *S. lividans* amongst others, it has been unsuccessful when using with *S. coelicolor M1152* and *Streptomyces* sp. KY 40-1.¹²⁴ In the case of *S. lasaliensis* it is possible that the failure is in the transfer of the plasmid between *E. coli* and *S. lasaliensis* resulting in a lack of antibiotic resistance to apramycin when overlaid. Alternatively, the double-strand breaks induced by the Cas9 protein could be toxic to the strain. As the full genome sequence is not available, it is possible that there are no off-target effects within the BGC, but off-target nuclease activity may affect other regions of the genome, potentially including essential genes, causing toxicity.

To examine this, a conjugation was attempted using a second system, CRISPR/Cas9-CodA(sm).¹⁹³ Before generating a construct for a specific genome deletion in this plasmid which expresses *S. pyogenes* Cas9, conjugation was performed using the unmodified plasmid. Once again, no exconjugants were obtained, suggesting that the Cas9 protein, which is constitutively expressed in both systems, is toxic to *S. lasaliensis*.

2.5 TAR capture of the echinomycin BGC

With the failure of the CRISPR-Cas9 approach in *S. lasaliensis*, it was decided to attempt the capture of the BGC of echinomycin and introduce it into a heterologous host. This approach has been recently used in the Corre lab for the characterisation of scleric acid biosynthesis, whereby the BGC was captured from an intractable host and genome edited using CRISPR-Cas9 in a heterologous host.¹⁹⁴

To begin, cluster specific capture vectors were constructed from the vectors pCAP03 and pCAP1000. pCAP03 uses 60 bp homology arms from *S. lasaliensis* genomic DNA to capture the cluster *via* homologous recombination in yeast. The pCAP03 vector was linearized using NdeI and XhoI, and a 128 bp sequence of synthetic DNA, shown in Figure 62, was introduced using Gibson assembly to give the pCAP03 capture vector. The correct assembly of the vector was confirmed by restriction digest and sequencing.

gBlock for pCAP03 (168 bp)

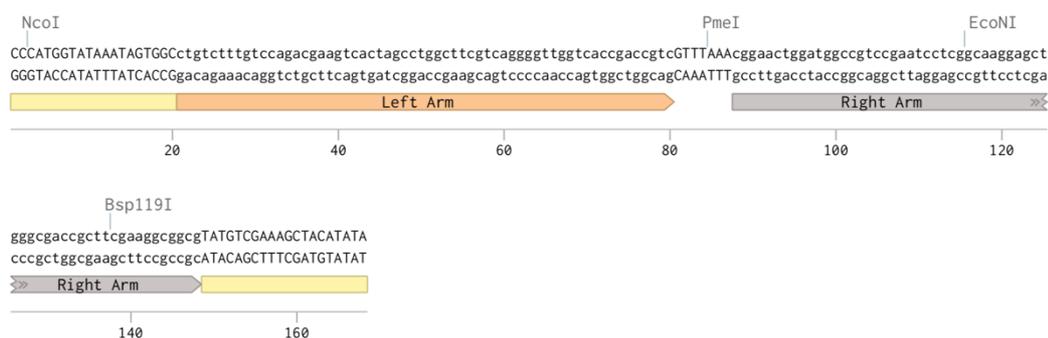


Figure 62 - Sequence for the synthetic DNA used to create the pCAP03 capture vector. The 20 bp regions of homology required for Gibson assembly are shown in yellow, the homologous arms for recombination in yeast are shown in orange and grey, flanking the PmeI cut site.

This 128 bp region contained two 60 bp homology arms from the termini of the echinomycin BGC, flanked by 20 bp regions of homology with the vector to facilitate Gibson assembly. The 60 bp homology arms were separated by a unique PmeI restriction site to allow the vector, shown in Figure 64, to be linearized in a selective manner. pCAP03 contains a *URA3* gene, encoding *URA3* which converts 5-fluoroorotic acid (5-FOA), **77**, to the toxic anti-metabolite 5-fluorouracil, **78**, of uracil, **79**, as shown in Figure 63. *URA3* is under the control of the strong promoter *ADHI*

from *Schizosaccharomyces pombe*, which allows up to 130 bp between the TATA box, and the initiation of transcription. Using small homology arms allows the selection against vectors which have not undergone homologous recombination, and therefore still express URA3 and are sensitive to 5-FOA. However, the shorter arms can lead to false positives where homologous recombination has occurred, but not within the BGC of interest.

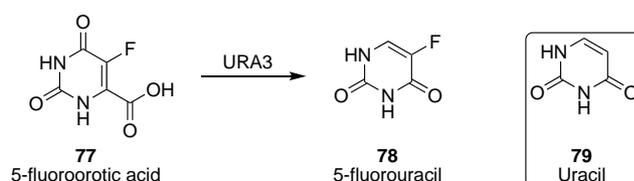


Figure 63 - The action of URA3, converting 5-fluoroorotic acid, 77, to 5-fluorouracil, 78, an anti-metabolite of uracil, 79.

pCAP1000 uses 1000 bp homology arms to give a greater chance of recombination occurring within the desired region of the genome containing the BGC. These 1000 bp arms, which precede and succeed the echinomycin cluster, were cloned from the genomic DNA of *S. lasaliensis*, and their identities confirmed by sequencing. The primers contain 25 bp regions of homology to the vector to facilitate Gibson assembly. The pCAP1000 vector was digested using SpeI and XhoI, to give 2 DNA fragments, which were purified by agarose gel electrophoresis. The two fragments and the two homology arms were assembled using Gibson assembly to give the pCAP1000 capture vector, shown in Figure 65. The correct assembly of the vector was confirmed by restriction digest and sequencing. In pCAP1000, the homology arms flank the *pADHI* and *URA3* genes, and therefore if the vector undergoes homologous recombination these genes will be lost completely, rendering the transformant colonies insensitive to 5-FOA. The pCAP1000 capture vector was similarly linearised by digestion with PmeI.

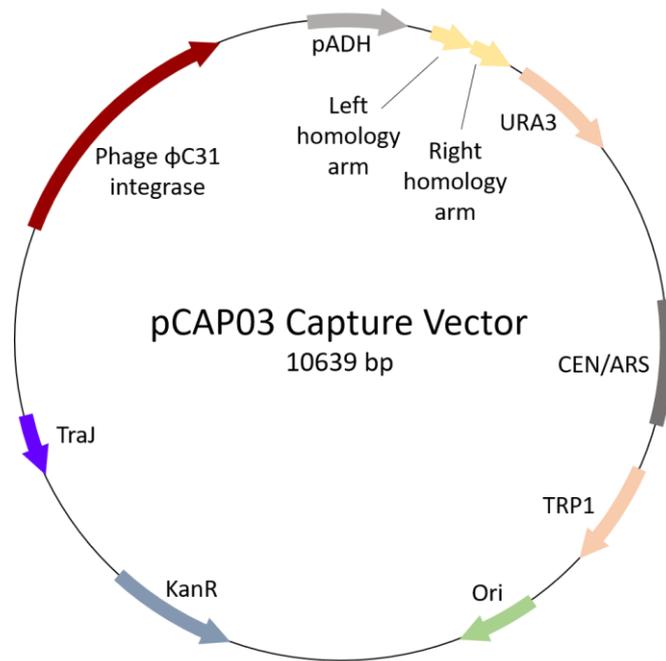


Figure 64 - Plasmid map of the pCAP03 capture vector , containing homology arms for the *Ecm* cluster between the promoter *pADH* and the *URA3* gene.

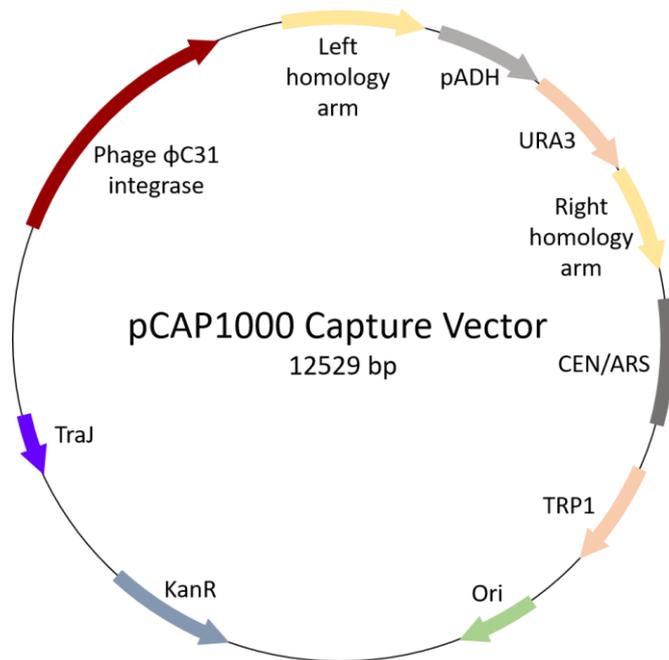


Figure 65 - Plasmid map of the pCAP1000 capture vector , containing homology arms for the *Ecm* cluster flanking the *pADH* and *URA3* genes.

Genomic DNA of *S. lasaliensis* was digested using NsiI and MseI, which have restriction sites adjacent to the echinomycin BGC, but not within it, as illustrated in Figure 66. This reduces the size of the genomic DNA so it can be transformed into yeast.

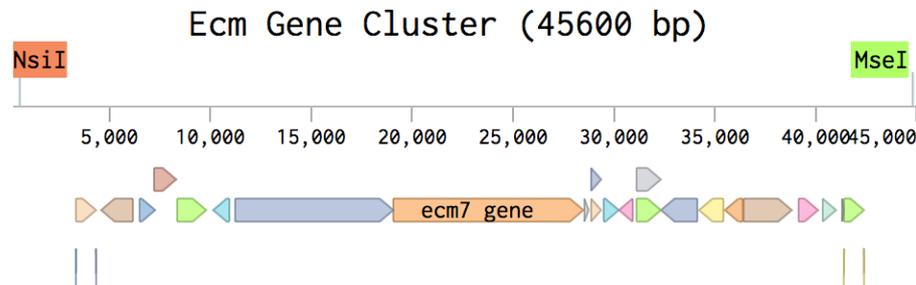


Figure 66 - The echinomycin gene cluster (*Ecm*), flanked by the restriction sites of NsiI and MseI , which can be used to digest *S. lasaliensis* genomic DNA.

The digested genomic DNA and the linearized capture vectors were transformed into *Saccharomyces cerevisiae* VL6-48N spheroplasts. The transformed spheroplasts were plated onto selective agar containing 5-FOA. Transformation with the pCAP03 vector containing the gaburedin cluster was used as a positive control, while transformation with the pCAP03 capture vector without any genomic DNA was used as a negative control.¹⁹⁵

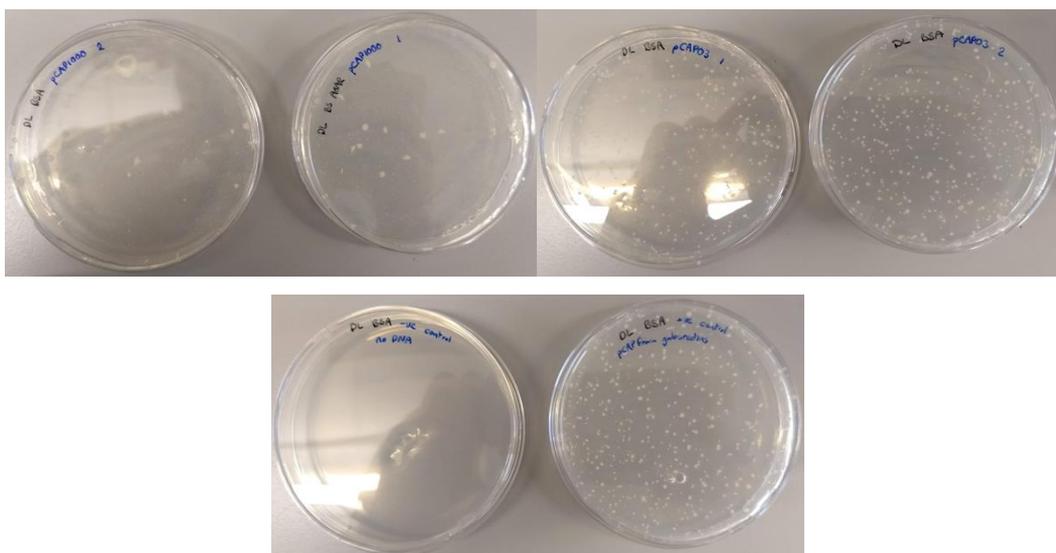


Figure 67 - Yeast transformants from the pCAP03 and pCAP1000 capture vectors on agar containing 5-FOA to select against cells containing empty capture vector. The pCAP1000 capture vector (top left) gives far fewer colonies than the pCAP03 capture vector (top right). The negative control which had no *S. lasaliensis* genomic DNA gives no colonies, while the positive control transformed with the gaburedin gene cluster in pCAP03 gives a comparable number of colonies to the pCAP03 capture vector for *Ecm*.

As expected, the positive control gave many colonies as the gaburedin cluster disrupts the expression of URA3, while the negative control gave no colonies, as homologous recombination could not occur and URA3 is expressed. Both pCAP03 and pCAP1000 capture vectors gave colonies that grew on selective agar, pictured in Figure 67. The pCAP03 capture vector gave more colonies, potentially due to a greater degree of recombination from areas of DNA which shared some homology with the 60 bp region, but not more extensive homology with the 1000 bp arms of the pCAP1000 capture vector. Sixty three colonies were re-plated onto selective agar containing 5-FOA to confirm the disruption of the URA3 selection system and prevent any residual DNA on the plates from transformation interfering with the screening process. 42 were chosen from pCAP03 plates and 21 from pCAP1000 plates. Yeast colonies were grouped into batches of 6, and the DNA extracted from these groups before screening by PCR for the presence of *ecm1* from the echinomycin BGC. This gene was amplified by each group of 6 colonies, as shown in Figure 68, indicating all groups contained at least one colony with the captured gene cluster.

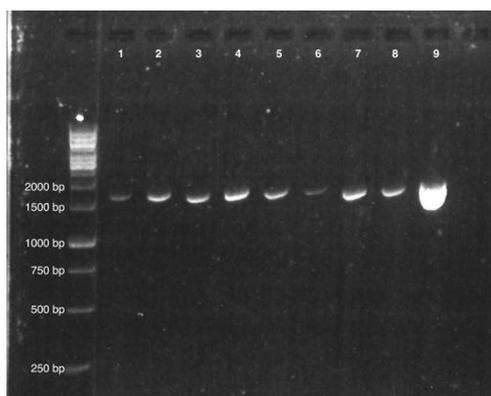


Figure 68 - Agarose gel of PCR products when screening for *ecm1* in grouped batches of extracted yeast transformant DNA (lanes 1-8) and a positive control from *S. lasaliensis* genomic DNA (lane 9).

However, when the PCR screen was repeated on the DNA isolated from individual colonies, no cluster positive colonies could be identified. It was reasoned that this may be due to a smaller amount of DNA from one colony as opposed to six, but repeated PCR attempts with increased levels of template DNA still gave no amplification.

Given the unreproducible nature of the *ecm1* PCR screen, it was decided to transform the batched genomic DNA into *E. coli* Stbl4, from which a colony PCR would be easier and greater amounts of the plasmid could be extracted. DNA from pCAP03 colonies 1-6, and pCAP1000 colonies 1-6 and 7-12 was used to transform electrocompetent *E. coli* Stbl4. This gave just 4 colonies from the DNA of pCAP1000 1-6 yeast colonies. Screening these for the sequence of the condensation domain of *ecm6* (*ecm6C1*) showed that 3 contained this sequence. To validate the capture of the cluster by colonies which amplified *ecm6C1*, the DNA was also screened for the presence of the homology arms, and for the first adenylation domain of *ecm7* (*ecm7A1*). The locations of these genes in the BGC is shown in Figure 69. However, none of the colonies amplified any of these 3 DNA fragments, suggesting that the amplification of *ecm6c1* was a false positive.

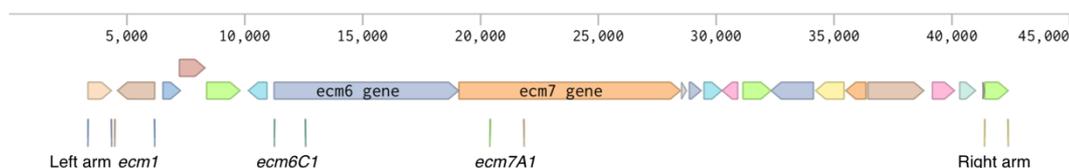


Figure 69 - DNA fragments that were screened for in yeast transformants and their location within the *ecm* gene cluster.

This was potentially due to a partial capture of the gene cluster, so the PCR product from the amplification of *ecm6C1* was sequenced. This indicated that the PCR product was indeed from the amplification of *ecm6C1*. However, neither homology arm is within 7 kb of *ecm6C1*, it is unclear how only this portion of DNA has been captured by the vector, without the DNA either side being retained. It is possible that the capture vector has undergone rearrangement in *E. coli* despite the Stb14 strain being designed to clone unstable DNA. It suggests the whole cluster must have been captured at some point to introduce *ecm6C1* into the pCAP1000 capture vector. In light of this, a number of yeast transformants were re-plated and DNA extracted. The DNA, which had previously given a negative result for the presence of *ecm1*, was then screened for the presence of both homology arms, as well as *ecm6C1* and *ecm7A1*, the locations in gene cluster are shown in Figure 30. However, none of the DNA screened gave a positive result. This would suggest that any rearrangement is either taking place in the yeast, or there is a deficiency in the PCR conditions employed to screen the yeast DNA, despite the successful amplification of these DNA fragments from *S. lasaliensis* genomic DNA under identical conditions.

2.6 *In vitro* reconstitution of Ecm6 module 1

2.6.1 Protein constructs and expression

In parallel with efforts to reconstitute and edit the echinomycin BGC *in vivo*, the components of module 1 of Ecm6 were cloned and part of the activity reconstituted *in vitro*. Chemical probes based on a range of amino acids can offload the Q2CA containing intermediate catalysed by the C domain of module 1 *in vivo*. Reproducing

this *in vitro* would demonstrate the ability to use the C domain as a biocatalyst and open the door to engineering echinomycin biosynthesis.

The first module uses the standalone A domain, Ecm1, and its ACP partner, FabC, to load Q2CA. The first module then uses C-A-T-E domains to add a D-serine residue. To investigate peptide bond formation by these enzymes, these enzymes were cloned and expressed using a pET151-TOPO system.

TOPO cloning relies on the action of DNA topoisomerase I. Topoisomerase I from the *Vaccinia* virus binds DNA, and specifically cleaves the DNA backbone on a single strand after 5'—CCCTT sequence. A bond is formed between the 3'-phosphate of the broken strand and Tyr274 of the topoisomerase. An overhang on the vector can insert between the blunt ends of a PCR product containing complementary bases to the overhang.¹⁹⁶ The topoisomerase reaction can then be reversed, releasing the topoisomerase, and ligating the PCR product insert into vector, as shown in Figure 70.¹⁹⁷

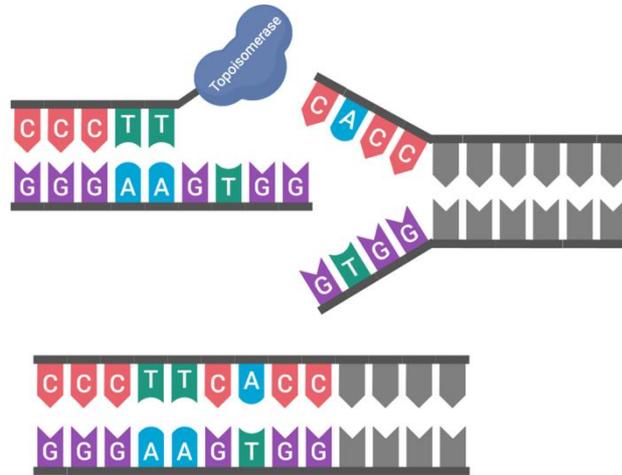


Figure 70 - Vector contains DNA Topoisomerase I after a 5'-CCCTT sequence, with a GTGG-5' overhang which can insert between a PCR product containing a complementary sequence. The 5'-CACC can then react to release the topoisomerase and form a new backbone bond.

The *fabC* gene was cloned from *S. lasaliensis* genomic DNA, using primers which contained an 5'-CACC sequence to enable insertion into the pET151-TOPO vector. The PCR product was purified and ligated into the vector to produce pET151-FabC. This gives a construct with *fabC* under the control of a T7 promoter, with an N-terminal His tag sequence encoded, shown in Figure 71. This construct was initially transformed into *E. coli* TOP10, and the plasmid extracted and sequence verified by sequencing. The correct construct was then transformed into *E. coli* BAP1. This *E. coli* strain contains a copy of the *sfp* gene, encoding the promiscuous phosphopantetheinyl transferase from *Bacillus subtilis*.¹⁹⁸ This allows the ACP produced from *fabC* to be post-translationally modified with a 4'-PPant arm *in vivo* to produce an active *holo*-ACP.

FabC could be expressed after induction of the T7 promoter with IPTG, to yield a soluble protein of around 12.5 kDa, pictured in Figure 72.

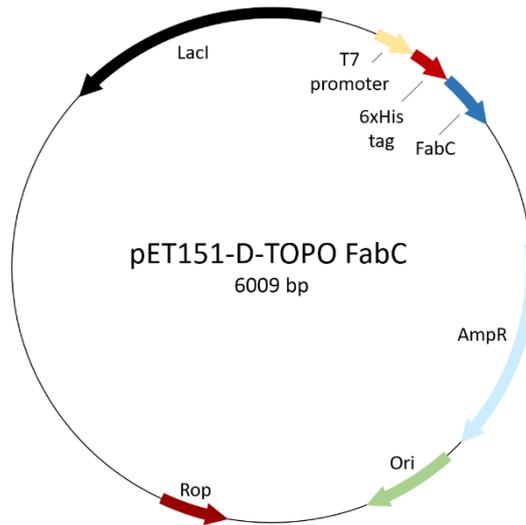


Figure 71 - Plasmid map of pET151-TOPO-FabC.

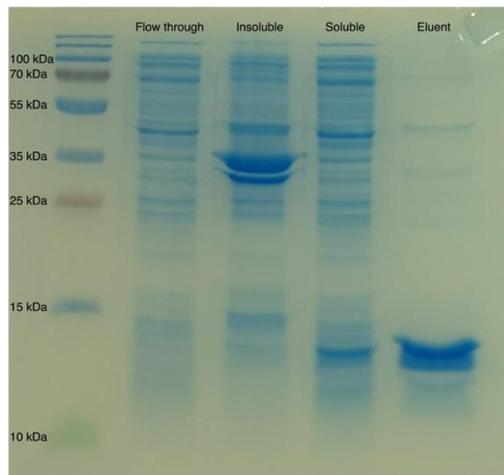


Figure 72 - SDS-PAGE gel from the expression of FabC (at approx. 12.5 kDa in eluent). The flow through is the fraction which does not bind to a nickel affinity column, the insoluble the protein that is pelleted after cell lysis, while the soluble fraction is protein in the supernatant after cell lysis. The eluent is the fraction which binds to the Ni-NTA column and can be eluted using imidazole to displace the protein.

An identical approach was used for the standalone adenylation domain, *Ecm1*. *Ecm1* was cloned from *S. lasaliensis* genomic DNA using primers containing a 5'-CACC sequence, ligated into pET151-TOPO, and transformed into *E. coli* TOP10. The sequence of the plasmid was confirmed, and the correct plasmid, shown in Figure 73, transformed into *E. coli* BL21 for protein expression.

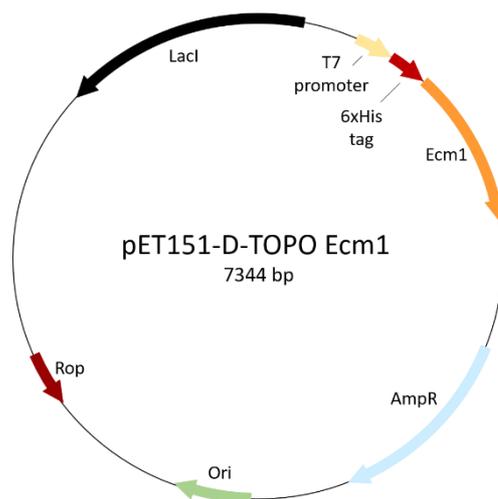


Figure 73 - Plasmid map of pET151-TOPO-Ecm1.

Ecm1 could also be expressed to give a soluble His-tagged protein of 60.8 kDa, which is shown in Figure 74. Both proteins were purified by nickel affinity chromatography and desalted using a PD-10 column, to afford FabC in a yield of 13 mg L⁻¹, and Ecm1 in a yield of 23 mg L⁻¹.

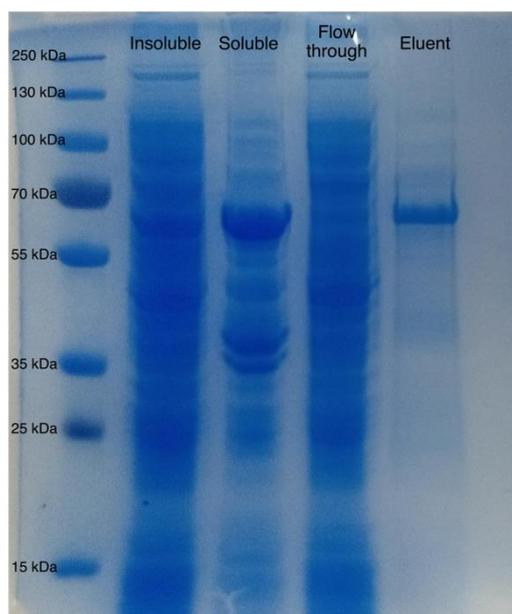


Figure 74 - SDS-PAGE gel of the fractions from the Ni-NTA purification of the adenylation domain Ecm1.

A construct for Ecm6 module 1 was initially designed using the domain boundary predictions of PKS/NRPS Analysis to include C-A-T-E domains.¹⁹⁹ This

fragment of *ecm6* was cloned by PCR, again using 5'-CACC sequence for insertion into pET151-TOPO. The construct was confirmed by sequencing and transformed into *E. coli* BAP1. Expression using conditions identical to FabC and Ecm1 yielded no protein of the correct size which bound to the Ni²⁺ affinity column, the fractions which are pictured in Figure 75.

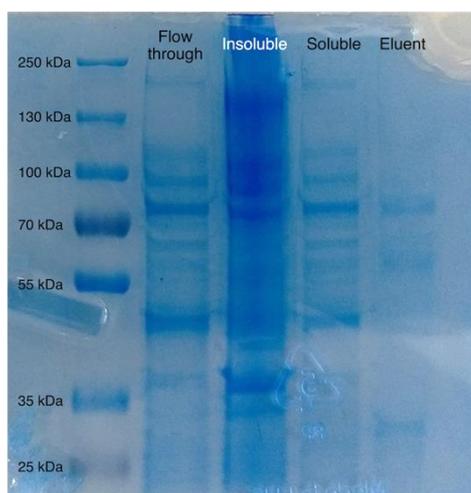


Figure 75 - SDS-PAGE gel from the expression of Ecm6m1 , showing a clear absence of soluble or eluted protein at 150 kDa.

Changing expression conditions to those used by Watanabe *et al.* to express full length Ecm6 gave no significant improvement in protein solubility or expression, with only a small band present at a size which may correspond to the expected size of 147 kDa, shown in Figure 76. The smaller bands which appear to increase in the induced culture could be products of proteolytic degradation of the large NRPS module.

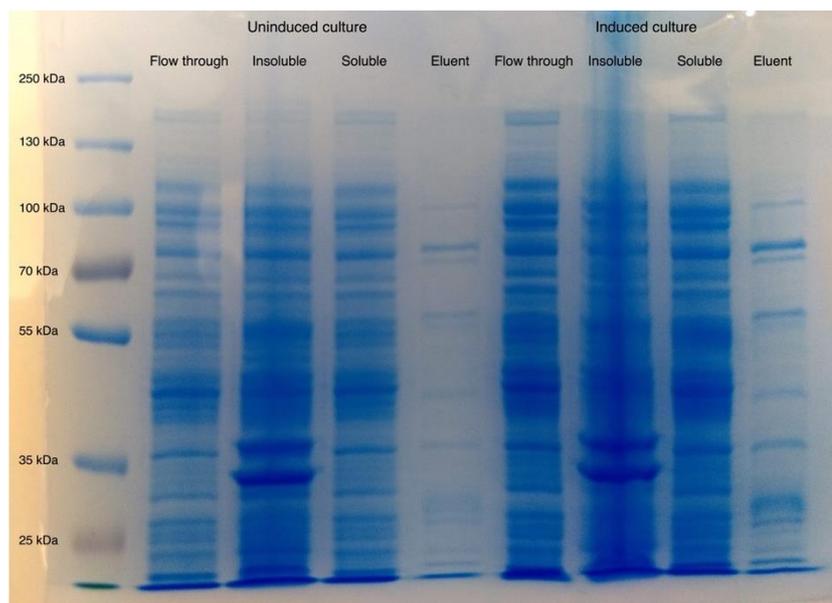


Figure 76 - Ecm6m1 induced using conditions developed by Watanabe et al. Potentially a small amount of soluble protein around 150 kDa in the induced culture.

Given the lack of protein being produced from this construct, a second construct of Ecm6 module 1 was designed. Using a homology model produced by Phyre2, flexible regions were identified at domain boundaries and identified as possible end points for the construct.²⁰⁰ This new construct was cloned using PCR with primers containing NdeI and XhoI restriction sites. The PCR product was purified, digested with NdeI and XhoI, and purified to remove the excised region and restriction enzymes. A pET28a vector was also digested by NdeI and XhoI to produce a linearised vector. Ligation of these two DNA fragments using T4 DNA ligase was attempted a number of times, without success.

Investigating condensation domain activity typically uses analogues of the acceptor amino acids, such as SNACs, and produces a product no longer tethered to the NRPS, the A, T, and E domains play no catalytic role. Given this, a shorter construct containing only the condensation domain of Ecm6 was designed, as shown in Figure 77.

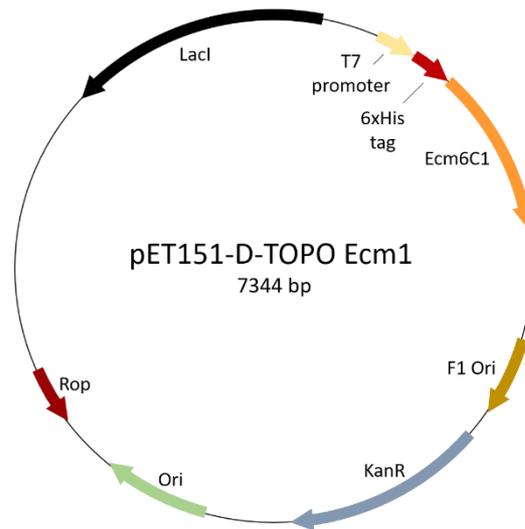


Figure 77 - Plasmid map of pET28a(+)-Ecm6C1.

The Phyre2 homology model was again used to determine the domain boundaries, identifying flexible linker regions. The C domain was cloned by PCR using primers containing BamHI and XhoI restriction sites, and a stop codon to terminate protein translation. The PCR product was purified, digested by BamHI and XhoI, and purified. This was ligated into a similarly digested pET28a vector using T4 ligase and transformed into *E. coli* TOP10. The sequence of the plasmid insert was confirmed and transformed into *E. coli* BL21. The Ecm6C1 was expressed after induction with IPTG, to give a soluble protein of 50.6 kDa, shown in Figure 78, in a yield of 17 mg L⁻¹.

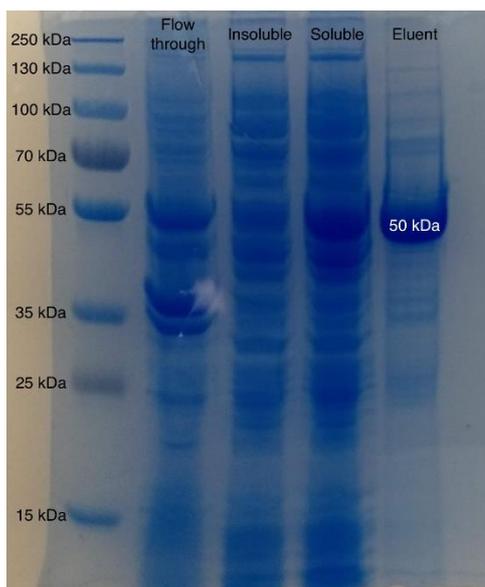


Figure 78 - SDS-PAGE gel of the fractions from Nickel affinity purification of Ecm6C1, visible at 50 kDa in the eluted fraction.

2.6.2 *In vitro* activity assays

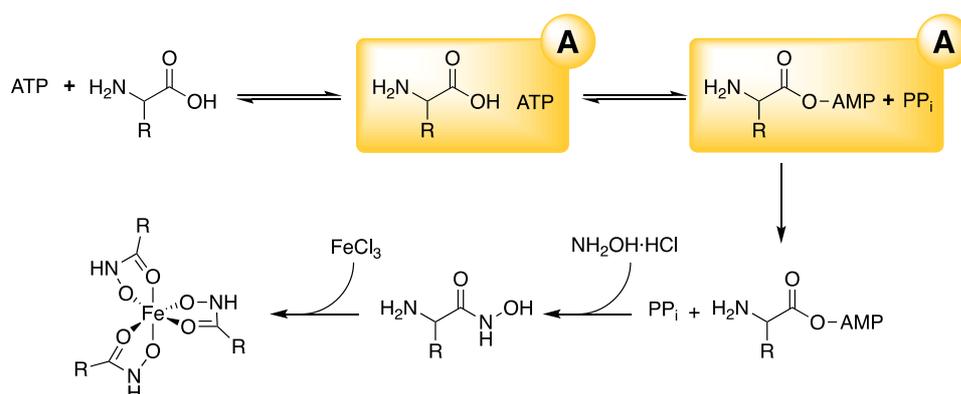


Figure 79 - Formation of a hydroxamic acid from the amino adenylate product of the A domain Ecm1, which can form a chromogenic complex with Fe(III).

To test the activity of the adenylation domain *in vitro*, a colourimetric assay which depends on the formation of the acyl adenylate was chosen.¹⁰⁹ The activation of the carboxylic acid forms an acyl adenylate, which then reacts with hydroxylamine to form the hydroxamic acid. This can then coordinate to Fe(III) to produce a coloured species which can be detected, as shown in Figure 79.

This assay showed a clear colour change between active Ecm1 and the denatured protein, as shown in Figure 80, but the change of colour could not be reliably

distinguished from the colour of FeCl_3 in a UV-Vis spectra at 540 nm or the literature wavelength of 420 nm. A lack of sensitivity is a noted limitation of this assay, so a different assay system was sought.

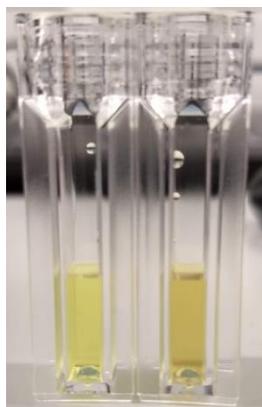


Figure 80 - Colour difference observed in the hydroxylamine assay. Clear change in colour in the sample containing active Ecm1 (right), compared to the sample containing denatured Ecm1.

The initial half-reaction of the adenylation domain produces inorganic pyrophosphate (PP_i) as well as the acyl adenylate. By using an inorganic pyrophosphatase to produce inorganic phosphate (P_i) a Malachite green based assay can be used to monitor A domains.¹⁰⁷ The amount of P_i produced can be quantified by comparison to a standard, and hence the amount of PP_i and the activity of the A domain.

Performing this assay on Ecm1, using yeast inorganic pyrophosphatase (YIPP) and a commercially available Malachite green phosphate detection kit from R & D Systems gave no difference between active and denatured enzyme reactions. Given the activity observed but not quantified in the hydroxylamine-based assay, it appeared that the assay conditions were the cause of the problem. Assay components were tested individually to identify the cause of any interference with the assay reagents. It was found that ATP alone produced comparable absorbance readings at 620 nm (A_{620}) to the complete reaction mixture. To identify if hydrolysis of the ATP prior to detection was the issue, fresh ATP was tested, but produced an identical result. Reducing the concentration from 5 mM to 1 mM did reduce the A_{620} , but the absorbance continued to rise as the colour was allowed to develop. Given the reagents contain ammonium molybdate in 3 M sulfuric acid, it is likely that the reagent is significantly hydrolysing

ATP during its 10-minute incubation step. However, adding a solution of 34% trisodium citrate as a buffer solution completely inhibited the development of any colour in the assay, presumably by preventing formation of the phosphomolybdate complex. Given the proprietary nature of the phosphate detection kit reagents, avoiding this problem proved intractable. An alternate detection reagent was therefore prepared according to Baykov *et al.*, combining the acidic molybdate solution with Malachite green directly, before mixing with the enzymatic reaction.²⁰¹ While the use of sodium citrate in this method did reduce the absolute value of the A_{620} , it also suppressed background hydrolysis of ATP to allow the function of the Ecm1 A domain to be assessed.

Ecm1 was shown to be active under the assay conditions, as shown in Figure 81. The A_{620} were normalised by subtracting a blank from the samples, which were then compared. As expected, the denatured Ecm1 showed no increased absorbance, consistent with no production of PP_i . In contrast, the active enzyme showed an increased absorbance, showing the enzyme is active and generating the acyl adenylate and PP_i .

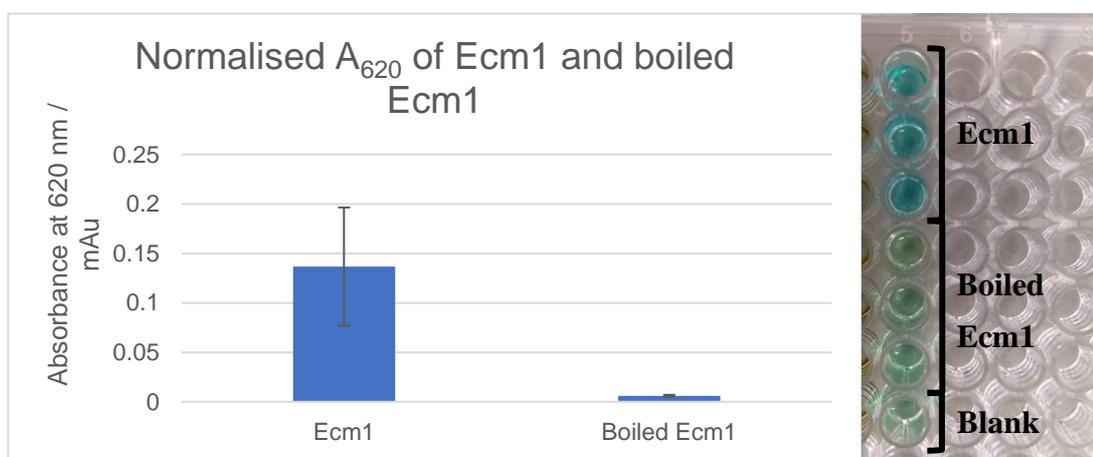


Figure 81 - Graph showing the absorbance of phosphomolybdate-malachite green complex in samples containing Ecm1 compared to denatured Ecm1. Ecm1 2.5 μ M, 25 mM Tris pH 8.0, 15 mM $MgCl_2$, 2.25 mM ATP, 150 mM $NH_2OH \cdot HCl$, 3 mM Q2CA, 37 $^{\circ}C$, 1 hour. The assay is shown on the right.

As the A domain has been shown to be active, the condensation domain activity with chemical probes could be tested *in vitro*. Combining Ecm1 with FabC and quinoxaline-2-carboxylic acid (Q2CA) should result in the production of Q2CA-S-FabC thioester, the natural donor substrate for Ecm6C1. Chemical probes can then

be used as the acceptor substrates, as illustrated in Figure 82, to investigate their ability to form amide bonds *in vitro* and compare to the activity observed *in vivo*.

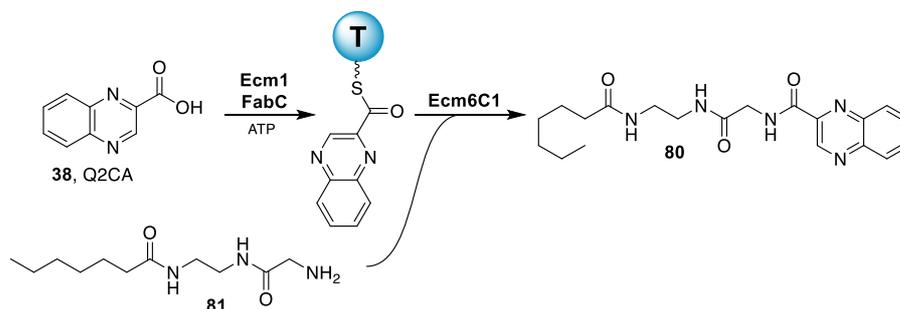
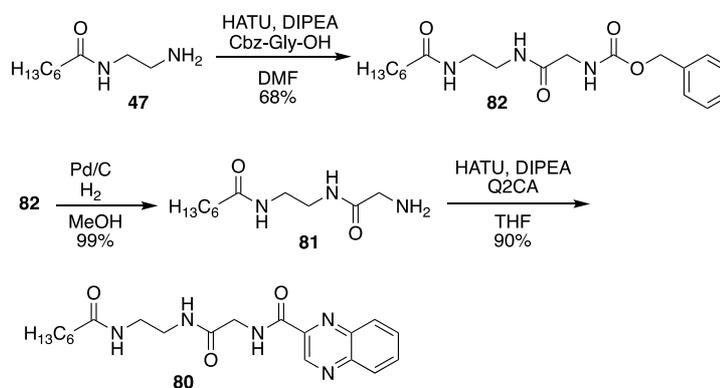


Figure 82 - Schematic of the C domain assay using FabC, Ecm1, and Ecm6C1 with chemical probes to test substrate specificity of Ecm6C1.



Scheme 4 - Synthesis of glycine-based chemical probe, 81, and synthetic standard for C domains assays, 80.

The C domain Ecm6C1 was incubated with Ecm1, FabC, Q2CA, and a chemical probe based on glycine, **81**, synthesised as shown in Scheme 4. When incubated at 37 °C for either one hour or three hours, no product of peptide bond formation can be detected in the EtOAc extracts of these assays by UV detection at 254 nm on an analytical HPLC at a retention time matching the authentic standard, **80**. This is in contrast to the amide bond formation observed *in vivo* which results in intercepted intermediates as previously described. It is possible that the lack of peptide bond formation is due to the excised condensation domain not being functional outside of a multidomain NRPS. It is also possible that the probes are poorer substrates for the acceptor site than an amino acyl thioester, and any amide bond formation is reduced to an undetectable level. Further work is needed to establish the cause of the lack of C

domain activity observe, and how this relates to the observed *in vivo* data, which will be discussed in chapter 6.

3.1 Chemical probing of colibactin assembly

3.1.1 Discovery of colibactin

Colibactin was first identified as a natural product of interest in 2006. It was observed that certain commensal *E. coli* strains contained a genomic island which could induce megalocytosis, growth of cells without mitosis, in eukaryotic cells. This behaviour was observed to be dependent on cell-cell contact, and separation of cells by a 0.2 μM membrane was sufficient to prevent it. Culture supernatant and cell lysate was also unable to cause megalocytosis. Transposon mutants in a 54 kb genomic island lost their pathogenic effects, indicating this region was responsible. Indeed, introducing the island into *E. coli* DH10B on a bacterial artificial chromosome (BAC) conferred the ability to induce megalocytosis.²⁰²

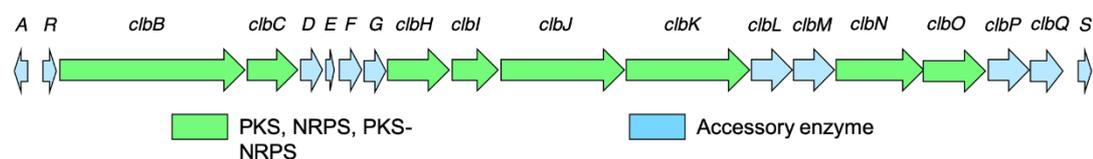


Figure 83 - Biosynthetic gene cluster coding for colibactin, the *pks* island.

Gene	Function	Gene	Function	Gene	Function
<i>clbA</i>	PPTase	<i>clbG</i>	Acyltransferase (AT)	<i>clbM</i>	Inner membrane transporter
<i>clbB</i>	PKS-NRPS	<i>clbH</i>	NRPS	<i>clbN</i>	NRPS
<i>clbC</i>	PKS	<i>clbI</i>	PKS	<i>clbO</i>	PKS
<i>clbD</i>	3-hydroxybutyryl-CoA dehydrogenase	<i>clbJ</i>	NRPS	<i>clbP</i>	Inner membrane peptidase
<i>clbE</i>	Acyl carrier protein (ACP)	<i>clbK</i>	PKS-NRPS	<i>clbQ</i>	Type II thioesterase
<i>clbF</i>	Acyl-CoA dehydrogenase	<i>clbL</i>	Amidase	<i>clbR</i>	Lux-R transcriptional regulator
<i>clbS</i>	Self-resistance protein				

Table 3 - The functions of the genes encoded by the colibactin BGC.

The genomic island, referred to as the *pks* island, contains genes for encoding for 3 NRPS, 3 PKS, 2 NRPS/PKS, and 11 accessory enzymes, shown in Figure 83, with assigned functions shown in Table 3. Mutagenesis showed that all these genes, with the exception of an efflux pump, were required for megalocytosis. It was shown that the product of this gene cluster, dubbed colibactin, caused double-strand breaks (DSBs) in the DNA of eukaryotic cells, blocking the G₂/M transition. Colonisation with *pks*⁺ *E. coli* strains, and the subsequent DNA damage, was postulated to be linked to the development of colorectal cancer.²⁰³²⁰³ However, the structure of the product was unknown, and the inability of culture supernatant or lysate to cause megalocytosis had shown that the compound responsible was not stable. This prevented the use of standard methods to isolate and characterise natural products as colibactin would not withstand extraction and extensive purification. Prediction of the encoded product was also hampered by the unusual enzymology of the assembly line, including a number of novel transformations in a non-colinear fashion. The known PKS-NRPS assembly line up to the beginning of the chemical probing of colibactin assembly described in this work is shown in Figure 84.

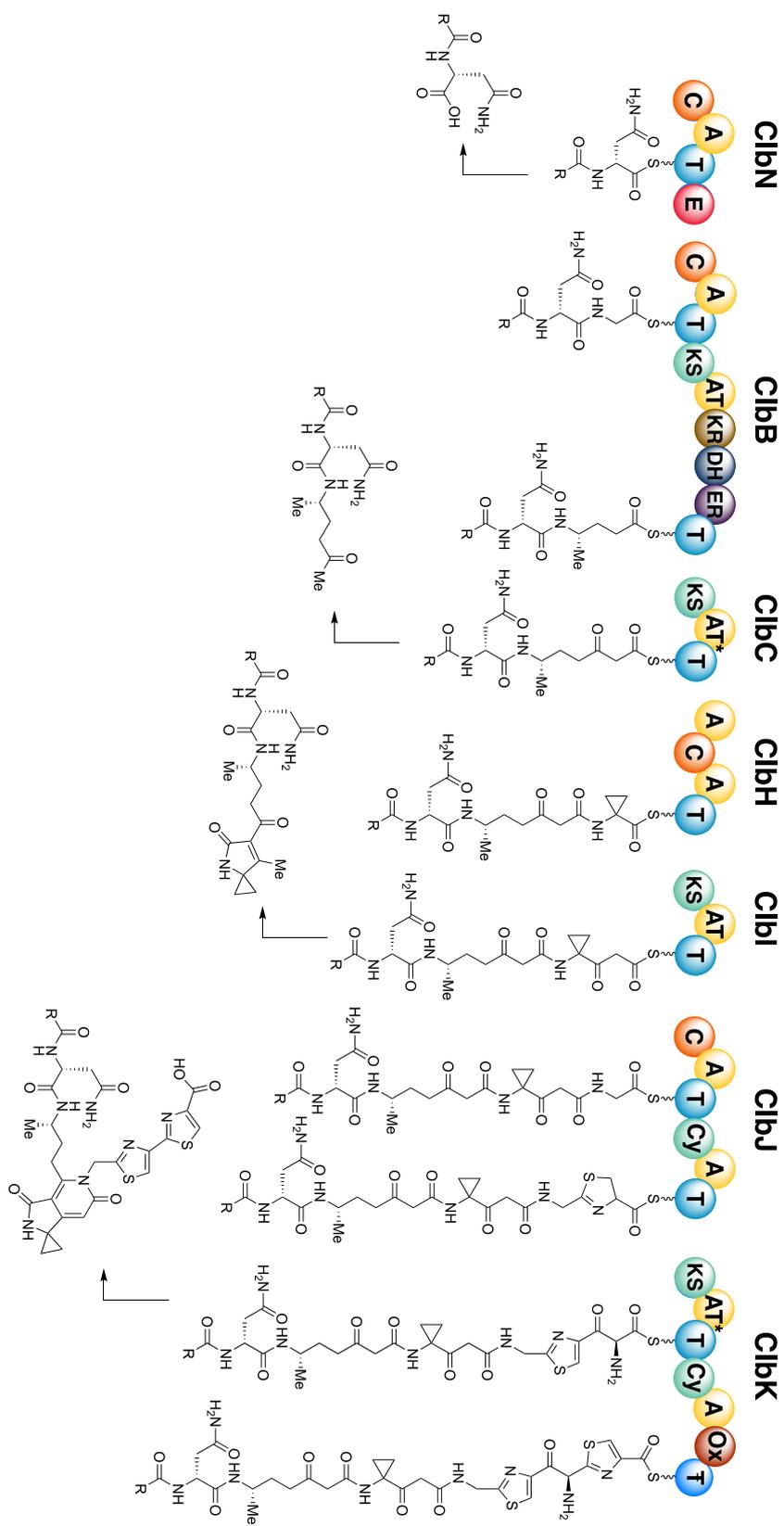


Figure 84 - The elucidated steps of colibactin biosynthesis, with selected identified precolibactins, up to

2017.

3.1.2 Unravelling colibactin biosynthesis

The first development in the quest to determine the structure of the genotoxin colibactin was the solving of the structure of ClbP, a putative peptidase, spurred by work on similar enzymes, ZmaA and XcnG, from the biosynthesis of zwittermicin and the xenocoumarins.²⁰⁴ ClbP shares around 30% identity with ZmaA and XcnG, and they are able to complement the *pks* island lacking *clbP*. It was determined that ClbP is an inner-membrane protein, and the active site identified by analogy with the known enzyme AmpC, an *E. coli* β -lactamase.²⁰⁵ This allows the maturation of the final colibactin product outside the cytoplasm and far from the host DNA. ClbN was identified as an NRPS containing a starter C domain, which may be able to acylate an amino acid to form a lipopeptide, as in xenocoumarin biosynthesis.

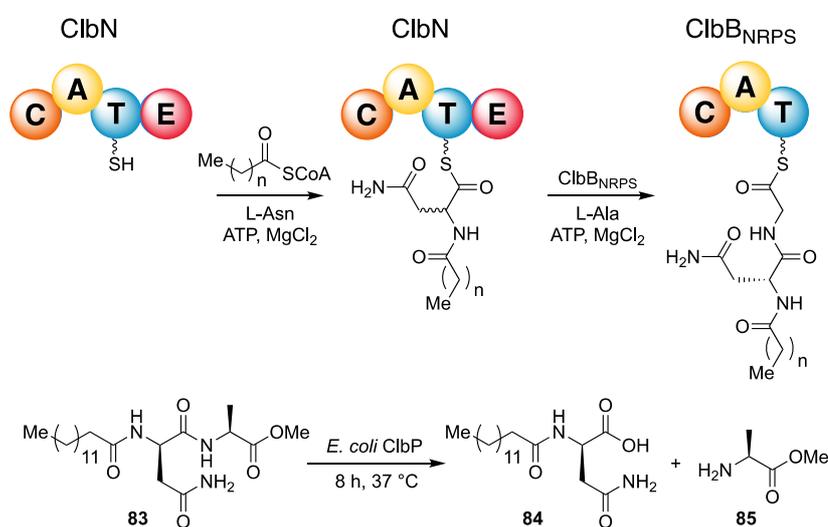


Figure 85 - *In vitro* activity of ClbN and ClbB shows production of a prodrug motif which can be cleaved by *E. coli* expressing peptidase ClbP.

This was further examined *in vitro*, as shown in Figure 85.²⁰⁶ It was shown that ClbN, with a C-A-T-E domain architecture, was able to activate L-asparagine, which is loaded onto the T domain. The C domain can catalyse the amide bond formation between L-Asn and fatty acyl-CoA thioesters, preferentially C₁₄. This is necessary as the cluster does not contain any fatty acid activating enzyme. Deuterium exchange proved the activity of the E domain, to give an *N*-Acyl-D-Asn prodrug motif as in xenocoumarin biosynthesis, which may therefore provide a substrate to ClbP.

It was reasoned the next module must be an NRPS, to form an amide bond which could be cleaved by ClbP. The only ^DCL domain in the cluster was present in ClbB, a two module PKS-NRPS. The A domain of ClbB preferentially activates L-Ala, and crucially accepts substrates from ClbN. This model substrate, **83**, can be cleaved by ClbP to give the cleaved prodrug motif, **84**, shown in Figure 85. This *in vitro* work was supported by simultaneous *in vivo* work in *E. coli* which resulted in the isolation and characterisation of the cleaved prodrug motif *N*-myrostoyl-D-Asn, **84**.²⁰⁷

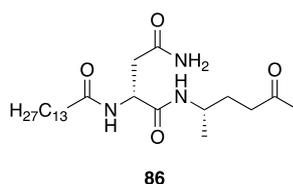


Figure 86 – Intermediate 86 identified from comparative metabolomics, produced by ClbN, ClbB and a second PKS.

Metabolomics of *clb+* strains compared to Δclb and $\Delta clbP$ strains resulted in the identification of a new intermediate species **86** shown in Figure 86, which accumulated in the $\Delta clbP$ strain.²⁰⁸ This was predicted to be the product of extension by a full length ClbB, and a second PKS module, one of ClbC, I, or O, followed by hydrolysis and decarboxylation. The intermediate, **86**, could be cleaved by ClbP, resulted in the formation of a cyclic imine from spontaneous cyclisation. Additionally, based on the fragmentation patterns and molecular networking, further intermediate structures could be proposed.

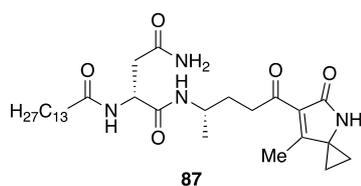


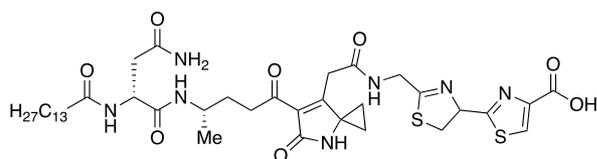
Figure 87 - Spiro-cyclopropane containing intermediate 87 isolated from $\Delta clbP$ mutant, and dependent on *clbH*, *C*, and *I*.

Further examination of species accumulated in $\Delta clbP$ strains resulted in the discovery of an intermediate containing a spiro-cyclopropane moiety, **87**, identified as a likely alkylating agent, shown in Figure 87.²⁰⁹ The deletion of ClbP was designed to

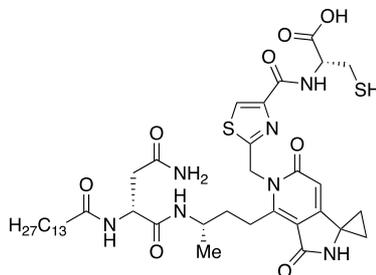
amass intermediates with the prodrug motif, precolibactins, which may have greater stability than the final compound, and the prodrug motif can be used to identify metabolites. The structure was confirmed by chemical synthesis and was shown to be processed by ClbP *in vivo*. This would result in an α,β -unsaturated iminium next to the cyclopropane, giving a potent electrophile.

In vivo work published in parallel also identified the cyclopropane species **87**, and showed that deletions in *clbH*, *C*, or *I*, prevented formation.²¹⁰ Feeding experiments with ¹³C labelled L-methionine showed this is used to form the 1-aminocyclopropanecarboxylic acid (ACC). However, the cluster does not possess any enzymes with homology to those used in known routes to ACC, suggesting novel enzyme catalysis at work. ClbH was suggested as an ACC synthase.

Vizcaino *et al.* also utilized isotopic labelling, combined with the molecular networking approach, to identify a novel precolibactin, shown in Figure 88. Using the predicted specificity of the adenylation domains of the NRPS combined with [U-¹³C] amino acids, the incorporation of Asn, Ala, Met, Gly, Cys, and Ser were shown. This resulted in the postulation of the most complex colibactin biosynthetic pathway then reported. Additionally, protein homology to the zwittermicin biosynthetic enzymes was used to predict the interaction of the ClbH A₁ domain with standalone ACP, ClbE. This was then proposed to interact with ClbDF to generate an aminomalonate extender unit as in zwittermicin biosynthesis, but evidence could not be found for this. The then largest precolibactin, precolibactin A, **88**, was identified, although the structure was later revised to that of **89**. The thiazole-thiazoline was proposed to take part in DNA intercalation, as in bleomycin, while the cyclopropane was again proposed as an electrophile, similar to those already known.²¹¹ Weak inter-strand cross-links were formed using the identified intermediate **87**, but there was no evidence of the DSBs characteristic of colibactin.



Proposed precolibactin A, **88**



Precolibactin A, **89**

Figure 88 - Proposed structure of precolibactin A 88, identified by metabolomics, and the correct structure 89 identified by synthesis.

Given the low yield and instability of colibactin intermediates, TAR was used to heterologously express the cluster, and increase the abundance of metabolites. In a $\Delta clbP$ mutant, [2,2-D₂]glycine was used to verify the involvement of the NRPS ClbJ in the biosynthesis of 2 intermediates, which were isolated and characterised. These intermediates diverge from those previously identified, containing a pyridine-2-one moiety, which is also seen in the corrected structure of precolibactin A, **89**.²¹²

The postulated action of the enzymes of ClbDEF were confirmed before the end of 2015. The PKS enzymes ClbO and ClbK are involved in the incorporation of an unknown polyketide building block. Bioinformatics show ClbD and ClbF have homology to known dehydrogenases ZmaG and ZmaI from the zwittermicin pathway, which form aminomalonyl-ACP (AM-ACP) from seryl-ACP. Freestanding ACP ClbE is homologous to the freestanding ACP in zwittermicin biosynthesis. The PPTase ClbA was shown to be promiscuous, capable of modifying ACP and PCP domains, and even replace PPTases in unrelated clusters.²¹³ It was shown that combining these enzymes with ClbH, which activates L-Ser at A₁, gives a mass corresponding to glycyl-ACP.⁹⁵ The reaction scheme to AM-ACP is shown in Figure 89. The glycyl-ACP would be expected from the spontaneous decarboxylation of AM-ACP.²¹⁴ It was speculated that AM-ACP could be involved in ACC biosynthesis, but this was ruled out as deletion of AM biosynthetic genes did not prevent production of ACC-

containing products. ClbK and O contain AT domains which are degraded, lacking serine in the conserved GSHxG motif required to transfer extender units to the ACP. However, the standalone AT ClbG has a functional GSHxG motif, but an altered specificity motif. Rather than a HAFH motif, corresponding to malonyl specificity, it possesses a YPYF motif. This is similar to the GPFH motif of ZmaF, a standalone AT that loads aminomalonnate in the biosynthesis of zwittermicin.

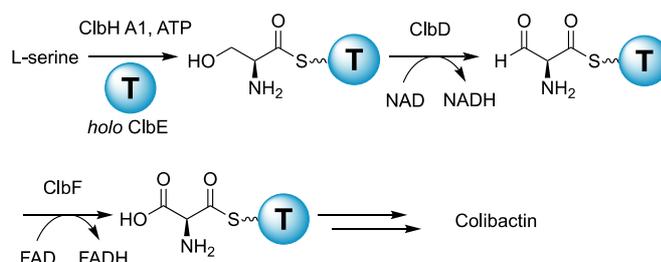


Figure 89 - Formation of aminomalonyl-ACP from seryl-ACP by dehydrogenases ClbD and ClbF.

This hypothesis was rapidly tested to determine where aminomalonnate could be incorporated, given it is essential for genotoxicity, but has not be observed in any precolibactins. It was shown that ClbG can transfer AM to PKSs lacking functional AT domains such as ClbC, K, and O, and even to the *cis*-AT domain containing ClbI.²¹⁵ Deleting *clbG* in a $\Delta clbP$ mutant did result in an increased titre of a precolibactin, **90**, shown in Figure 90.

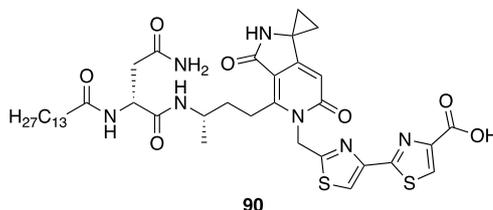
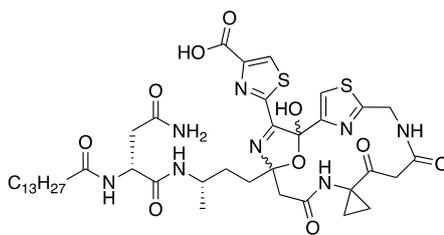


Figure 90 – Precolibactin, 90, which accumulated in a $\Delta clbG\Delta clbP$ *E. coli* strain.

ClbC and ClbI are predicted to use malonnate extender units in the formation of colibactin, but ClbG could not load ClbC with a malonyl unit, and ClbI demonstrated an *in vitro* preference for an aminomalonyl unit. It was speculated that the type II thioesterase, ClbQ, may have a proofreading role to ensure the correct extender unit is incorporated, or be dependent of the differing intracellular levels of malonyl and aminomalonyl. ClbK showed it was capable of self-acylation with aminomalonnate from ClbE, without a need for ClbG, despite a degraded AT.²¹⁵

Further metabolomics work used multiplex automated genome editing (MAGE) to site-selectively inactivate individual protein domains. MAGE allows allelic replacement on *E. coli* chromosomes or BACs using single-strand synthetic oligonucleotides. The metabolomic signatures of these PKS or NRPS domain inactivations were used to determine the effect of the domains on metabolite production. These could support the involvement of the domain in transformations and determine the order of steps in a multidomain protein. The focus was on ClbO and ClbL, as these enzymes had yet to be included in the biosynthetic model. In a ClbP point-mutant background, domains downstream of ClbH were inactivated, as the biosynthesis prior to ACC incorporation was established. Cyclopropyl-containing intermediate production was abolished by inactivation of any of the C-A₂-T domains of ClbH. In ClbJ, the timing of oxidation was unclear as it lacks an oxidase (Ox) domain, leaving the possibility of spontaneous oxidation, by the oxidase of ClbK, or by another enzyme. In the inactivated Ox domain mutant ClbK, there was an increase in thiazoline with a reduction in monothiazole products. This suggests the thiazoline is processed by ClbK Ox, which is then the substrate for downstream processes.²¹⁶

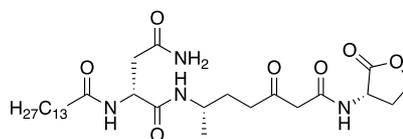
Precolibactin-886, **91**, shown in Figure 91, was the first reported metabolite to incorporate aminomalonate, through ClbK, in a $\Delta clbP\Delta clbQ$ strain.²¹⁷ This would require incorporation of AM between thiazole rings, suggesting bis-thiazole products are a result of module skipping. In mutants where ClbH-A₁ or ClbK-PKS were inactivated, preventing incorporation of aminomalonate, this module skipping behaviour was shown to be possible.



Precolibactin-886, **91**

Figure 91 - Structure of precolibactin-886, **91, the first identified product of the colibactin pathway to contain an aminomalonate moiety.**

The biosynthesis of the ACC unit was found to occur on ClbH, which uses SAM as a substrate.¹⁰⁰ Bioinformatic predictions were inconclusive for ClbH-A₂, suggesting the activation of a non-proteinogenic amino acid as a possibility. Neither L-methionine nor ACC were activated *in vitro*. L-Ser and SAM could be activated by the full length ClbH construct, while the A₂ domain could activate SAM as part of a tridomain C-A₂-T construct. However, the presence of ClbI was essential for the formation of ACC from SAM. Identification of a lactone by-product, **92**, shown in Figure 92, produced from the pathway, allowed the confirmation that the aminobutyryl group of [U-¹³C, ¹⁵N]Met is incorporated into the lactone.



Lactone metabolite, **92**

Figure 92 - Lactone metabolite, **92, produced by *E. coli* expressing clbN, B, C, H, and clbA.**

This suggests that amide bond formation occurs first, followed by cyclopropane formation, for which the Ser178 of the ClbI KS is essential, potentially acting as a general base or a nucleophile. This represents a novel biosynthetic route to ACC, shown in Figure 93.

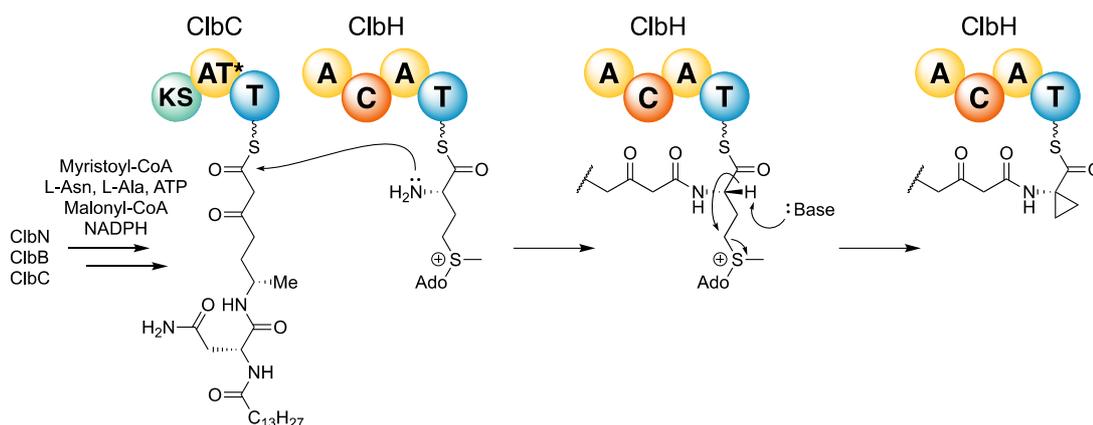


Figure 93 - Biosynthesis of 1-aminocyclopropanecarboxylic acid (ACC) from SAM, catalysed by ClbH.

It was also shown that while $\Delta clbP$ and $\Delta clbP\Delta clbQ$ mutants were used to aid the isolation and identification of precolibactins, these mutants do not exhibit cytopathogenicity, presumably due to the absence of the mature colibactin. ClbQ was shown to be a low specificity type II thioesterase, capable of hydrolytic release of both correct and incorrect intermediates from the pathway.²¹⁸ Cyclised pyridone-containing precolibactins were found not to be substrates of ClbQ, while the linear intermediates were, suggesting the pyridone-containing precolibactins are off-pathway products, not true intermediates. It was also demonstrated that ClbQ can hydrolyse AM-ACP, releasing aminomalonate. Aminomalonate-containing precolibactins, such as precolibactin-886, therefore accumulate in $\Delta clbQ$ mutants.

Chemical synthesis was also used to determine the likely structure of precolibactins, and therefore aid identification of late stage intermediates.²¹² $\Delta clbP$ strains accumulate advanced *N*-acylated precolibactins, which readily cyclise in mild basic conditions to form pyridone products. Synthetic precolibactin analogues were prepared bearing a *N*-terminal Boc group, rather than an amide. The more electron-rich carbamate immediately cyclised to form an imide, and upon deprotection underwent cyclodehydration to form an unsaturated imine. These differing cyclisation patterns are illustrated in Figure 94. This result indicates presence of ClbP alters the pathway of released intermediates, indeed feeding a linear precursor to *E. coli* expressing ClbP results in an unsaturated imine.

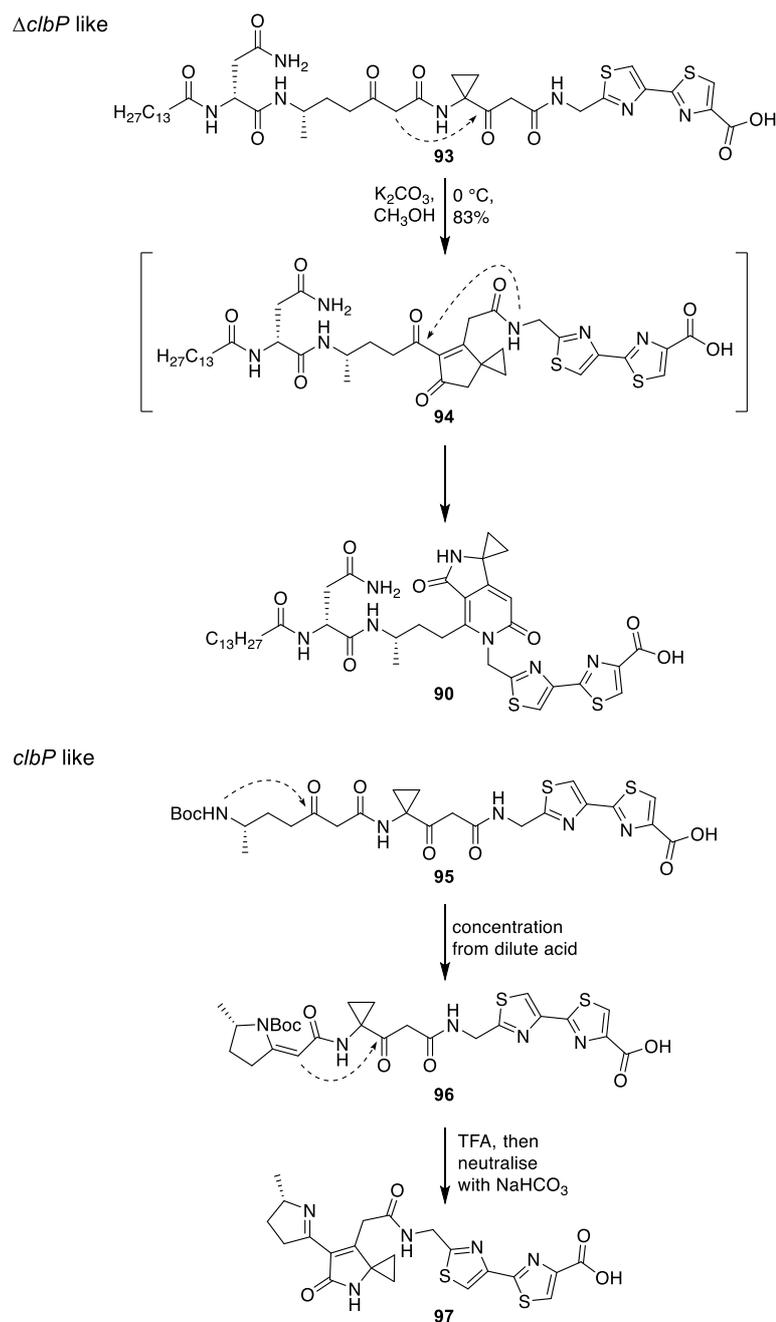


Figure 94 - Differing cyclization patterns of precolibactins containing the *N*-Acyl-d-Asn prodrug motif present in *ΔclbP* strains, and those lacking the prodrug motif in strains containing *clbP*.

The ability of a number of analogues to damage DNA was also tested. The unsaturated imine, **98**, shown in Figure 95 was capable of causing damage at nanomolar concentrations, with extensive cleavage observed by DNA electrophoresis. Pyridone compounds were inactive up to 0.5 mM.²¹²

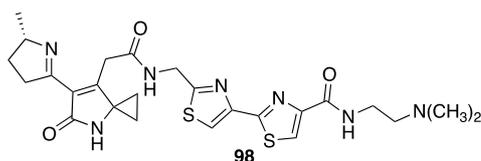


Figure 95 - Synthetic precolibactin mimic, 98, capable of damaging DNA at nM concentrations.

This damage is likely a result of alkylation by the cyclopropane, as replacing the group with a gem-dimethyl moiety resulted in no DNA-damaging activity up to 0.5 mM. Severe DNA alkylation can result in DSBs through incomplete repair, possibly explaining colibactins known activity.²¹⁹ A dimeric imine was also found to be capable of cross-linking DNA. The altered reactivity of intermediates in $\Delta clbP$ strains led to the suggestion that the most advanced structure, precolibactin-886, **91**, may be an oxidation product of a putative precolibactin-888, shown in Figure 96, which could lead to an unsaturated imine in the presence of ClbP.

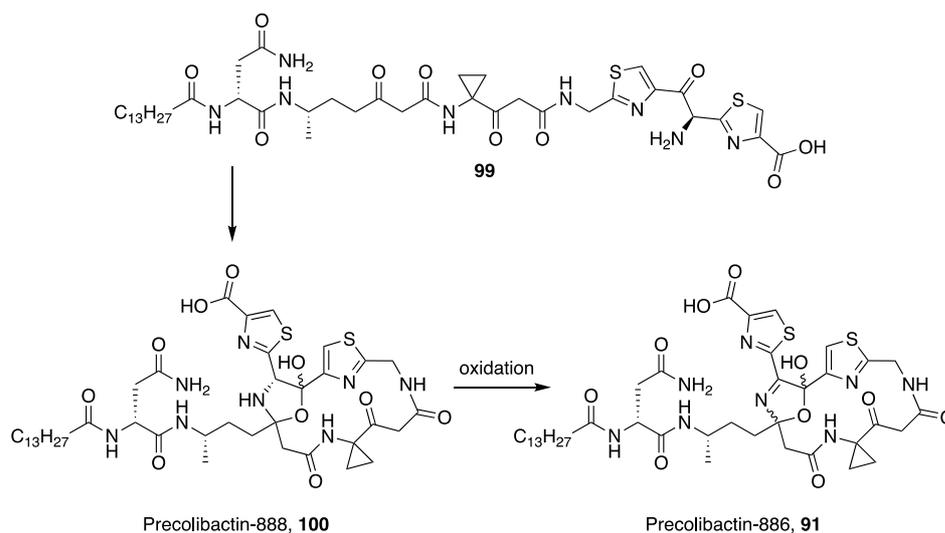
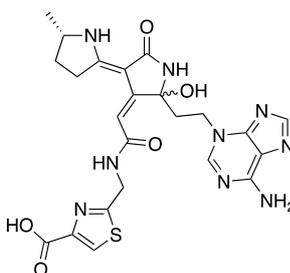


Figure 96 - Putative biosynthetic route to precolibactin-886, 91, from a linear precolibactin, 99, via precolibactin-888, 100.

Despite the evidence that synthetic colibactin warheads alkylate DNA, evidence for DNA alkylation *in vivo* remained lacking. Unknown DNA adducts are hard to identify due to their low abundance within a complex mixture. Using untargeted DNA-adductomics, in which MS³ is performed on ions which lose 2'-deoxyribose or one of the four DNA bases during MS² fragmentation, accurate mass and MSⁿ can give structural information on the adducts.²²⁰ This approach identified adenine adducts from HeLa cells transiently infected with *pks+* *E. coli*.²²¹ Labelling with ¹³C-containing

Ala, Met, Gly, and Cys confirmed these adducts were related to the colibactin pathway. Based on the mass spectrometric data, a structure was proposed, **101**, which was confirmed by the synthesis of authentic standards with identical retention times and MS² fragmentation, shown in Figure 97. These contain carboxylic acids, the standards of which proved poor alkylating agents, and do not require the use of all biosynthetic enzymes, suggesting they are degradation products of a larger molecule. These DNA adducts were also separately identified from the incubation of *pks+* *E. coli* with plasmid DNA.²²²



DNA adducts caused by colibactin, **101**

Figure 97 - DNA adduct, 101, identified from HeLa cells treated with *pks+* *E. coli*, and confirmed by *in vitro* alkylation of DNA.

3.2 Chemical probes for the investigation of colibactin biosynthesis

Previous work in the group, conducted by Dr Ina Wilkening, Candace Ho and Jedzrej Gozdzik, had resulted in the identification of a number of known colibactin intermediates from the use of chemical probes to intercept both PKS and NRPS intermediates (unpublished data). A selection of these are presented in Figure 98.

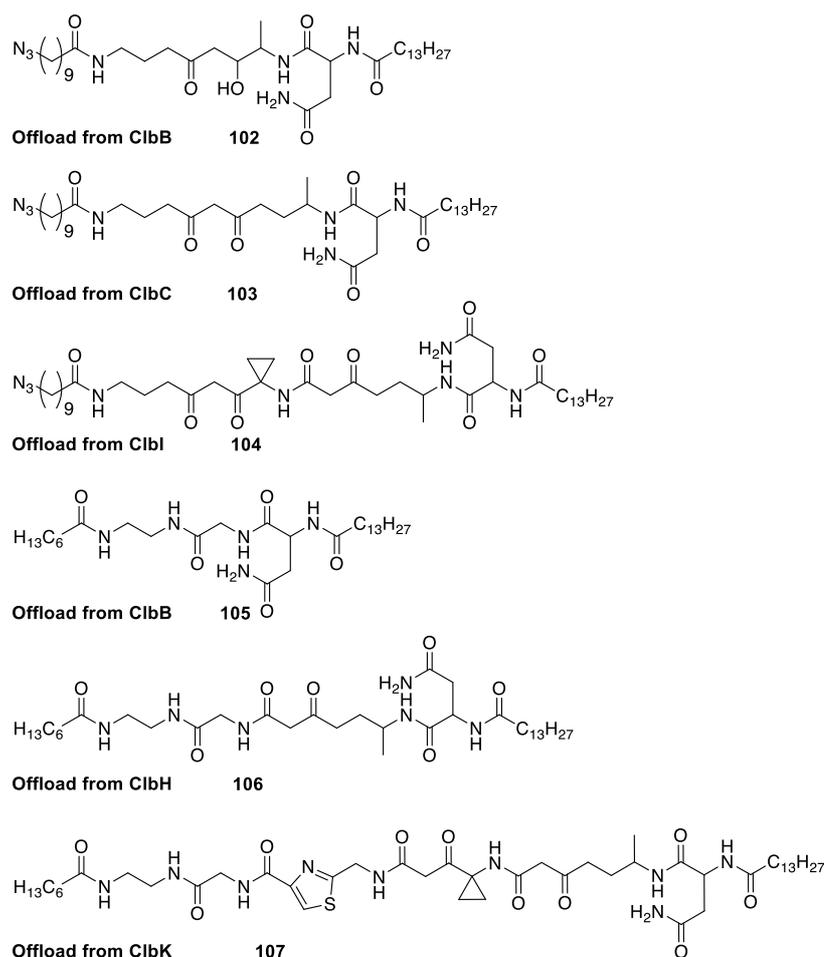


Figure 98 - Offloaded intermediates detected in previous work using chemical probes for both PKS and NRPS systems.

At the time of these investigations, the role of ClbL and ClbO remained unelucidated, and other details, such as the timing of oxidation of the second thiazole ring, were unclear. To try and clarify these unknown timings and extensions, a number of new chemical probes were synthesised and tested to try and gain additional information about the biosynthesis.

3.2.1 Cysteine-based chemical probes

Using a cysteine-based probe, **75**, previously tested in the echinomycin producer *S. lasaliensis*, it was hoped to identify the oxidation of the second thiazoline to a thiazole, and when this occurs in an aminomalonate containing precolibactin. NRPS ClbJ was known to install the first thiazoline, followed by incorporation of aminomalonate and a second thiazoline by PKS-NPRS ClbK.

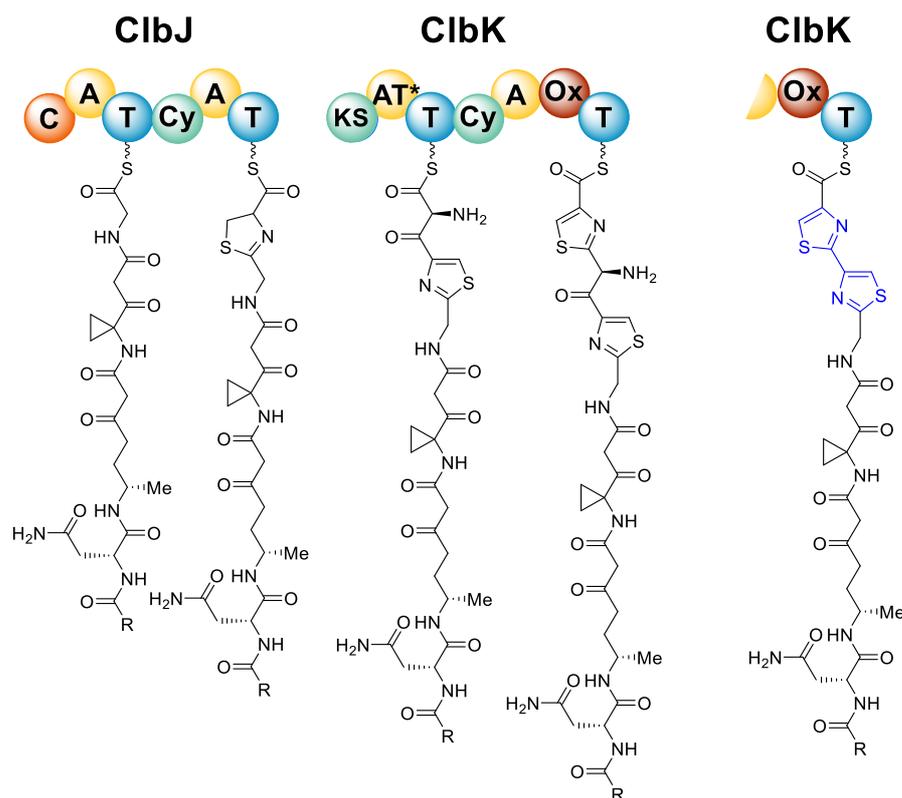


Figure 99 - Installation of thiazole moieties by ClbJ and ClbK, along with aminomalonate incorporation, with the possibility of module-skipping in ClbK.

In the bis-thiazole products identified, shown in blue in Figure 99, the oxidation of the first thiazole, formed by ClbJ, was found to be oxidised by ClbK before incorporation of a second thiazole by module skipping. However, the thiazoline could possibly undergo chain extension through the incorporation of an aminomalonate unit before reaching the T domain of ClbK, and thus be the true substrate. Additionally, by intercepting the biosynthetic pathway at ClbK, it may be possible to isolate a linear intermediate of colibactin containing an aminomalonate unit. The probe was synthesised as described in chapter 2.

Feeding experiments were performed in liquid culture of LB media, with *E. coli* strains grown to OD_{600} of 0.4 to 0.6, and 5 mL of LB media containing appropriate selection were supplemented with chemical probes at a 1 mM or 2 mM concentration. These cultures were incubated at 37 °C, 180 rpm for 24 hours before extraction with ethyl acetate (EtOAc). The extracts were concentrated and redissolved in MeOH before analysis on an Orbitrap Fusion LC-MS². In addition to the *E. coli* MG1655

pBAC *pks+* strain, feedings were also conducted in the pBAC *pks* Δ *clbA* strain lacking the phosphopantetheinyl transferase, and unable to produce colibactin. Production of colibactin and abolishment was confirmed by the presence of the prodrug motif, which can be seen in Figure 100. Strains containing insertional mutations in biosynthetic genes of interest were also examined, including *clbO(KS)*, *clbO(ACP)*, *clbQ*, *clbP*, and *clbL*.

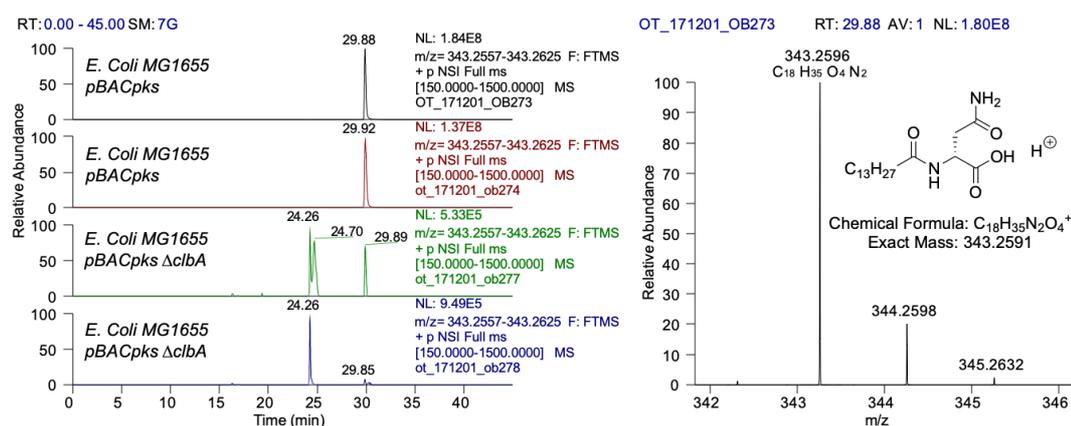


Figure 100 - Detection of the *N*-myristoyl-*D*-Asn colibactin prodrug motif used to verify production levels in *pks+* *E. coli* and the Δ *clbA* mutant strain. The prodrug motif elutes at 29.9 mins in *pks+* *E. coli*, and is absent in the Δ *clbA* strain.

Given the notable toxicity of the cysteine probes to *S. lasaliensis*, initial feeding experiments were performed at 1.0 mM concentrations. These found no intermediate species which were present in strains fed with chemical probes, and absent in the appropriate control experiments, in which the producing strain lacked chemical probe, or chemical probe was fed to a non-producing strain.

Due to the lack of offloaded species identified from these experiments, and the *E. coli* strains used appearing to tolerate the cysteine probe, the feedings were repeated at a 2.0 mM concentration. The feeding experiments revealed putative offloaded intermediates corresponding to the interception of the pathway after the installation of a thiazoline heterocycle, but once again no intermediates sufficiently advanced to allowed identification of the oxidation state of the first heterocycle in an aminomalonate containing colibactin could be detected.

While these intermediates, shown in Figures 101 and 102, had high-resolution masses corresponding to the molecular formula and weren't present in the relevant

control samples, only MS² fragments consistent with the probe partial structure could be assigned. It is possible that the cyclization pattern is different to that predicted which could explain the higher mass fragments, it is also possible that these ions are not from the intermediates postulated.

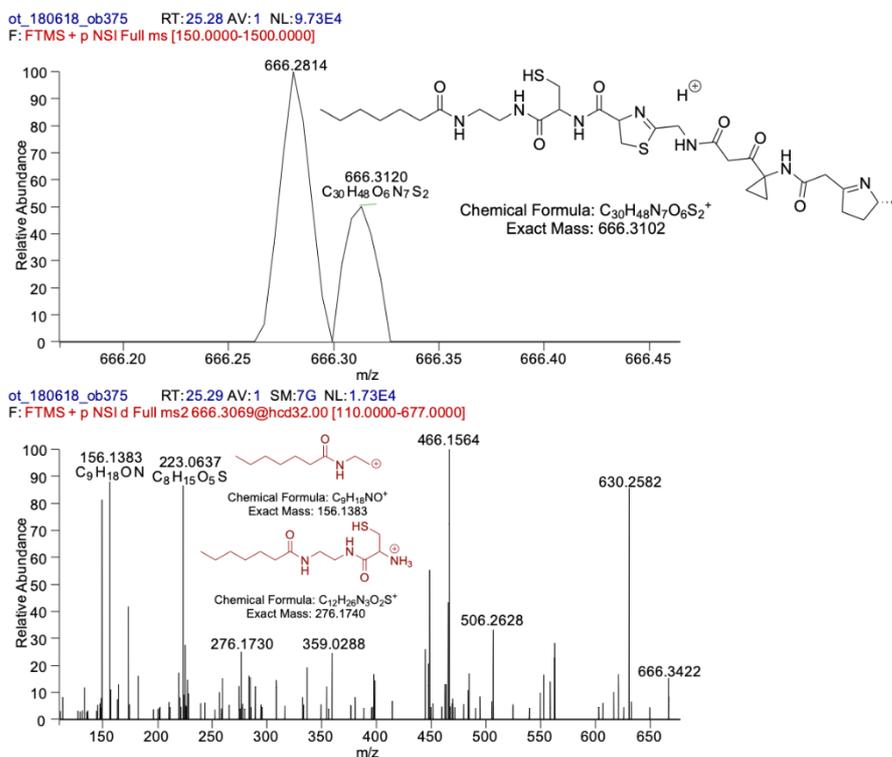


Figure 101 - Putative intermediate from *pks+* *E. coli* corresponding to offloading from ClbK, followed by dehydration after processing by ClbP.

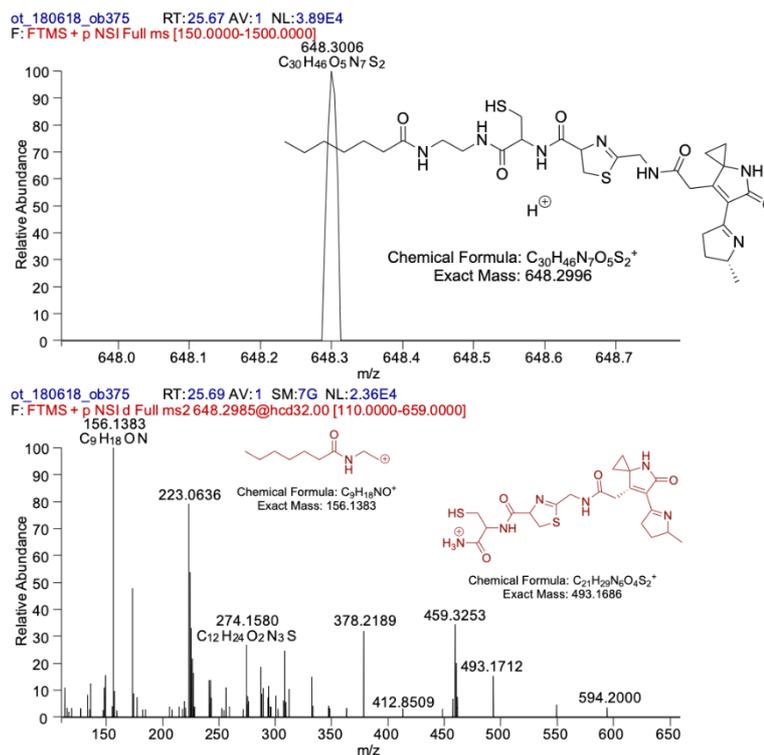


Figure 102 - Putative intermediate from *pks+* *E. coli* corresponding to offloading from ClbK, followed by double dehydration after processing by ClbP.

The absence of other identified intermediates could potentially due to a number of factors. As shown in Figure 103, the cysteine probe, **75**, tends to form the disulfide species during the course of the feeding experiments and subsequent extraction.

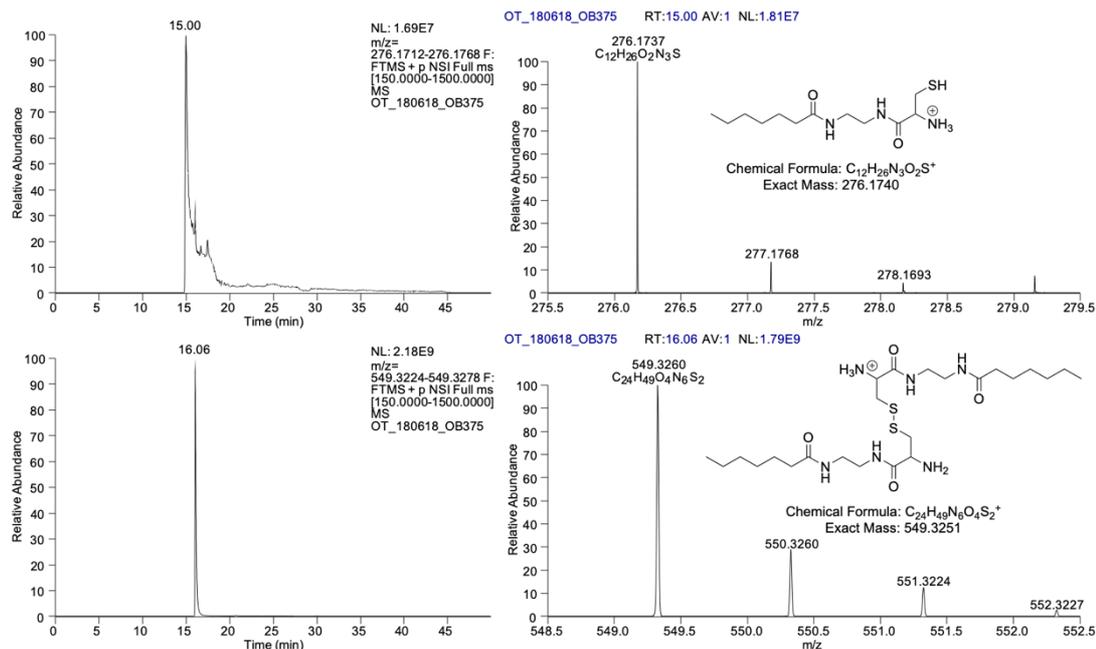


Figure 103 - High resolution mass spectra of the cysteine probe from culture extract (top), and the disulfide species from culture extract (bottom). The doubly-charged disulfide species can also be detected at a retention time of 16.1.

This will reduce the concentration of the species which mimics an aminoacyl extender unit. Also, disulfide formation after any offloading could result in the formation of a mixed disulfide, creating a species of unknown mass which will be far lower abundance therefore harder to detect, and also no longer subject to targeted MS² fragmentation.

These probes have a negligible impact on the production of NRPs in systems they have been used in such an echinomycin and vancomycin. This suggests that competition between the chemical probes and pathway substrate for intermediate species is dominated by the natural pathway. Combined with this low efficiency, colibactin is a natural product which is produced in small amounts. The amount of flux that passes through the NRPS modules are therefore low, reducing the opportunity to offload intermediates, and contributing to the low abundance of offloaded species.

Precolibactins have also been shown to demonstrate a range of spontaneous reactions after release from the biosynthetic enzymes, as described above for the case of pyridone and imine containing compounds. These cyclodehydration reactions will increase the complexity of the mixture being analysed, proving difficult in the targeted

approach used here. These reactions would also decrease the abundance of the already low abundance species, again making their detection and identification in a complex mixture more challenging.

3.2.2 Aminomalonate-based chemical probes

Previous work in the group by Dr Ina Wilkening and Jędrzej Goździk resulted in the synthesis of chemical probes based on the aminomalonate moiety, shown in Figure 104.

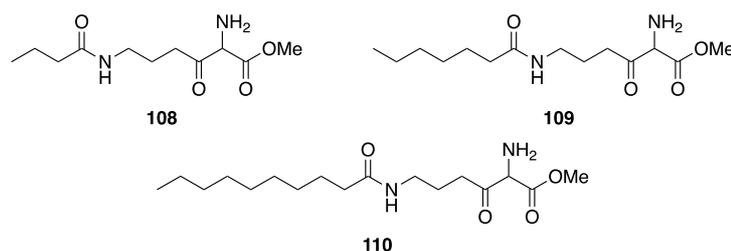


Figure 104 - Aminomalonate probes, 108 – 110, previously synthesized, protected with a methyl ester bearing C4, C7, or C10 acyl chains.

These probes were intended to undergo ester hydrolysis *in vivo* to generate an active probe species. This aminomalonate species was postulated to be able to act as a probe for PKS systems, as shown in Figure 105a., by mimicking a malonate extension unit, undergoing decarboxylative Claisen condensation and offloading the resulting intermediate. It was also proposed the decarboxylation could occur spontaneously after ester hydrolysis. This would then generate a probe containing an α -amino group, potentially mimicking a glycine residue, and allowing the interception of NRPS systems, as in Figure 105b.

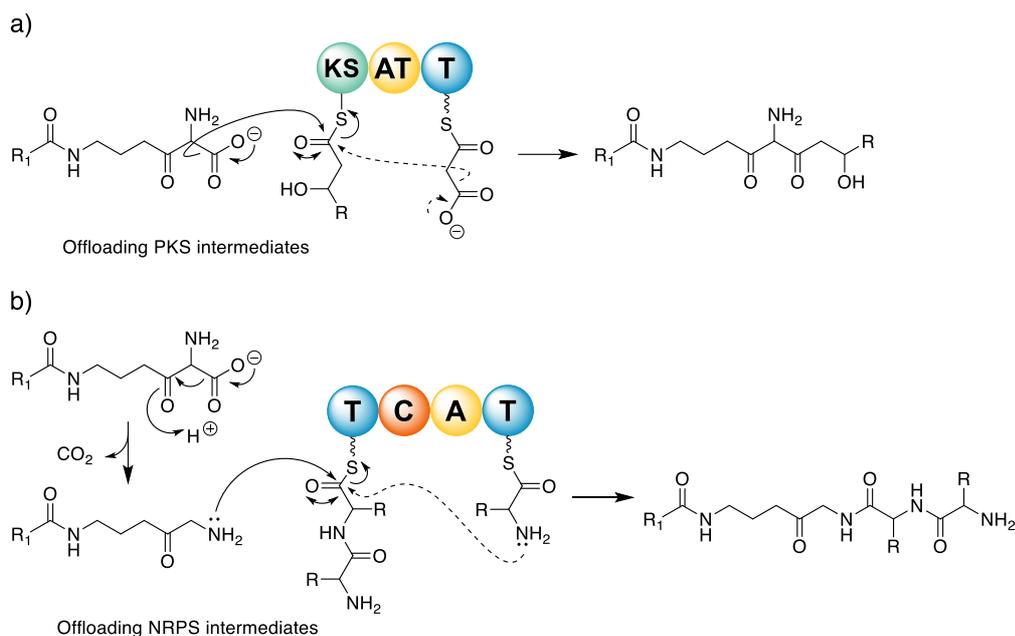
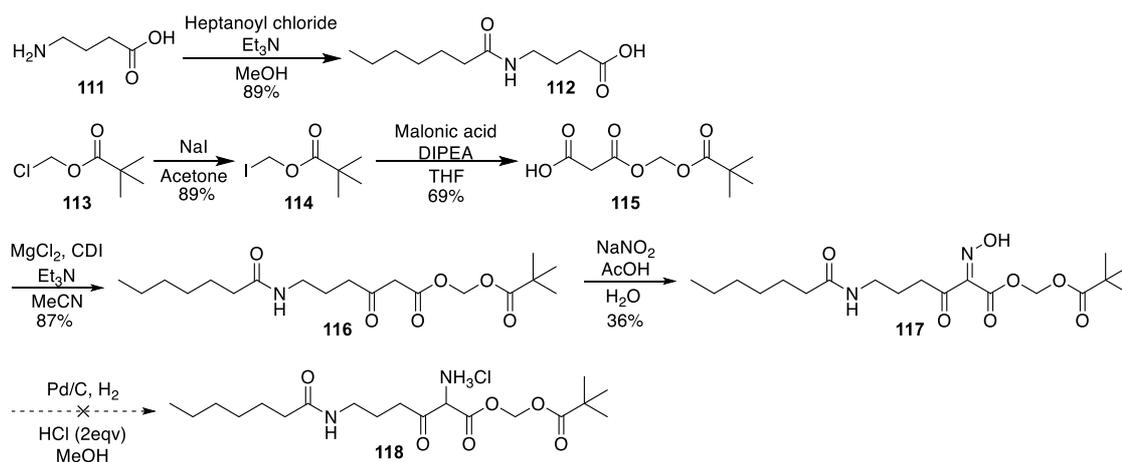


Figure 105 – Putative dual action of aminomalonnate-based chemical probes. a) Offloading of PKS modules via decarboxylative Claisen condensation. b) Decarboxylation followed by offloading from NRPS modules using the primary amine.

However, the initial aminomalonnate probes suffered from poor ester hydrolysis *in vivo*, resulting in a large amount of inactive probe and therefore poor intermediate capture. In PKS probes it has been shown that replacing the methyl ester with a more labile acetoxymethyl ester resulting in almost quantitative hydrolysis.¹⁵² However, these probes were significantly less stable, and liable to hydrolyse over time. It was therefore attempted to synthesise an aminomalonnate probe bearing a pivaloyloxymethyl ester, which while bench stable, would more readily hydrolyse and allow their testing as dual-action chemical probes. A synthetic route was devised, shown in Scheme 5.



Scheme 5 - Attempted synthetic route towards an aminomalonate probe protected with a pivaloyloxymethyl ester.

Initially, heptanoyl chloride, was used to couple with γ -aminobutyric acid, **111**, to give carboxylic acid, **112**. In parallel, a Finkelstein reaction was performed on pivaloyl chloride, **113**, using sodium iodide to afford the iodide, **114**, in excellent yield. Malonic acid was used to displace the iodine to give the malonic pivaloyloxymethyl ester, **115**. This was combined with the acid, **112**, which was activated by 1,1'-carbonyldimidazole, in a magnesium-mediated decarboxylation reaction, to afford the β -ketoester, **116**.²²³ The oxime functionality was then introduced at the α -position using sodium nitrite in acidic conditions, generating nitrous acid *in situ*, to form the oxime, **117**.

However, the final reduction proved troublesome and could not be optimised. Reduction of the oxime to the amine using Pd/C and H₂ were unsuccessful, and no product could be isolated. An ion corresponding to the [M+H]⁺ adduct could be detected in mass spectrometry, but attempts to purify the product showed that it degraded on silica. Attempts to work up the reaction mixture also resulted in degradation of the product. HPLC purification of the reaction mixture after filtration through Celite showed that separation of the product from other reaction components was possible, shown in Figure 106.

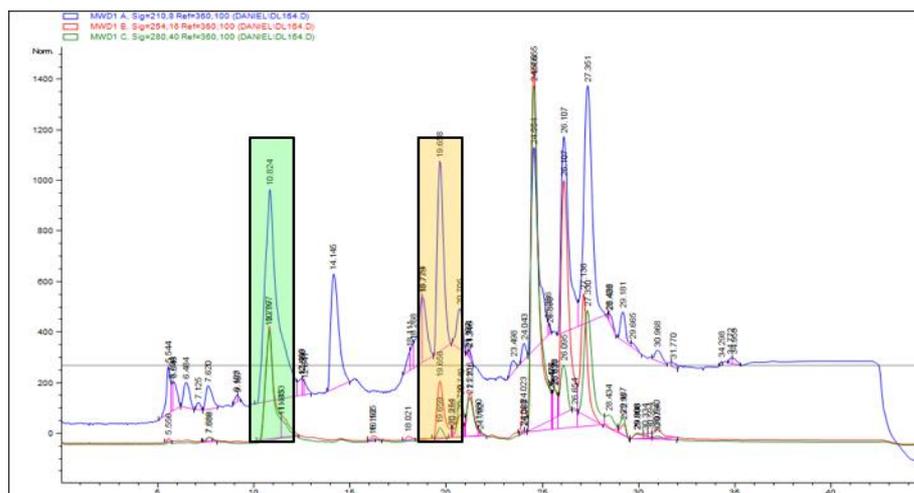


Figure 106 - UV traces at 210 nm, 254 nm, and 280 nm from the HPLC separation of the reaction mixture from the reduction of oxime, **117. Starting material is highlighted in orange, while the amine product is highlighted in green.**

However the UV trace, shown in Figure 106, showed a number of other compounds were present even before the complete consumption of the starting oxime. Mixing the starting material with 3 equivalents of hydrochloric acid showed minimal degradation after 6 hours, suggesting hydrolysis of the pivaloyloxymethyl ester is not a significant contributor to these side products. Indeed, adding a large quantity of HCl after filtration through Celite reduced the amount of decomposition which occurred.

The number of possible alternate strategies for the reduction of the oxime, **117**, were limited by the presence of the ketone which will be reduced before the oxime in a range of common reduction conditions. Zn/AcOH was explored, but the reaction proceeded very slowly, measured by the consumption of starting material. None of the desired product could be identified from the resulting complex reaction mixture. ^1H NMR of the mixture formed from the Pd/C and H_2 reduction showed the splitting of the *tert*-butyl group into a mixture of singlets. Combined with the necessity of keeping the compound acidic, and the amine therefore protonated, it is likely that the nucleophilic amine intramolecularly attacks the pivaloyl carbonyl group. This would explain the splitting of the ^1H NMR signals, and the need for acidic conditions, but the product of the reaction has not been isolated.

Given the requirement of an extremely acidic environment to prevent degradation of the probe, it was decided not to be compatible with the biological work and corresponding pH conditions for which it was designed. Alternate probes were therefore designed, as shown in Figure 107, with the synthesis shown in Scheme 6.

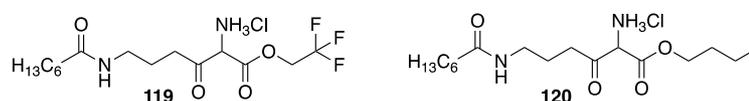
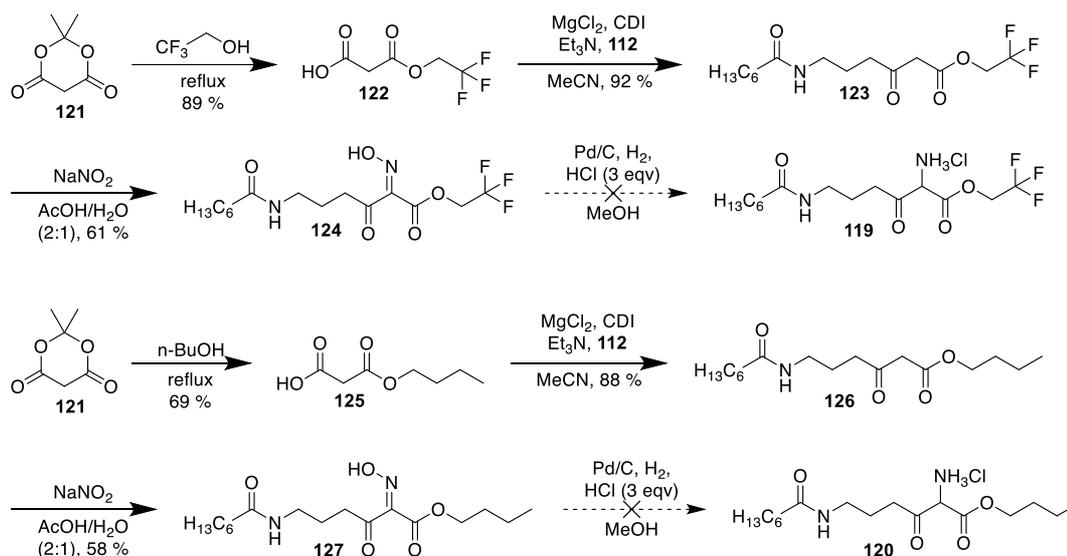


Figure 107 - Aminomalonate-based probes containing trifluoroethanol or butanol esters.



Scheme 6 – Attempted synthetic route to aminomalonate probes protected by trifluoroethanol, 119, or butanol esters, 120, instead of the potentially reactive pivaloyloxymethyl ester.

Both of these probes were hypothesised to be more likely to undergo ester hydrolysis than the equivalent methyl esters, while not allowing the intramolecular attack of the amine moiety which was likely problematic for the pivaloyloxymethyl ester. Butanol esters were proposed to act as better substrates for intracellular esterases, while the fluorine in trifluoroethanol should exert an electron-withdrawing effect which should generate a more labile ester.²²⁴ As shown in Scheme 6, malonic half esters of trifluoroethanol, **122**, and butanol, **125**, could be generated by reaction with Meldrum's acid, **121**, before decarboxylative coupling to carboxylic acid **112**. Installation of the oxime proceeded well as before to give the α -oxime- β -ketoesters **124** and **127**. Reduction of the oxime moiety was again problematic. The Pd/C catalysed reactions were characterised by incomplete conversion and messy reaction

mixtures. ^{19}F NMR of the mixture resulting from the reduction of **124** showed evidence of trifluoroethanol, suggesting a degree of hydrolysis or transesterification. The desired product could not be purified from the reaction mixture due to the need to keep the amine protonated, given the known ability of α -amino- β -ketoesters to undergo spontaneous dimerisation, resulting in pyrazine formation after oxidation, as shown in Figure 108.²²⁵

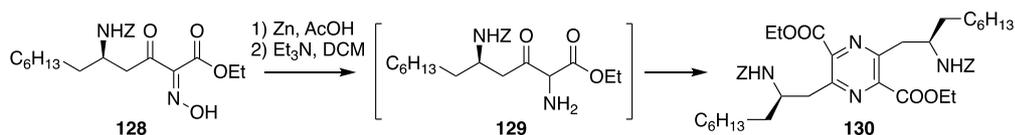
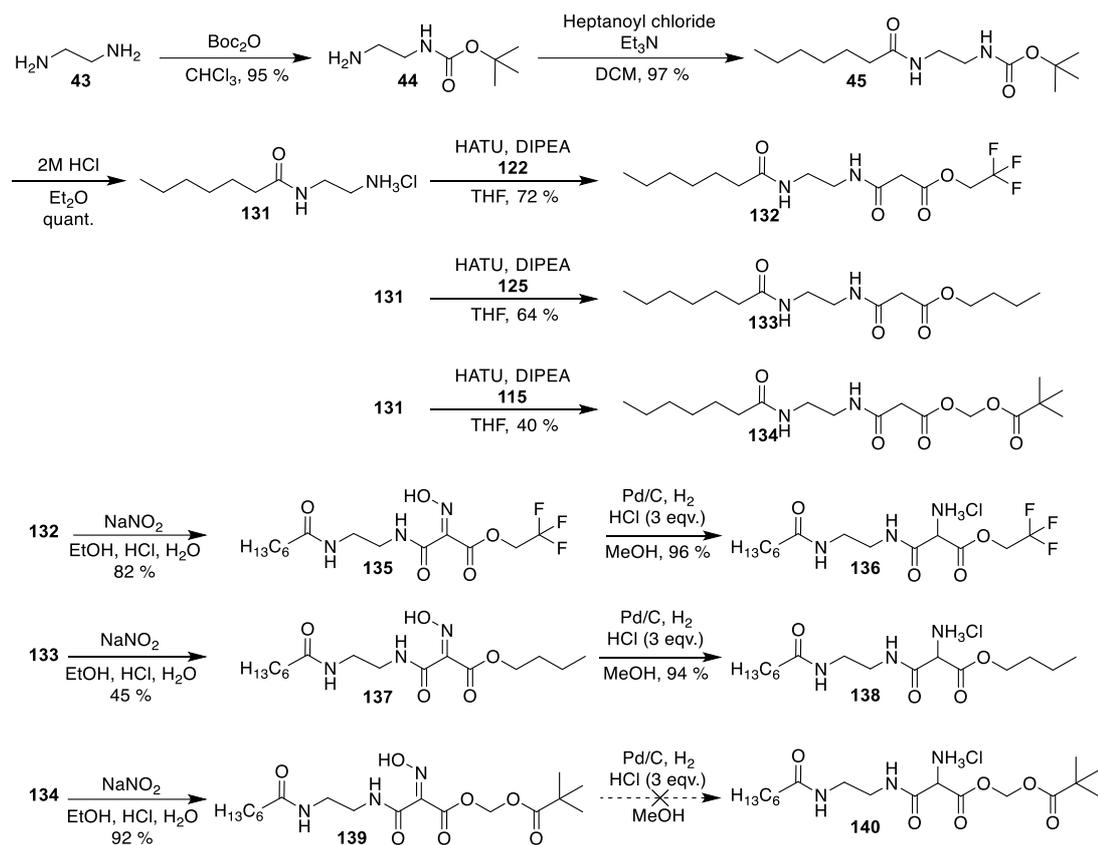


Figure 108 - Formation of pyrazines from cyclisation and oxidation in air of α -amino- β -ketoesters in basic conditions, from Buron *et al.*²²⁵

To produce a stable probe which could be used to investigate aminomalonate usage in the colibactin biosynthetic pathway, it was decided to replace the β -ketone with an amide. Following the synthetic route shown in Scheme 7, both the butanol and trifluoroethanol esters could be synthesised, but the pivaloyloxymethyl ester remained synthetically inaccessible.



Scheme 7 - Synthesis of aminomalonate probes protected by trifluoroethanol and butanol esters, containing an amide in place of the β -keto functionality.

As with previous NRPS probes, ethylenediamine, **43**, is monoprotected with Boc anhydride, and acylated with heptanoyl chloride. The Boc group is then removed with HCl. The amine hydrochloride salt, **131**, is then coupled to the appropriate malonic half ester using HATU to give esters **132**, **133**, and **134**. The oximes are then formed using sodium nitrite in EtOH, HCl and H₂O. Using AcOH as with the β -ketoester compounds resulted in no reaction. The amide containing oximes **135** and **137** could then be reduced using Pd/C and H₂ to give the aminomalonate based probes **136** and **138**. The reduction of the pivaloyloxymethyl ester compound **139** resulted in none of the desired product, likely due to intramolecular reactions of the newly formed amine.

Feeding of aminomalonate probes **136** and **138** were hoped to allow the offloading of both PKS and NRPS intermediates from the colibactin pathway. At the same time, the probes should mimic the aminomalonate extender unit that is utilised

by ClbK and was hypothesized to be installed by ClbO. Feeding experiments with both probes resulted in no detectable offloading from either PKS or NRPS modules. Poor hydrolysis, which was an issue for previous aminomalonate probes, could be an issue for these probes. Figure 109 shows the comparison of the ion intensities of the probes in extracts with the intensities of the decarboxylated probes, **81**, showed that the trifluoroethanol ester **136** showed a greater degree of hydrolysis than the butanol ester **138**.

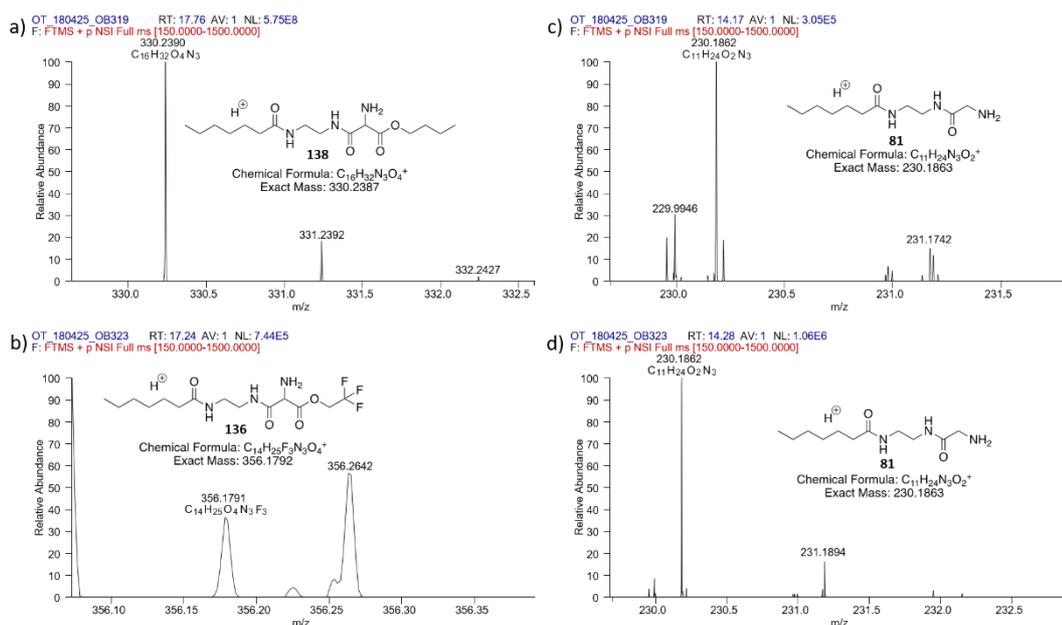


Figure 109 - Mass spectra of the probes 138 (a) and 136 (c), and their decarboxylated derivatives 81 (c and d). The intensity of the butanol ester 138 is several orders of magnitude higher than the trifluoroethanol ester 136, suggesting worse hydrolysis *in vivo*.

Offloading from the PKS by the hydrolysis product of the probes could also be limited by the short half-life of the product. The short lifetime of the aminomalonate species in the pathway has been noted in *in vitro* work, marked by the detection of glycyl-ACP rather than AM-ACP.⁹⁵ In synthetic work it has been noted that the half-life of α -amino- β -ketoesters is pH dependent, and less than one minute at pH 7.0.²²⁶ This would naturally reduce the amount of probe that can undergo decarboxylation in the ketosynthase (KS) active site, and thus reduce the likelihood of offloading pathway intermediates.

A short half-life would increase the amount of active NRPS probe that is formed *in vivo*, but the total amount would still be significantly lower than in feeding experiments using the equivalent glycine-based probe, **81**.

While the replacement of the ketone with an amide is expected to increase the stability of the carboxylic acid formed, this could reduce the amount of decarboxylation and therefore offloading from the PKS pathway, and the amount of the NRPS active species in the culture, resulting in poor offloading from the NRPS pathway. While these issues could potentially be mitigated by increasing the concentration of the probe, this could lead to toxicity issues, and is likely less effective than feeding with separate probes specifically designed for PKS and NRPS systems.

3.2.3 PKS and NRPS chemical probes

To investigate the utility of using chemical probes designed for PKS pathways in combination with NRPS chemical probes, feeding experiments were conducted using two probes simultaneously; glycine-based probe **81**, and fluoromalonate-based chemical probe, **141**, previously synthesized in the group, shown in Figure 110.

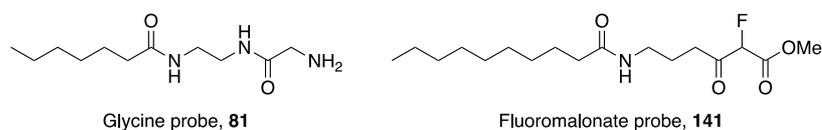


Figure 110 - The glycine-based chemical probe 81 and the fluoromalonate-based chemical probe 141, prepared by Robert Jenkins, used to investigate colibactin biosynthesis.

Given the lack of late-stage intermediates detected which could help assign roles to biosynthetic enzymes of unknown function, *E. coli pks+* and $\Delta clbA$ were used for these experiments. While ClbJ does incorporate glycine, the glycine probe **81** has previously proved a useful tool to offload from a range of NRPS modules, including those which it is not the cognate substrate for. The fluoromalonate probe has proved a useful tool for the investigation of a number of polyketide synthetase systems, notably in the thiotetronate antibiotic thiolactomycin.²²⁷

In spite of these uses in other systems, the feeding experiments using both probes led to the detection of just one putative intermediate, shown in Figure 111, from the fluoromalonate probe. Even in this case, the only MS² fragment which can

be assigned is derived from the probe itself. As a result, the nature of the compound cannot be verified, but given the mass is absent in *pks+* strains without the probe, and the $\Delta clbA$ strain with the probe, it is likely to be related to the colibactin pathway. Based on the exact mass and the colibactin biosynthetic pathway, it is likely a result of interception of chain extension after ClbK, with ClbP processing and subsequent dehydration. This would suggest that the fluoromalonate probe can compete with aminomalonate-ACP at the KS of ClbO.

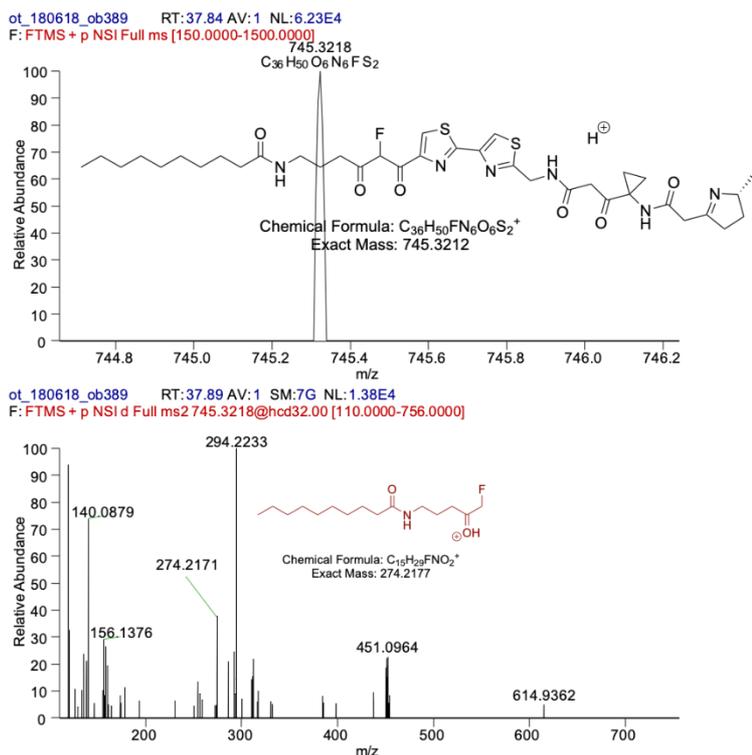


Figure 111 - The only putative intermediate detected in feedings conducted with the fluoromalonate probe 141 and glycine probe 81. Only a single fragment in the MS² spectra attributable to the probe can be assigned.

It was reasoned that as the *E. coli* strains that produce colibactin are grown for 24 hours, in comparison to the five days of *Streptomyces* strains, there may be reduced opportunity for offloaded intermediates to accumulate. The feeding experiments with the cysteine-based probe, **75**, the glycine-based probe, **81**, and fluoromalonate-based probe, **141**, were therefore repeated with 72 hours incubation.

In the case of the cysteine-based probe, there were no captured intermediates detected, suggesting that the problems with the probe are too pronounced to be solved

with longer incubation. However, in feedings using the glycine-based probe, **81**, new intermediate species could be detected, shown in Figures 112.

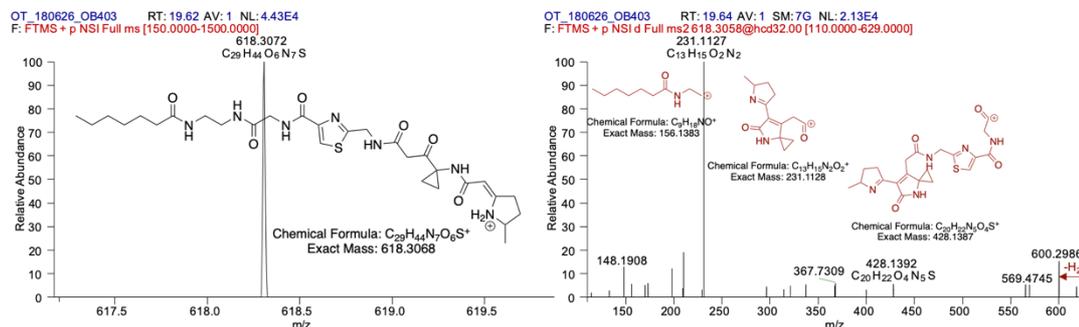


Figure 112 - Putative offloaded intermediate and MS² fragmentation from feeding with the glycine-based probe **81 over 72 hours.**

The new intercepted intermediate would correspond to the offloading at the NRPS of the PKS-NRPS ClbK, after undergoing module skipping as seen in Figure 99, without the incorporation of amino malonate. MS² fragments would support this structure, with additional dehydration during fragmentation.

The chemical probing of colibactin to elucidate unknown steps in the assembly was partially successful in identifying a possible extension of ClbO, a PKS module of unknown function. While the probes based on an aminomalonnate-like moiety, **136** and **138**, were unable to capture any intermediates, the species captured by the cysteine-based probe **75** and glycine-based probe, **81**, are consistent with the colibactin pathway. The conclusions of this work will be discussed further in chapter 6.

3.3 Following work

Following the work described above to elucidate late-stage steps in colibactin biosynthesis, further work both *in vitro* and *in vivo* was conducted to determine the roles of ClbL and ClbO and solve the structure of colibactin.

While the structure of precolibactin-886, **91**, did not require the use of the PKS ClbO, a compound was identified in cell extracts at *m/z* 970 which was lost in the $\Delta clbPQO$ triple mutant strain. Precolibactin-969 is shown in Figure 113. This structure could potentially incorporate a second aminomalonnate from ClbO. To increase the titre of this product, the self-resistance gene *clbS* was deleted, preventing hydrolysis of the

cyclopropane ring by ClbS.^{228,229} In a $\Delta clbPQS$ strain, the metabolite was identified and shown to contain an additional C_3HNO_2 compared to precolibactin-886, consistent with aminomalonate incorporation. This compound can be hydrolysed by ClbP, and was shown to be able to nick and break DNA in the presence of Cu(II). However, this does not account for the role of ClbL, or the role of ClbQ, which was shown to be essential for megalocytosis activity.

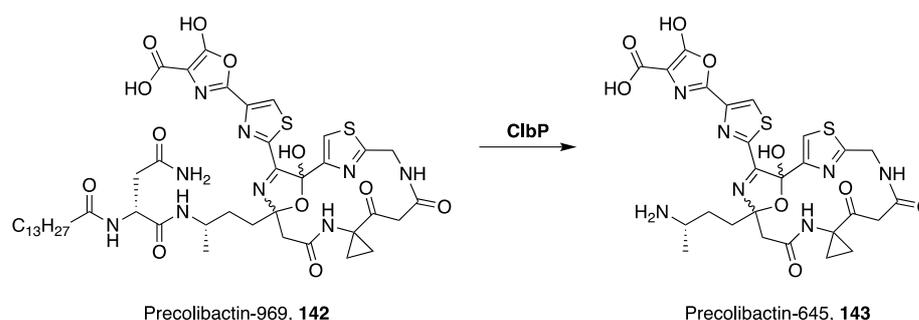


Figure 113 - Precolibactin-969, 142, identified from $\Delta clbPQS$ *E. coli* strain, and the structure of proposed colibactin-645, 143, produced by the action of ClbP on 142.

The role of ClbL was elucidated when it was noticed that while no new precolibactins could be identified in a $\Delta clbL$ strain, a new indole-containing metabolite, **144**, was produced in strains expressing ClbL shown in Figure 114. A combination of MS² and ¹³C labelling showed this was a product of the colibactin pathway, confirmed by synthesis of an authentic standard. This contained no amine or carboxylic acid, as would be expected from ClbLs predicted role as an amidase.

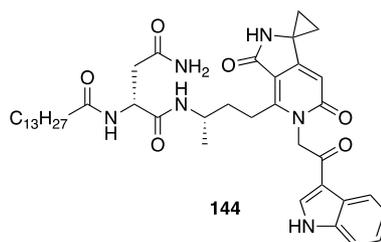


Figure 114 – 144, a metabolite which accumulates in the ClbL expressing *E. coli* strain, but not in $\Delta clbL$ strains.

It was determined that an aminoketone could act as a nucleophile towards a ClbI-bound thioester.²³⁰ This could be reconstituted *in vitro*, suggesting ClbL catalyses amide bond formation. An aminoketone could be derived from the

decarboxylation of an aminomalonate moiety, and indeed ClbL preferentially used a mimic of the aminoketone which would derive from the decarboxylation of the ClbO product, as in Figure 115. Combining the products of ClbO and ClbI would give the structure, **145**, two cyclopropane warheads, consistent with the colibactin-induced interstrand cross-links. Additionally, a new DNA-adduct, **146**, could be identified which was consistent with the proposed structure, shown in Figure 115.

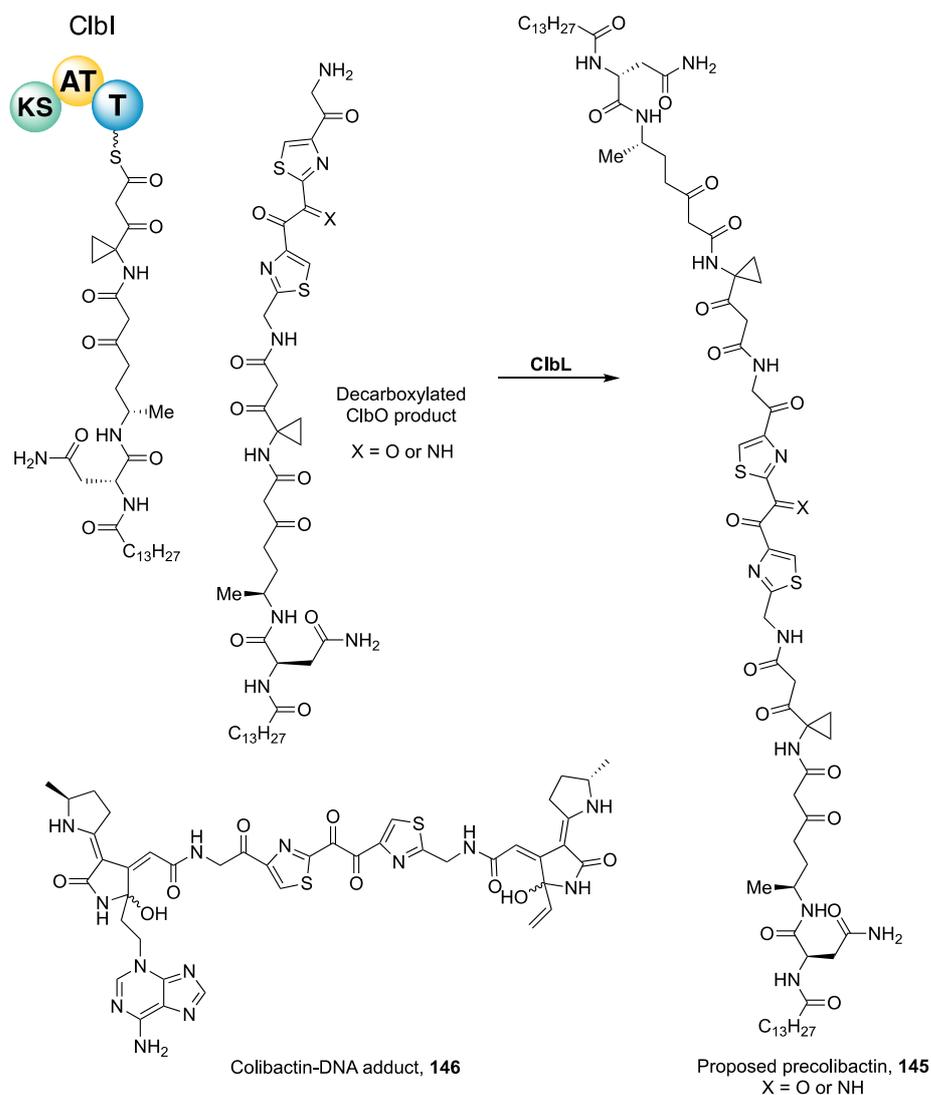
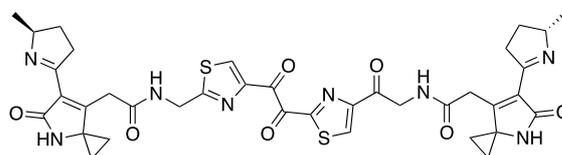


Figure 115 - Action of amidase ClbL which catalyses amide bond formation between the decarboxylated product of ClbO chain extension and thioester-bound ClbI product to generate a pseudo-dimeric precolibactin, 145. A larger DNA adduct, 146, containing multiple warheads could be identified from untargeted DNA adductomics.

Simultaneously, a second approach towards the final structure of colibactin was being pursued. Given the struggle to isolate colibactin, an effort was made to elucidate the DNA-colibactin cross-linked structure. Adducts were studied by comparing unlabelled strains with labelled auxotrophic *E. coli* strains.¹⁰¹ Mass-shifts in Cys or Met fed cultures identified metabolite containing two thiazole and two cyclopropane rings. Ser and Gly feeding showed this metabolite incorporated glycine, and three carbons and one nitrogen from serine. Labelled glucose and ammonium nitrate confirmed the structure contained 37 carbon atoms, and 8 nitrogen atoms. Fragmentation showed the neutral mass loss of two adenine and two hydroxyl groups.

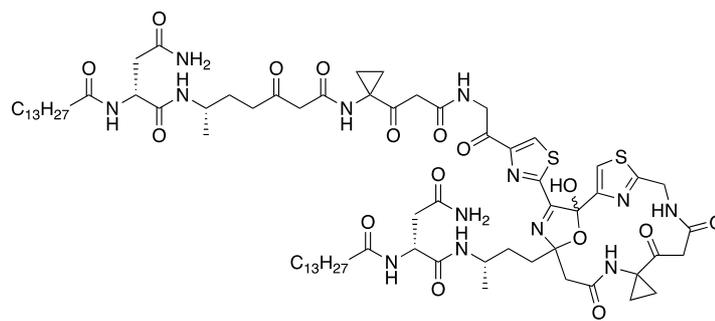
A structure was proposed, taking the labelling experiments, MS², and proposed roles of the biosynthetic enzymes into account, to give the structure of colibactin, **147**, shown in Figure 116. This showed the C36-C37 bond is susceptible to hydrolytic cleavage, after the oxidation of C37. The α -ketoimine was detectable in *clb+* cultures, but not $\Delta clbO$, *clbL* point mutants, or a *clb-* strain. This would explain the difficulty in isolating colibactin, and the degraded nature of the DNA adducts discovered. The structure was confirmed by chemical synthesis, which produced identical MS² to the natural colibactin, and produced identical DNA adducts.



Colibactin, **147**

Figure 116 - Structure of the genotoxic natural product colibactin, 147.

The expected linear precursor to colibactin, precolibactin-1491, could not be detected, but a precolibactin-1489 could be found shown in Figure 117, a predicted macrocycle related to precolibactin-886 and 969. This structure, supported by labelling experiments and MS², would require all genes in the *clb* cluster. It is proposed to be an off-pathway product of oxidation and macrocyclisation of the linear precursor. The full pathway is shown in Figure 118, and the significance of the assigned roles of ClbO and ClbL will be discussed in the context of chemical probing of the pathway in chapter 6.



precolibactin-1489, 148

Figure 117 - Precolibactin-1489, 148, macrocycle derived from the cyclisation and oxidation of linear precolibactin.

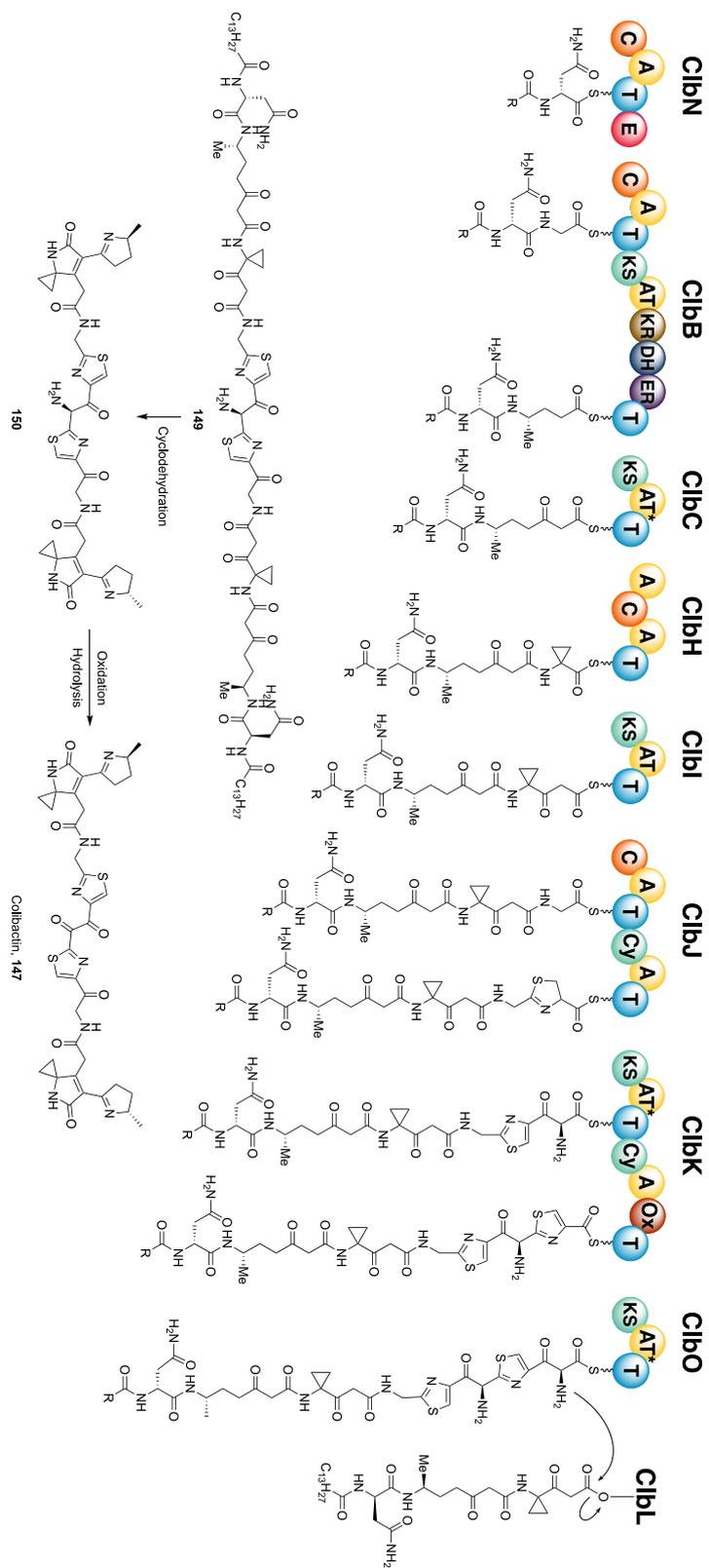


Figure 118 - Biosynthetic pathway of colibactin accounting for all PKS, NRPS, and PKS-NRPS modules (2019).

4. Investigation of vancomycin bio-assembly *via* chemical probes

4.1 Vancomycin discovery

Vancomycin was first isolated from cultures of *Streptomyces orientalis*, later renamed *Amycolatopsis orientalis*, a strain found in the soil of Borneo in 1956. It was the first glycopeptide antibiotic discovered. A semi-purified form was used by Eli Lilly and noted that bacterial resistance did not easily develop. The structure however would not be determined for over 20 years.

It was noted that vancomycin inhibited the cell wall mucopeptide, stopping growth of the cell and eventually leading to cell lysis. The binding was found to be specific to the D-alanyl-D-alanine terminus, and could form a complex with the nucleotide-*N*-acetyl muramyl pentapeptide precursor.^{231,232} However the method of binding of D-Ala-D-Ala, essential to the mechanism of action, could not be determined until the structure was solved.

The molecular formula could be determined by elemental analysis, and chemical degradation was used to obtain additional structural information. Acid hydrolysis of vancomycin was found to give aspartic acid and glucose. Milder hydrolysis was able to liberate leucine, which IR suggested possessed a secondary amine, *N*-methyl leucine. Oxidation of vancomycin using permanganate or nitric acid was able to give information on the connectivity of the phenol groups.²³³ Acid hydrolysis could also cleave vancosamine, the structure of which could be determined by NMR.^{234,235} It was determined that the three ring system was connected by ether linkages through oxidative cleavage with permanganate.²³⁶ Analysis of the NMR spectra suggested the incorporation of modified phenylglycine and β -hydroxytyrosine residues.²³⁷ All these elements could only be combined when the structure of a mild hydrolysis product, in which the asparagine residue is hydrolysed to give aspartic acid, was solved by X-ray diffraction in 1978 in complex with D-Ala-D-Ala to give the structure shown in Figure 119.²³⁸

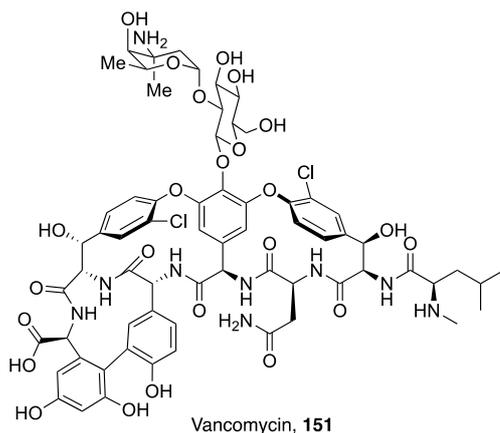


Figure 119 - The structure of glycopeptide antibiotic vancomycin, confirmed by X-ray diffraction in 1978.

4.2 Vancomycin activity

The crude isolate from *Amycolatopsis orientalis* was found to be effective against a wide range of Gram positive bacteria, notably penicillin resistant *Staphylococcus*.²³⁹ Vancomycin has since been deemed an essential medicine by the World Health Organisation, and is used for the treatment of methicillin-resistant *Staphylococcus aureus* (MRSA) and methicillin-resistant *Staphylococcus epidermis* (MRSE), along with *Clostridium difficile*. Use is also advised in the case of penicillin or β -lactam allergies, and as a drug of last resort against Gram positive bacteria.²⁴⁰

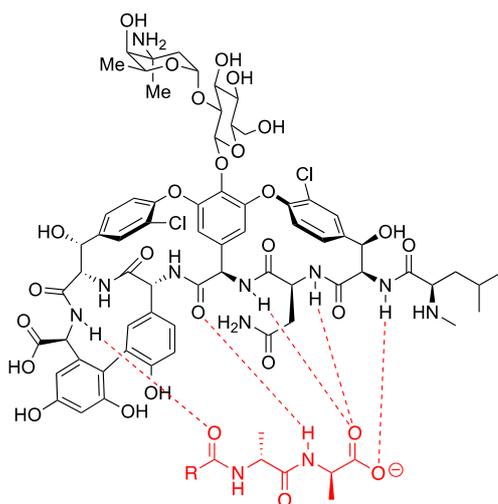


Figure 120 - Vancomycin binding the terminal D-alanyl-D-alanine motif of the mucopeptide through five hydrogen bonds.

Vancomycin acts to inhibit cell wall biosynthesis, which in Gram positive organisms is composed of a thicker, more permeable peptidoglycan layer. Gram negative bacteria are not susceptible as their cell wall is protected by an outer membrane.²⁴¹ Treatment with vancomycin leads to an accumulation of cell wall precursors, indicating it acts at a late stage. UDP-muramyl-NAc-pentapeptide forms a 1:1 complex with vancomycin, binding through five hydrogen bonds to the D-Ala-D-Ala motif, shown in Figure 120.²³¹ The additional steric bulk imparted by complexation with vancomycin prevents transglycosylation from proceeding, preventing the polymerisation of the precursor. Even in the event that transglycosylation could continue, transpeptidation would be inhibited as the terminal D-Ala residue is inaccessible, which is shown in Figure 121.^{242,243}

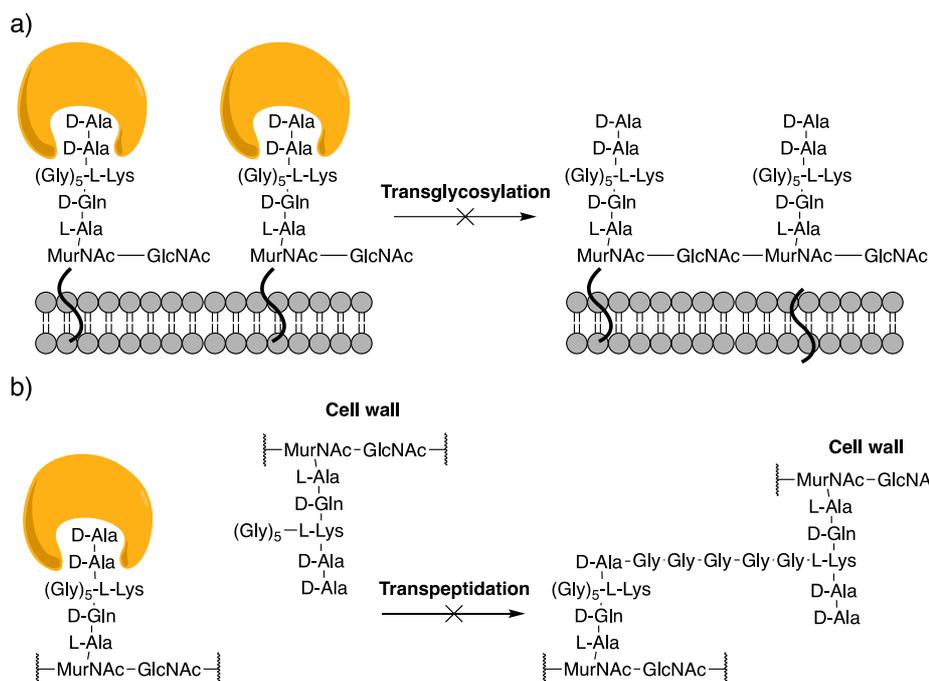


Figure 121 - Vancomycins inhibition of cell wall biosynthesis by preventing transglycosylation (a) and transpeptidation (b), with vancomycin represented by the orange shape.

4.3 Antimicrobial resistance to vancomycin

Resistance in the clinic to vancomycin was reported 30 years after discovery, in 1986.²⁴⁴ Currently there are six resistance loci that have been reported, *VanA-E*, *VanG*.²⁴⁵ Each of these six loci encode for one of two pathways for glycopeptide resistance. These are responsible for the alteration of the terminal D-Ala-D-Ala motif to either D-Ala-D-Lactate in the case of *VanA*, *B*, and *D*, or D-Ala-D-Ser for *VanC*, *E*, and *G*. Both of these changes reduce the binding affinity to vancomycin by reducing the number of hydrogen bonds, shown in Figure 122, thus conferring resistance.

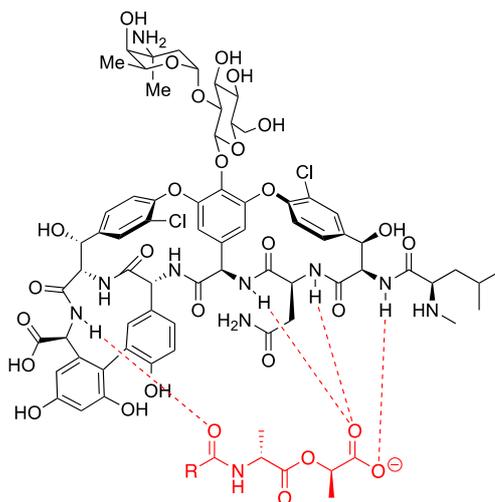


Figure 122 - The altered binding of vancomycin to D-Ala-D-Lac in resistant strains, with the loss of one hydrogen bond and a 1000-fold drop in binding affinity.

Resistant strains produce a suite of proteins to effect this change: ligases to couple D-Lac to D-Ala, dehydrogenases to produce D-Lac from pyruvate, and dipeptidases to hydrolyse D-Ala-D-Ala. *VanC* phenotypes produce a ligase, dipeptidase, and a serine racemase to enable production of D-Ala-D-Ser. *VanC* strains however remain susceptible to teicoplanin, shown in Figure 123, which does not induce the transcription of these genes, and subsequent cell wall remodelling.²⁴⁶

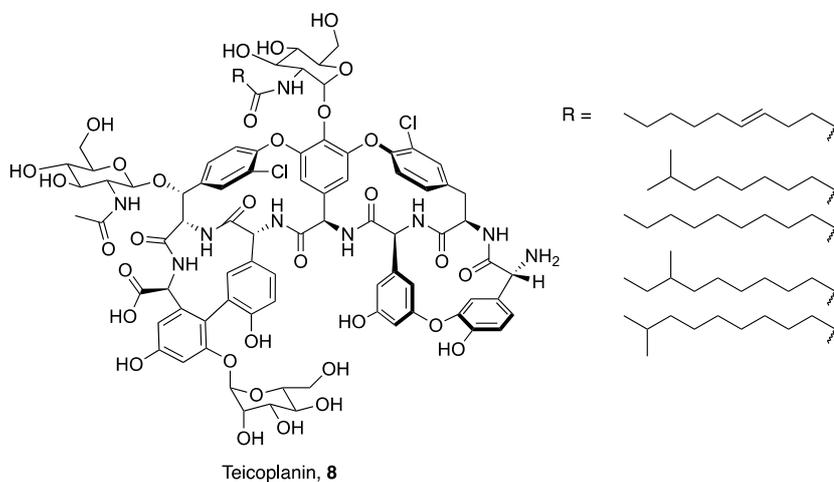


Figure 123 - The structure of glycopeptide antibiotic teicoplanin, which is sufficiently different to avoid triggering transcription of resistance genes in *VanC* phenotypes.

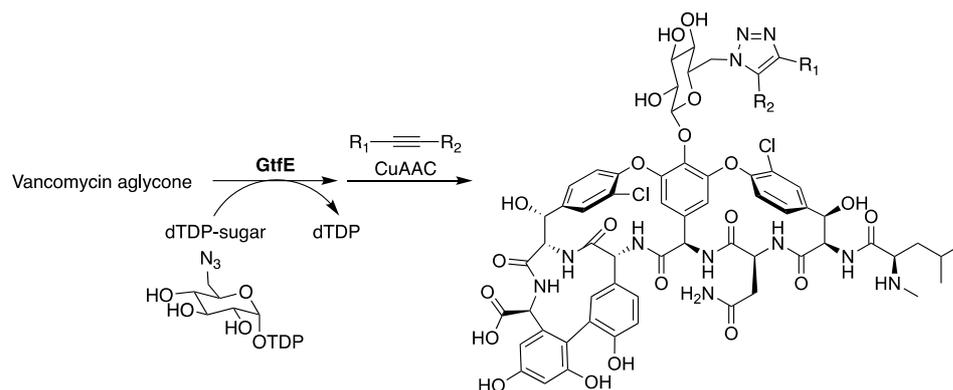


Figure 125 - Functionalised vancomycin derivatives using GtfE to install an azido-sugar, followed by a click CuAAC reaction to install further complexity.

The inherent substrate flexibility of glycosyl transferases also allow the generation of unnatural analogues through glycorandomization, exploiting either new aglycones or novel sugar moieties, shown in Figure 124.^{249,250} Novel sugars have been exploited to conduct site-selective chemical synthesis through Cu(I)-catalysed azide-alkyne cycloaddition (CuAAC) with azido-sugars, shown in Figure 125, to efficiently introduce a range of substituents including a β -lactam, to create a hybrid antibiotic with improved activity against *Enterococci*.²⁵¹ The use of acyltransferases has also been combined with glycosyl transferases to produce new lipoglycopeptides.²⁵²

The strain *Amycolatopsis balhimycina*, formerly *Amycolatopsis mediterranei*, produces the vancomycin-family glycopeptide balhimycin. In a mutant strain unable to produce β -hydroxytyrosine (β HT), supplementation of the strain with 3-fluoro- β -hydroxytyrosine, **154**, led to the production of a difluorinated balhimycin analogue, **155**, through a mutasynthetic approach, shown in Figure 126. 2-fluoro- β HT and 3,5-difluoro- β HT were also tolerated.²⁵³ This mutasynthesis could be extended to the AB biaryl ring system in $\Delta dpgA$ mutants, allowing further unnatural balhimycins to be produced.²⁵⁴

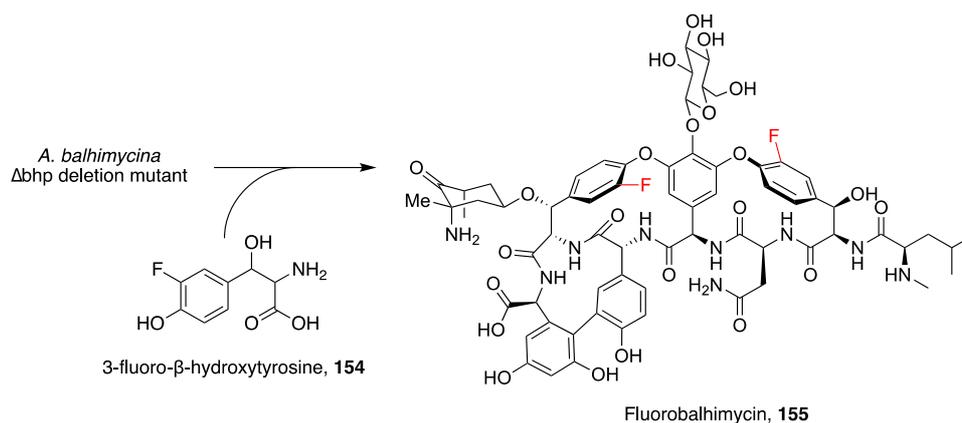


Figure 126 - Generation of fluorobalhimycin through a mutasynthetic approach, feeding 3-fluoro- β -hydroxytyrosine to a Δbhp mutant strain incapable of producing β HT.

4.4.2 Semi-synthetic derivatives of vancomycin

Some of the greatest success in vancomycin derivatisation so far has come from semi-synthetic routes, which have afforded two clinically used antibiotics, telavancin and oritavancin, shown in Figure 127. Similarly to teicoplanin, telavancin, **156**, has a hydrophobic chain, decylaminoethyl, attached to a sugar, in this case, vancosamine. It also possesses a hydrophilic phosphonomethyl(aminoethyl) group on amino acid seven to improve solubility and reduce toxicity.²⁵⁵ It is active against *VanB* phenotypes. Oritavancin, **157**, is a derivative of chloroeremomycin, containing a *N*-alkyl-*p*-chlorophenylbenzyl group attached to 4-*epi*-vancosamine. This facilitates dimerization and membrane anchoring, which can enable binding to either D-Ala-D-Ala or D-Ala-D-Lac, giving activity against *VanA*, *B*, and *C* resistance phenotypes.²⁵⁶

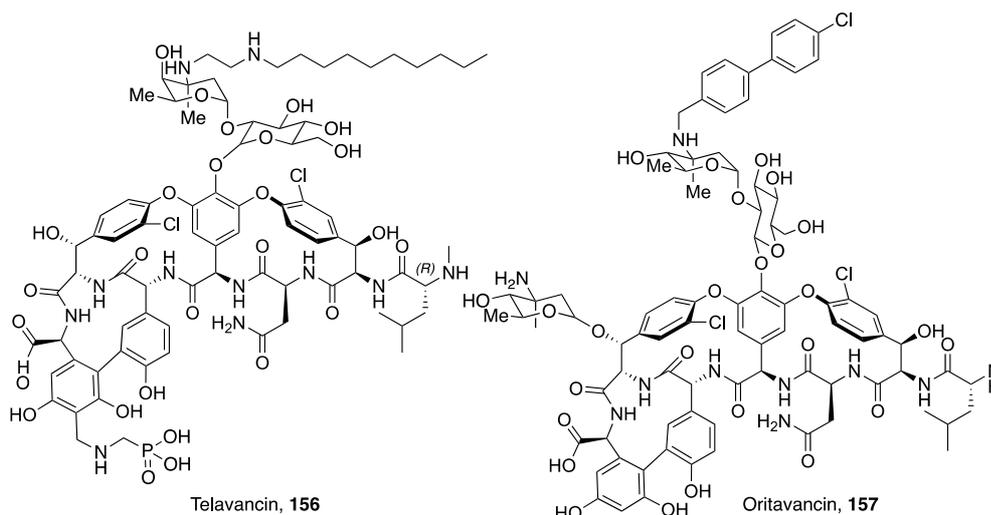


Figure 127 - Semisynthetic vancomycin derivatives, telavancin, 156, and oritavancin, 157. Both contain hydrophobic group attached to the amino-sugar, and polar groups for solubility appended to the scaffold.

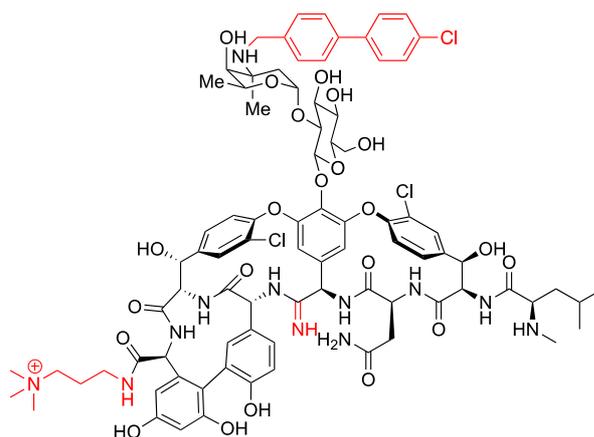
The chlorine atoms of residues two and six have been used as chemical handles for Suzuki-Miyura couplings. These showed that substitutions at less hindered amino acid two gave improved activity against vancomycin-resistant *Staphylococcus aureus* (VRSA) or VRE depending on the size of the substituent, while disubstituted products lost activity against VRSA, VRE and vancomycin-sensitive strains.²⁵⁷ In an alternate approach, the aryl chloride of residue two was converted to a boronic acid, facilitating a range of transformations, but a range of electron withdrawing or donating substituents failed to significantly affect activity. It was shown that a permethylated vancomycin aglycone was highly active against *VanB* resistant *Enterococcus faecalis*, approaching the activity of vancomycin against sensitive strains.²⁵⁸

N-bromosuccinimide has been used in conjunction with differing catalysts to achieve site-selective bromination of vancomycin, with catalyst-dependent selectivity.²⁵⁹ Methylation of the *N*-terminal leucine had little impact on activity,²⁶⁰ but modification of the carboxylic terminus could improve activity. Carboxamides of eremomycin proved more active against *S. epidermis*, while the coupling of a quaternary amines to vancomycin improved activity against VRE.^{261,262}

4.4.3 Total synthesis

The complexity of the vancomycin scaffold presents a formidable synthetic challenge due to the rigid, cross-linked backbone and atropisomerism of the aryl rings, and as such it has been a target for a number of synthetic groups. Three total syntheses of the aglycon have been reported, first in 1998 by the Evans group in 40 steps, followed shortly by the Nicolaou group.^{263,264} These were subsequently followed by the Boger group in 1999.²⁶⁵ While total synthesis often promises to allow access to novel analogues, this was only the case for one of these efforts.

The change from D-Ala-D-Ala to D-Ala-D-Lac results in a 1000-fold drop in binding affinity with vancomycin. Through the synthesis of D-Ala-D-Lac analogues, it was determined that the loss of a hydrogen bond is only responsible for a 10-fold loss of binding affinity, with the 100-fold loss due to the lone pair-lone pair repulsion.²⁶⁶ This prompted the synthesis of a vancomycin aglycone analogue lacking a carbonyl in residue 4, removing the H-bond and lone pair repulsion. This worsened binding against the natural substrate, but improved binding to the resistant D-Ala-D-Lac and imparted activity against the *VanA* phenotype.²⁶⁷ Introducing an amidine to this position allowed the aglycone to bind D-Ala-D-Ala almost as well as vancomycin, and improved D-Ala-D-Lac binding 600-fold.²⁶⁸ Proceeding to the amidine *via* a thioamide allowed the divergent synthesis of a range of amidine analogues, allowing clarification of the protonation state and hydrogen bonding character of the amidine.¹⁸¹ The protonated amidine prevents the lone pair repulsion, and acts as a weak reverse hydrogen-bond donor. Combining the amidine, only accessible through total synthesis, with a quaternary ammonium salt, and a hydrophobic (4-chlorobiphenyl)methyl addition to vancosamine, displayed in Figure 128, can improve the activity 6000-fold against VRE strains compared to vancomycin. The three independent, synergistic mechanisms of action also reduce the likelihood of resistance developing, with the MIC increasing 4-fold after 50 passages. The number of mechanisms is inversely correlated with the development of resistance, with **158** proving markedly more robust than current standards such as daptomycin and linezolid.



Boger's binding pocket modified vancomycin , 158

Figure 128 - Synthetic vancomycin analogue by Boger *et al.* incorporating an amidine in residue four, a C-terminal quaternary ammonium salt, and a disaccharide functionalised with (4-chlorobiphenyl)methyl group.

4.5 Biosynthesis of vancomycin

The understanding of the biosynthesis of vancomycin has been hindered by the lack of genomic information available until the late 1990s, and information of the antibiotic structure itself. Work on the related glycopeptides chloroeremomycin and balhimycin, shown in Figure 129, which differ only in glycosylation patterns, as well as vancomycin itself, has unveiled most aspects of the biosynthetic pathway to this important antibiotic. The 67 kb gene cluster responsible for the production of vancomycin has been recently reported and is shown in Figure 130 (NCBI accession number: HE589771.1).

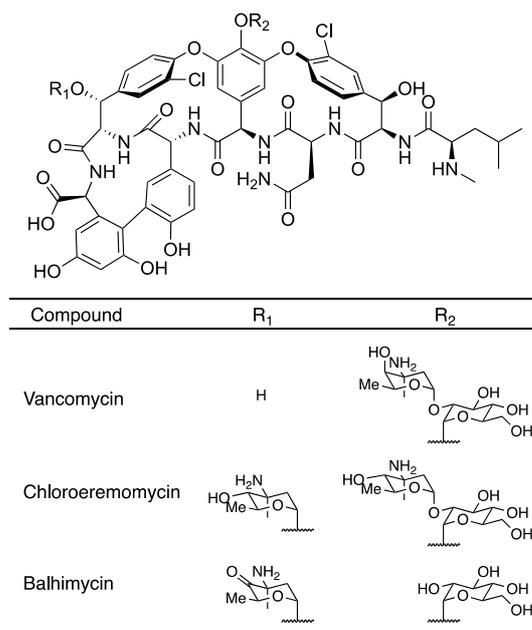


Figure 129 - The differing glycosylation patterns on the heptapeptide scaffold of the related compounds, vancomycin, chloroeremomycin and balhimycin, which have been used to elucidate the biosynthetic pathway.

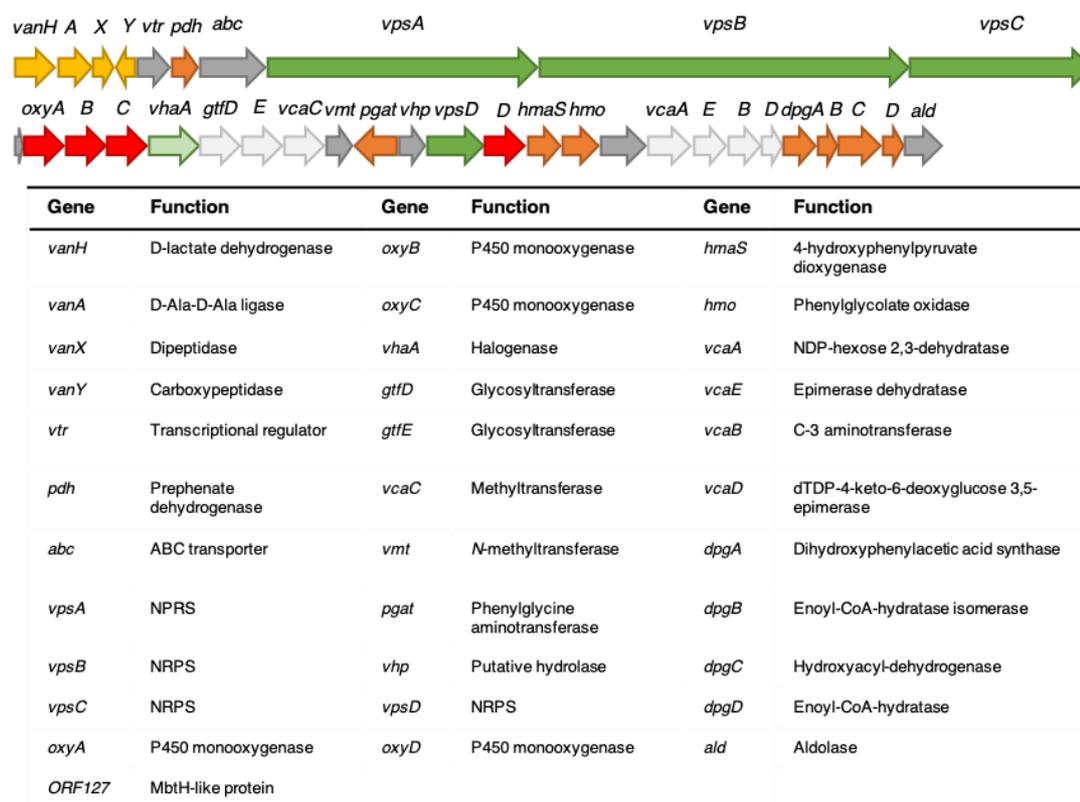


Figure 130 - The 67 kb gene cluster coding for the vancomycin biosynthetic enzymes and their assigned functions.

The balhimycin BGC and the chloroeremomycin BGC contain homologous genes for precursor biosynthesis, NRPSs, halogenase and oxidative enzymes. The clusters differ in the glycosyltransferases which account for the alternate glycosylation patterns observed.

4.5.1 NRPS assembly

While chemical degradation revealed the incorporation of a number of amino acids, including non-proteinogenic amino acids such as 4-hydroxyphenylglycine, the mode of assembly remained unknown until the sequencing of the chloroeremomycin gene cluster in 1998. Prior to this, the assembly of the 3,5-dihydroxyphenylglycine was found to be derived from acetate using [1,2-¹³C₂]acetate, and 4-hydroxyphenyl glycine and Cl-β-OH-Tyr was found to be produced from tyrosine from feeding with (2R, 3R)-[2-²H,3-²H₁,3-¹³C]tyrosine.²⁶⁹ This also allowed the identification of L-Tyr as the

precursor to Cl- β -OH-Tyr, and that β -hydroxylation occurs with retention of configuration. Sequencing of the chloroeremomycin cluster led to the identification of three NRPSs *cepA-C*, containing seven modules. Despite the four D-amino acids, only three epimerase domains were found.²⁷⁰ Potential oxidative enzymes, haloperoxidases, and glycosyl transferases were also identified. The *oxyA-C* genes, halogenase *bhaA*, and glycosyl transferases *bgtA-C* of the balhimycin cluster were previously identified from a cosmid library.²⁷¹ The NRPS genes of the balhimycin cluster, sharing an identical heptapeptide scaffold, was identified in 2002.⁸⁴ This identified an additional NRPS not bearing any homology to those in the *cep* cluster. *BpsD* is separate to the other NRPS, and inactivation of the gene prevents balhimycin production. *BpsD* was predicted to activate L-tyrosine and could be complemented with β -OH-Tyr.

4.5.2 Non-proteinogenic amino acid biosynthesis

In addition to L-leucine and L-asparagine, vancomycin also contains three non-proteinogenic amino acid, L-3,5-dihydroxyphenylglycine (Dhpg), L-4-hydroxyphenylglycine (Hpg), and both L- and D- β -hydroxytyrosine (β HT). These must be generated from available building blocks before incorporation into the heptapeptide scaffold.

As was shown in labelling studies, Dhpg can be derived from acetate units. This was suggested to proceed *via* a polyketide mechanism; this was confirmed by complementary *in vivo* and *in vitro* experiments.^{272–274} Inactivation of *dpgA* of the balhimycin cluster prevented production, which could be complemented by 3,5-diphenylacetic acid. This gene was translationally coupled to *dpgB,C* and *D*, which when expressed in *S. lividans* produced 3,5-dihydroxyphenylglyoxalic acid. The final step to produce Dhpg would be transamination; inactivation of the phenylglycine aminotransferase *pgat* eliminated production and led to the accumulation of 3,5-dihydroxyphenylglyoxalic acid, **161**. It was also shown that this transaminase was necessary for the biosynthesis of Hpg, and its action confirmed *in vitro*.²⁷⁵ The complete pathway to Dhpg is shown in Figure 131.

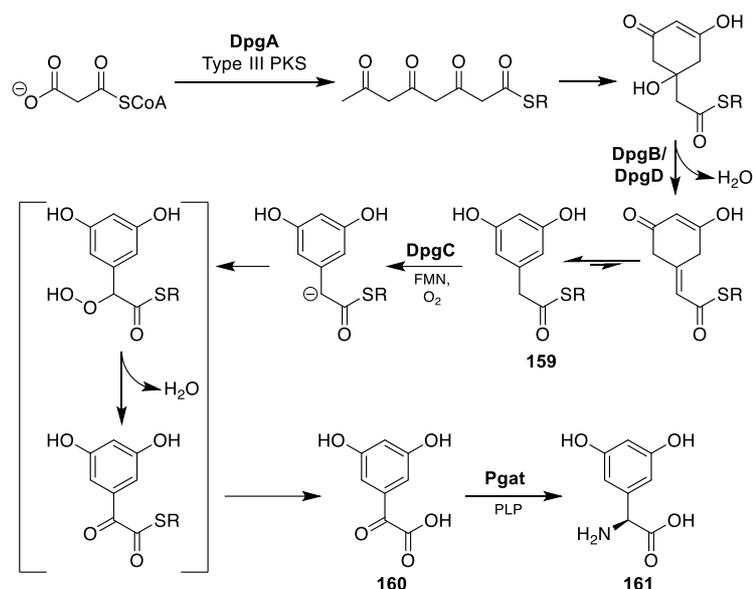


Figure 131 - Biosynthesis of Dhpg starting from malonyl-CoA, converted to 3,5-dihydroxyphenylacetyl-CoA by DpgA and DpgB/D. This is oxidised by DpgC before transamination by Pgat.

In vitro characterisation of the homologous proteins from the *cep* cluster found DpgA to be a type III polyketide synthase, which utilises only malonyl-CoA as a substrate, with DpgB and D acting as dehydrogenases to produce 3,5-dihydroxyphenylacetyl-CoA, **159**. DpgC, a cofactorless monooxygenase, then converts this to 3,5-dihydroxyphenylglyoxylate, **160**.

Hpg biosynthesis, shown in Figure 132, was investigated *in vitro* using enzymes from the chloroeremomycin pathway and the homologous vancomycin enzymes.^{276,277} This showed the biosynthesis can begin from either tyrosine, as identified in labelling experiments, or prephenate. It was assumed that *p*-hydroxyphenylpyruvate, **162**, could be generated from prephenate by a prephenate dehydrogenase, although the protein could not be produced in a soluble form. This can then be converted to *L-p*-hydroxymandelate, **163**, by the non-heme iron dioxygenase HmaS. Hmo, a hydroxymandelate oxidase, then acts to produce *p*-hydroxybenzoylformate, **164**, which is then transaminated to produce Hpg, **165**.

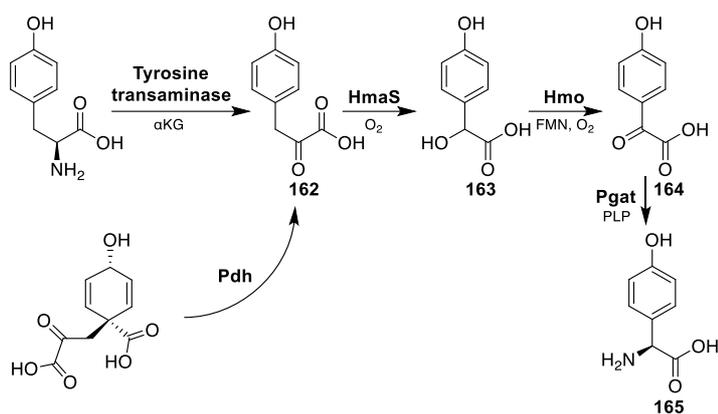


Figure 132 - Biosynthesis of Hpg beginning from either prephenate or tyrosine which are converted to *p*-hydroxyphenylpyruvate, followed by oxidation to *L-p*-hydroxymandelate by HmaS, then oxidation to *p*-hydroxybenzoylformate and transamination to Hpg by Pgat.

The final non-proteinogenic amino acid β HT, **166**, is produced by a NRPS and cytochrome P450. The additional NRPS identified from the balhimycin cluster, BpsD, was shown to activate L-tyrosine. Once loaded, this is hydroxylated by the P450 OxyD, in a manner analogous to the production of β HT in novobiocin biosynthesis.²⁷⁸ This NRPS-P450 complex was later structurally characterised.²⁷⁹ The amino acid is then hydrolytically released from the NRPS by the thioesterase Bhp, as shown in Figure 133, which is selective for L- β HT over L-Tyr.²⁸⁰

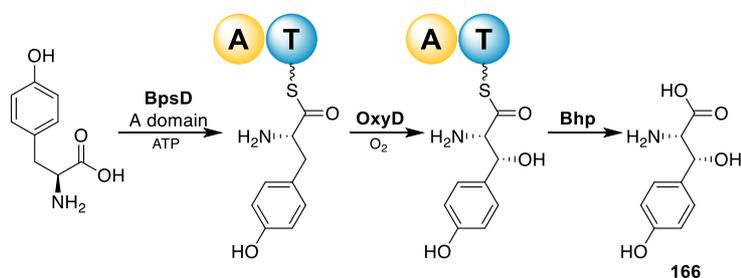


Figure 133 - Biosynthesis of β -hydroxytyrosine from L-tyrosine by the NRPS BpsD and P450 OxyD, before hydrolytic release by Bhp.

4.5.3 Tailoring of the heptapeptide

In glycopeptide NRPSs, only three epimerisation modules are encoded, but these cannot explain the presence of D-leucine in the heptapeptide. While this was noted in the identification of the NRPSs of chloroeremomycin, it was not until the

expression of a CepA A-T construct that this could be investigated.²⁸¹ This showed the first A domain had a six-fold preference for L-Leu over D-Leu, and it was suggested a racemase outside the cluster may be responsible. The genome sequencing of vancomycin producers identified eleven potential racemases.²⁸² Of these, one had high homology to a *Vibrio cholerae* PLP-dependent amino acid racemase found to be sufficient for D-Leu biosynthesis, but its action in vancomycin biosynthesis has not been experimentally confirmed.²⁸³

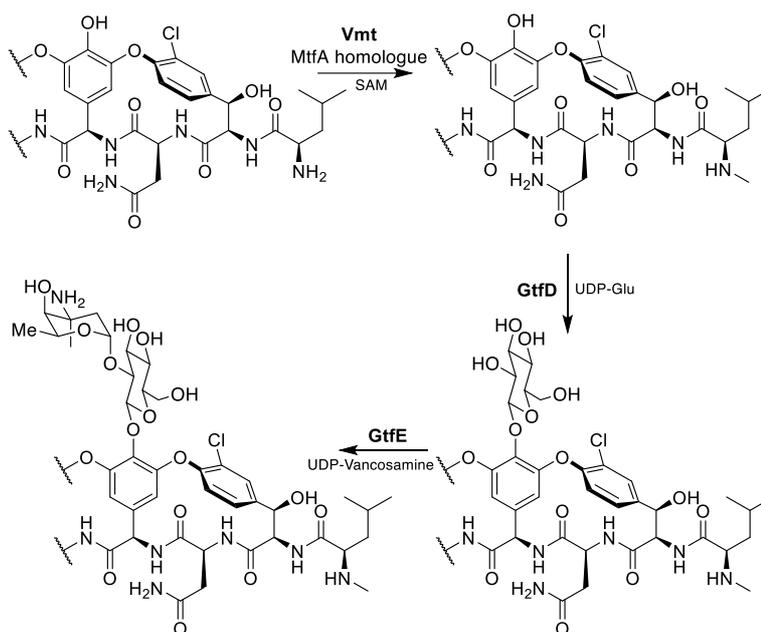


Figure 134 - Methylation and glycosylation of the vancomycin aglycone after oxidative cross-linking and release from the NRPS.

The timing of the *N*-methylation of the terminal leucine was established through the incubation of various possible substrates with MtfA. This found the preferred substrate was the cross-linked aglycone, indicating it occurs after the action of the Oxy enzymes and release from the NRPS.

The action of the glycosyl transferases was one of the first steps to be clarified. It was shown that GtfE transfers a glucose from an NDP-sugar to the vancomycin aglycone, as in Figure 134.²⁸⁴ The role of GtfD in the addition of vancosamine was confirmed with the demonstration that 4-*epi*-vancosamine could be transferred to glucosylated vancomycin.²⁴⁹

4.5.4 Oxidative cross-linking

The order of the oxidative cross-linking steps had been established by inactivation of the *oxyA-C* genes in *A. balhimycina*, showing the sequential cross-linking by OxyB, then OxyA, and finally OxyC, as in Figure 135. However, the exact substrates of the P450s were unknown, and are not yet fully clarified.¹¹³ Interrogation of the pathway was hampered *in vitro* by the inability to proceed past the initial cyclisation by OxyB. It was only with an understanding of the role of the X-domain that the cross-linking could be understood. The X-domain is common to all glycopeptide biosyntheses but is not found in other NRPS. While similar to a C-domain, it has changes which block the substrate binding channel, removing catalytic activity. However, it was found to bind the Oxy enzymes, recruiting the P450s to the NRPS for cross-linking.⁶⁴ While previously OxyB from the *van* cluster had been shown to act on hexa- and heptapeptide substrates, it was found that adding the X-domain increased activity.²⁸⁵ In the homologous P450, CepF of chloroeremomycin biosynthesis, cross-linking barely proceeded without the X-domain.

This breakthrough also allowed the activity of OxyA to be reconstituted *in vitro*, followed by the reconstitution of OxyC allowing the chemoenzymatic synthesis of the vancomycin aglycone.^{79,85}

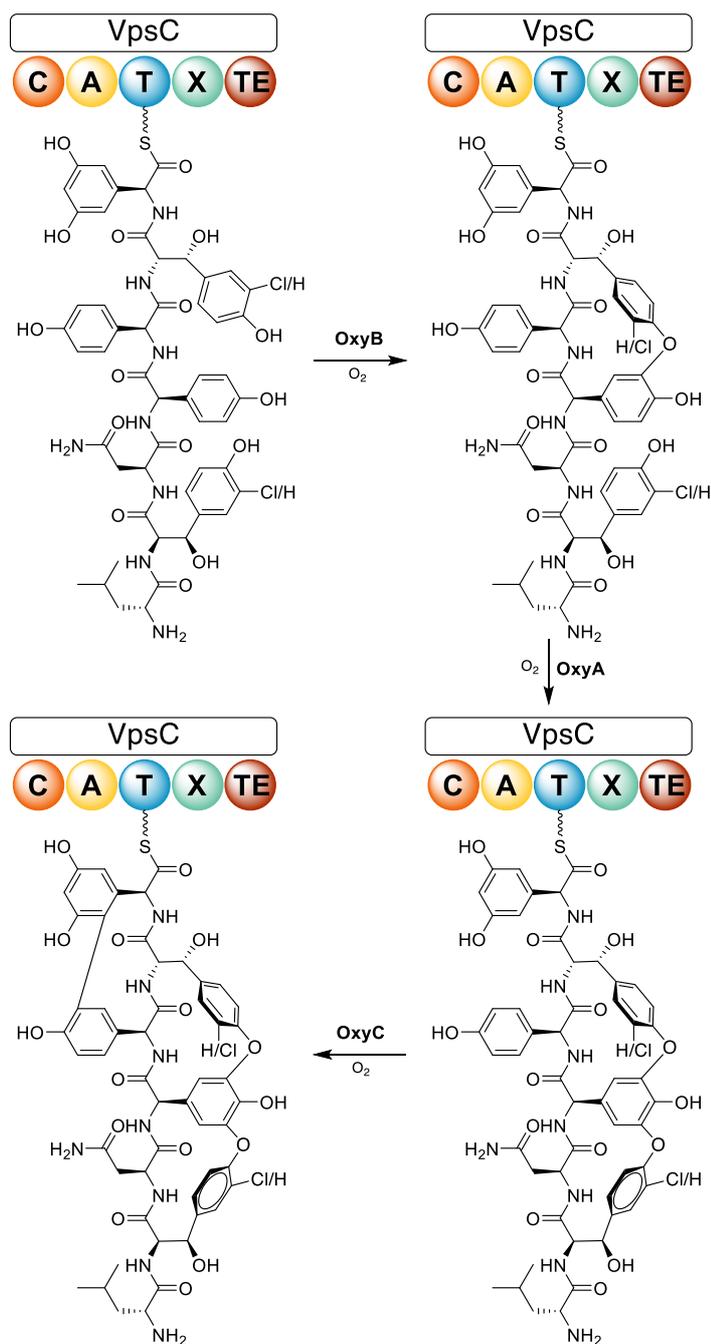


Figure 135 - Sequential cross-linking of the NRPS-bound heptapeptide by P450 monooxygenases OxyB, OxyA, and OxyC.

However, the timing of chlorination by the halogenase VhaA remains unclarified. This has been complicated by potentially conflicting evidence from a variety of experiments. When trying to understand the substrates of the Oxy enzymes, it was found chlorination on residue two reduced the rate of OxyB cross-linking of a hexapeptide, and chlorination on residue six abolished it completely.¹⁰⁴ However, this

occurred before the role of the X-domain was clarified. In the related teicoplanin biosynthesis, dichlorinated heptapeptides were found to be better substrates for OxyA, with a slight improvement also noted for OxyB, both in the presence of the X-domain.²⁸⁶ This was proposed to be due to better substrate orienting due to the bulky chlorine substituent, and would agree with the isolation of linear dichlorinated peptides from mutant *A. balhimycina* strains.²⁸⁷

4.5.5 Halogenation

Halogenase inactivation leads to the production of dechlorobalhimycin, and balhimycin production could not be restored with Cl- β -OH-Tyr. This would rule out the action of the halogenase on free amino acids before A-domain activation.^{288,289} This would suggest halogenation after activation of the amino acid or on the growing peptide chain as likely scenarios. *In vitro* halogenation showed dichlorination of a hexapeptide when β HT was incorporated, but not with tyrosine, suggesting this may be needed for recognition. Meanwhile, no halogenation of an unnatural *N*-methyl-Leu- β HT T-domain bound dipeptide was observed.²⁹⁰ Characterised FADH₂-dependent halogenases associated with NRPs act on enzyme-bound amino acids, which was found to be the case in teicoplanin biosynthesis.²⁹¹ However since the biosynthesis of teicoplanin is known to incorporate β HT *via* a different mechanism, this finding may not be applicable to vancomycin biosynthesis. The insertion of the module seven TE domain into module three of BpsA led to the offloading of chlorinated dipeptides, shown in Figure 136. The halogenase BhaA was found to be inactive on T-domain bound L-Tyr, possibly due the lack of β -hydroxyl group, or due to the dipeptide being the true substrate.

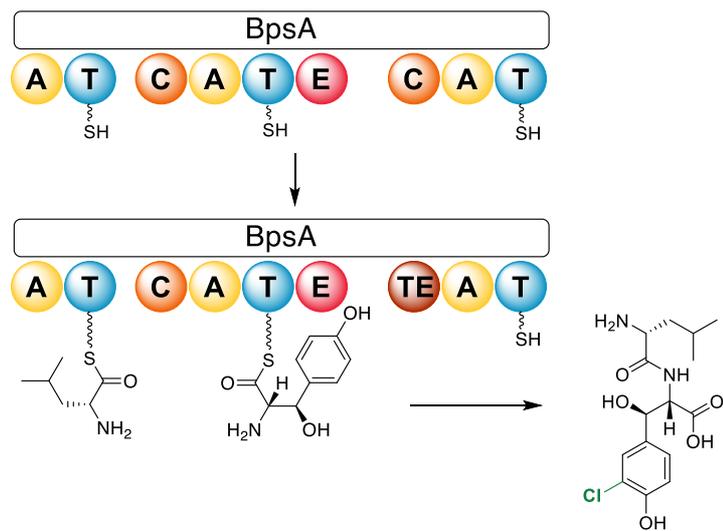


Figure 136 - The replacement of the C domain of module 3 with the TE domain of module 7 causes the offloading of dipeptides. These dipeptides were found to have been halogenated by VhaA.

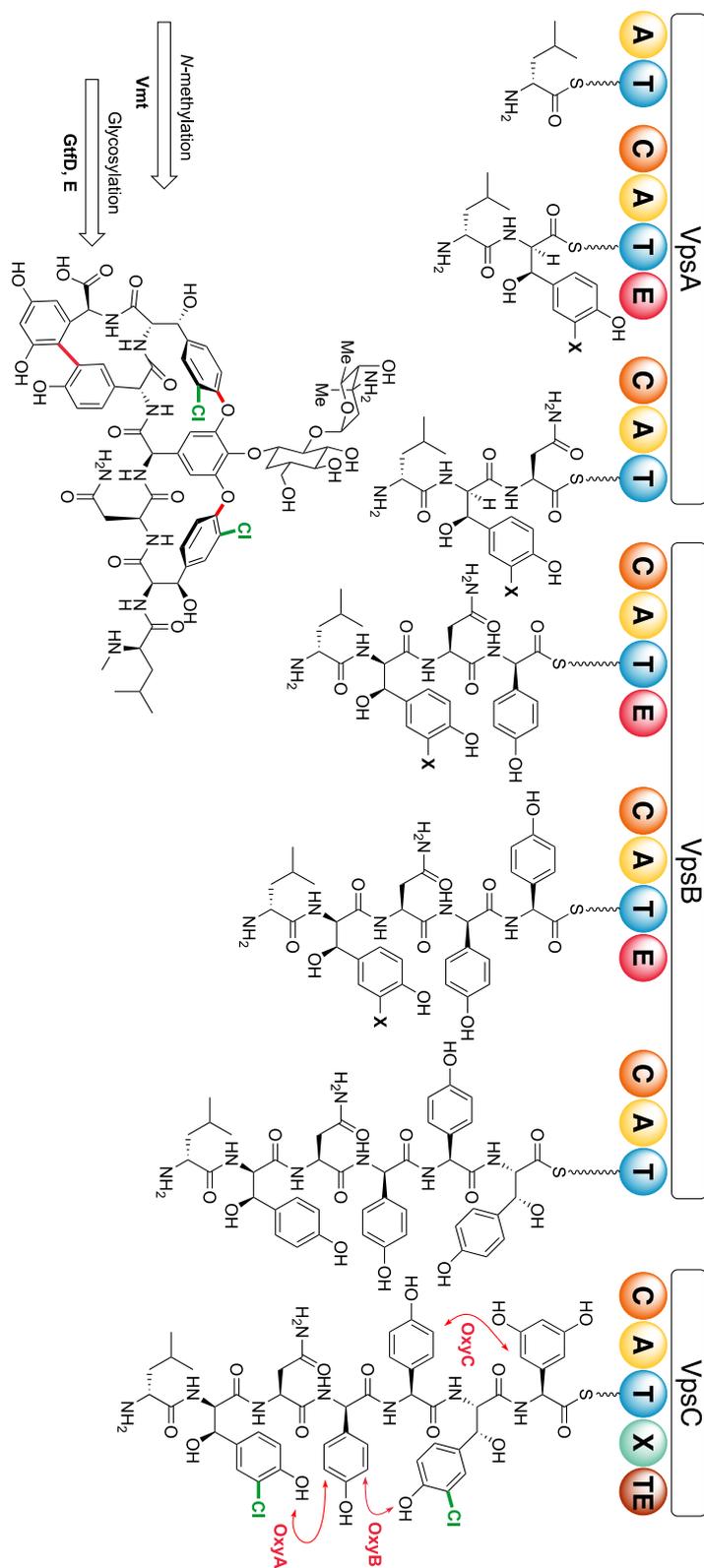


Figure 137 – Proposed biosynthesis of vancomycin, where X = H or Cl. The heptapeptide is assembled by the 7 module NRPS, followed by oxidative cross-linking, before hydrolytic release. The *N*-terminal leucine is then methylated, followed by glycosylation with glucose and vancosamine.

4.6 Chemical probes to investigate vancomycin biosynthesis

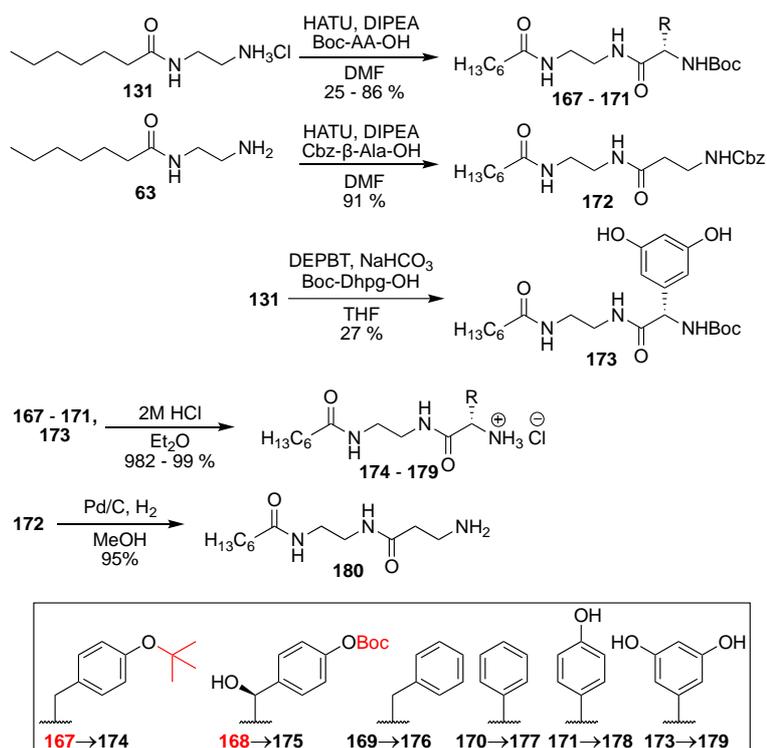
While crucially important compounds, the glycopeptide antibiotics have not been previously investigated using chemical probes of any sort. There are a number of unresolved questions in the biosynthesis of vancomycin, which chemical probes offer a valuable tool to study.

The enzyme responsible for the epimerisation of the first amino acid of the pathway from L-Leu to D-Leu is still unknown. The timing of the action of halogenase VhaA is also unknown, as shown in Figure 137. The introduction of the two chlorine atoms occurs while the peptide chain is bound to the NRPS, but the substrates are not known, or how the chlorination affects the cross-linking Oxy enzymes or *vice versa*.

4.6.1 Synthesis of chemical probes

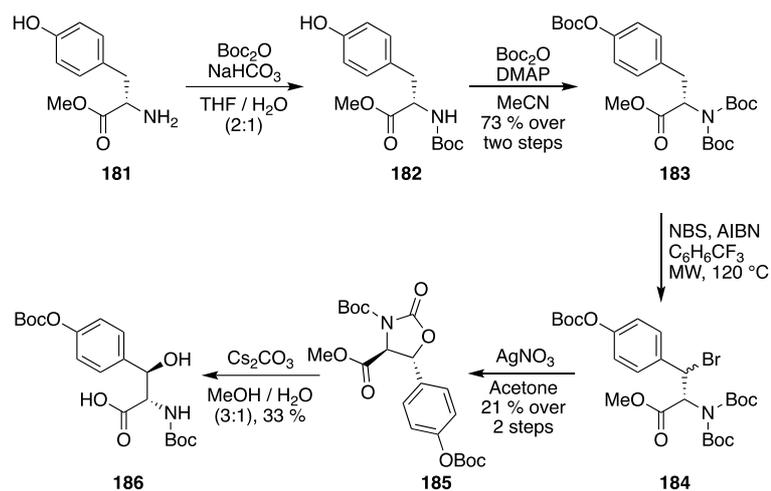
To investigate the yet unclarified steps in vancomycin biosynthesis, particularly the timing of halogenation, a number of chemical probes were devised and synthesized as part of my PhD work, in collaboration with the Cryle group at Monash University.

The probes were largely designed to contain aromatic amino acid side chains on the basis of aromatic amino acids naturally incorporated in vancomycin (β -hydroxytyrosine, 4-hydroxyphenylglycine, and 3,5-dihydroxyphenylglycine), and also to be based on non-cognate substrates (tyrosine, phenylglycine, and phenylalanine) to investigate the flexibility of different C domains at the acceptor site during glycopeptide assembly. Glycine and β -alanine probes were also tested based on their proven utility in probing echinomycin and colibactin biosyntheses, as detailed in the previous chapters.



Scheme 8 - Synthesis of chemical probes for the vancomycin NRPSs based on glycine, β -alanine, phenylglycine, phenylalanine, tyrosine, β -hydroxytyrosine, 4-hydroxyphenylglycine, and 3,5-dihydroxyphenylglycine.

The synthesis of chemical probes utilised for vancomycin biosynthetic investigation is summarised in Scheme 8. The common amine intermediate **131** was synthesized as described as in previous chapters, followed by its coupling to protected amino acids using HATU, followed by Boc deprotection using HCl or Cbz deprotection by hydrogenolysis. In the case of Tyr probe **167** and β HT probe **168**, the phenol protecting groups, shown in red in scheme 8, could also be removed under acidic conditions to afford **174** and **175**. In the case of the Dhpg (a gift from Cryle group), the amino acid is highly prone to epimerisation is basic conditions.²⁹² Therefore DEPBT was used as the coupling reagent, due the reduced epimerisation reported.²⁹³ While most amino acids were commercially available, β HT was prohibitively expensive and was synthesized using a modified procedure from Crich *et al.* as shown in Scheme 9.²⁹⁴

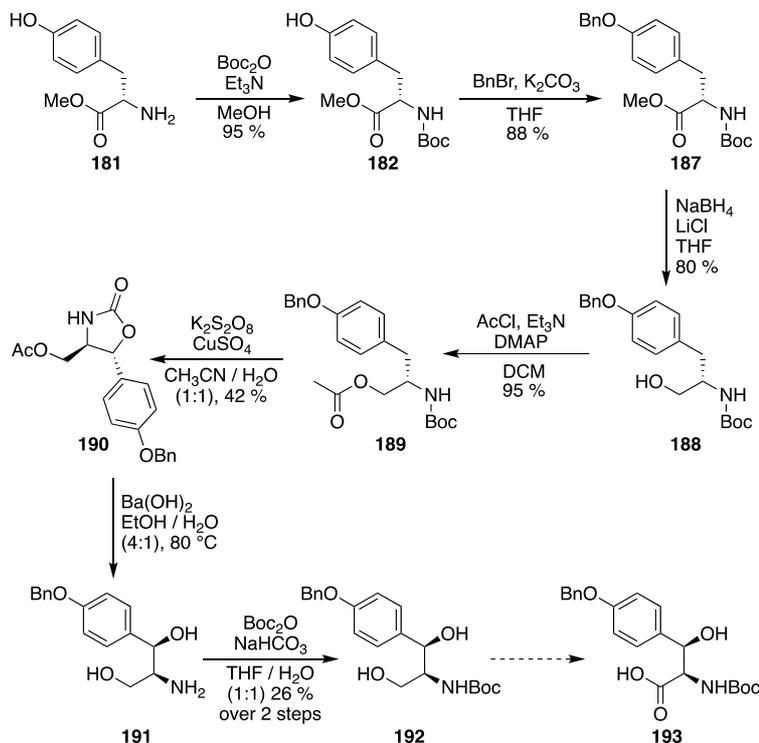


Scheme 9 - Synthesis of β -hydroxytyrosine from tyrosine methyl ester following an adapted procedure from Crich et al.²⁹⁴

(Boc)₂-Tyr(Boc)-OMe, **183**, was synthesised as reported. Given the ozone-depleting and toxic nature of the CCl₄ used as solvent in the reported procedure, alternative solvents were investigated for the radical bromination of intermediate **183**. Reactions conducted in cyclohexane consumed *N*-bromosuccinimide (NBS) as established by TLC but did not result in product formation, likely due to solvent bromination. Chloroform has also been suggested as a suitable replacement, but again no product formation was observed in this solvent.²⁹⁵ Using α,α,α -trifluorotoluene in the presence of azobisisobutyronitrile (AIBN) as a radical initiator gave a small amount of product, **184**, as detected by TLC in 1:4 EtOAc: petroleum ether.²⁹⁶ Microwave heating improved the consumption of starting material, but this did not seem to be due to thermal degradation of the Boc-protected starting material. Optimisation of the amount of NBS to 2 equivalents allowed the complete consumption of the starting material. The solution containing the intermediate bromo species **184** was then concentrated and mixed with AgNO₃ to abstract the bromine and form the oxazolidinone, **185**. This proceeds *via* a carbocation which is stabilised by the oxygen of the Boc carbamate group, and thus preferentially forms the *anti*-diastereomer. This was then hydrolysed using caesium carbonate to afford the Boc-protected amino acid, **186**, which could be coupled as in Scheme 8.

Given the poor yield obtained in the crucial radical bromination step in this route, alternative preparations were examined to allow the scalable synthesis of β HT.

A second route shown in Scheme 10, utilising a benzylic hydroxylation rather than a benzylic bromination, was attempted.²⁹⁷



Scheme 10 - Second synthetic route towards β-hydroxytyrosine proceeding through a benzylic hydroxylation based on the route developed by Shimamoto et al.²⁹⁷

Starting from L-tyrosine methyl ester, **181**, this could be protected at the amine functionality using Boc anhydride, followed by protection of the phenolic alcohol as a benzyl ether using benzyl bromide. The ester of **187** could then be reduced using LiBH₄, generated *in situ* from NaBH₄ and LiCl.²⁹⁸ The alcohol is then masked as an acetyl ester **189** using acetyl chloride. The benzylic hydroxylation can then be performed using K₂S₂O₈ and CuSO₄, to give the oxazolidinone **190**. This could then be hydrolysed using Ba(OH)₂, followed by Boc protection of the amino alcohol **191** without workup, to afford the protected amino alcohol **192**. While not yet completed, the selective oxidation of the primary alcohol to a carboxylic acid could be achieved using a PtO₂-mediated Heyns oxidation, or a sodium hypochlorite/ TEMPO oxidation.²⁹⁹ Once the carboxylic acid is formed, the probe could be coupled to common amine **131** and deprotected sequentially using hydrogenation and HCl to remove the benzyl ether and Boc protecting groups.

4.6.2 Interception and capture of NRPS intermediates in vancomycin bio-assembly

Feeding experiments with the newly prepared chemical probes were conducted by initially culturing *Amycolatopsis orientalis* in R5 media for 3 days at 30 °C with shaking; this seed culture was then used to inoculate R5 media or R5 agar. The agar contained the chemical probes at 2.0 mM concentration, while the probe was added to the liquid cultures in 0.5 mM aliquots beginning on the second day. The cultures were then incubated for 5 days at 30 °C with shaking. The liquid cultures were lyophilised, and then all cultures were extracted with MeOH, concentrated and redissolved in MeOH. Cultures lacking any chemical probes were used as controls. The organic extracts were initially analysed on a Bruker MaXis high resolution mass spectrometer, which detected putative intercepted intermediates. An Orbitrap Fusion high resolution mass spectrometer was used to subject the putative species to MS² fragmentation to support the proposed structures.

The ability to effectively intercept and offload NRP intermediates from the vancomycin NRPS varied between the tested probes. No differences were observed between extracts from liquid cultures and those grown on solid agar. A summary of the results of our experiments is provided in Table 4 and illustrated in Figure 138.

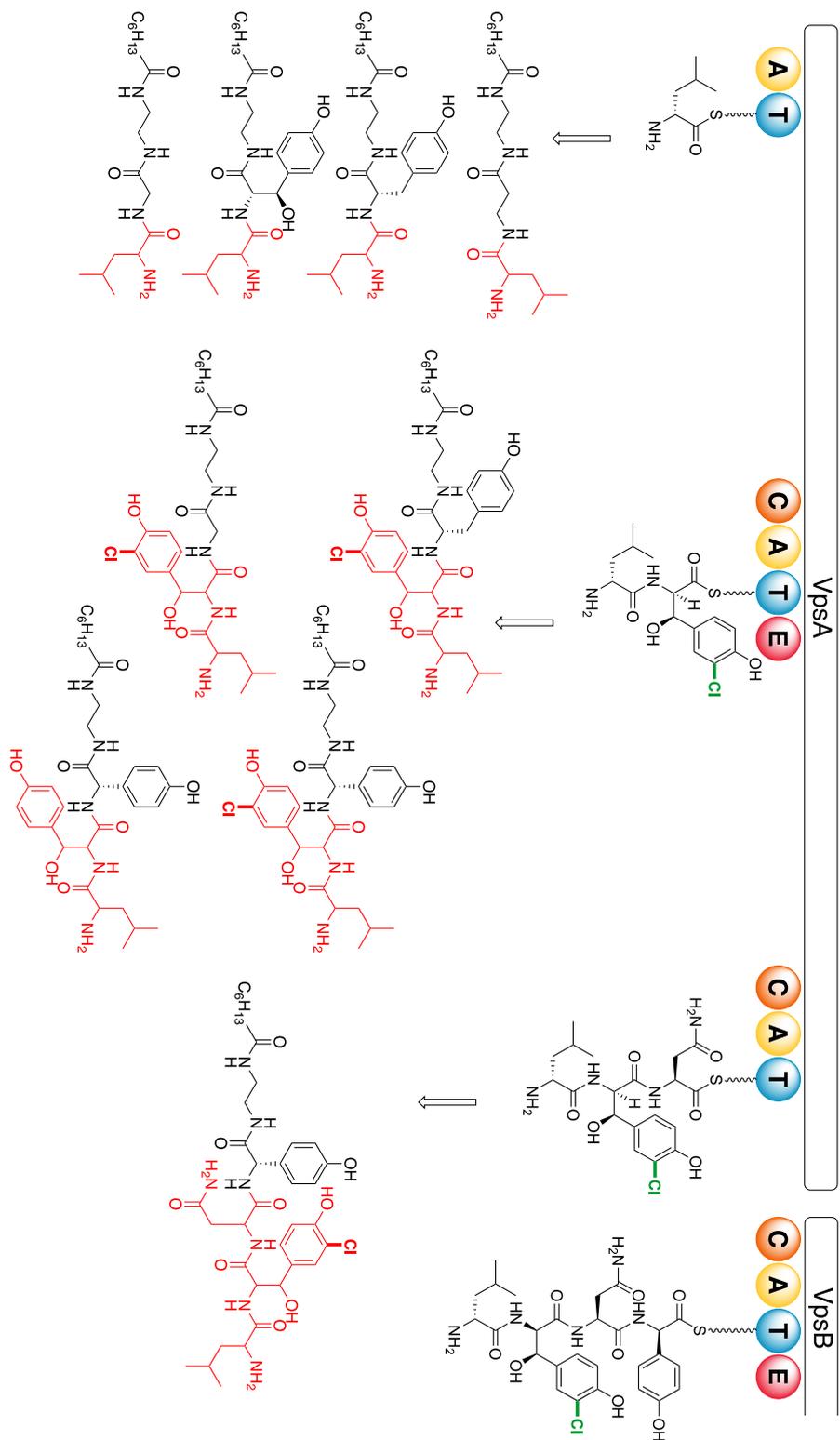
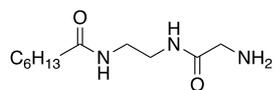


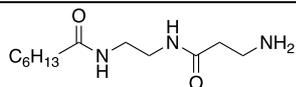
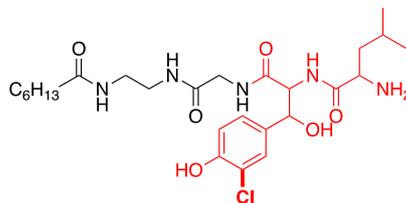
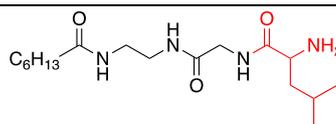
Figure 138 - Vancomycin biosynthetic intermediates intercepted by chemical probes in this work.

Probe structure

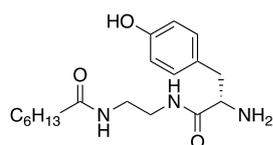
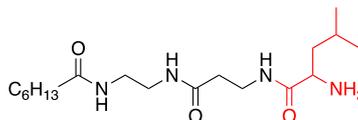
Captured intermediates



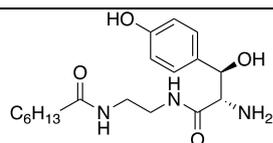
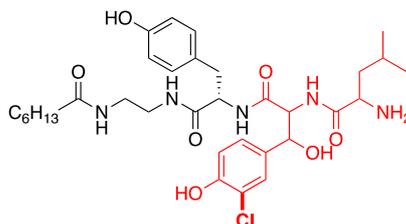
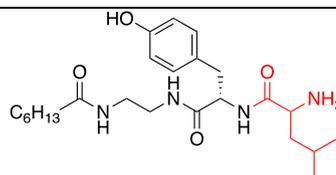
81



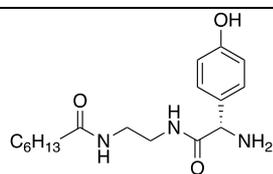
180



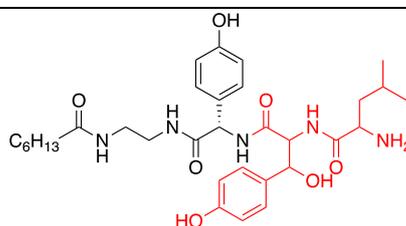
174



175



178



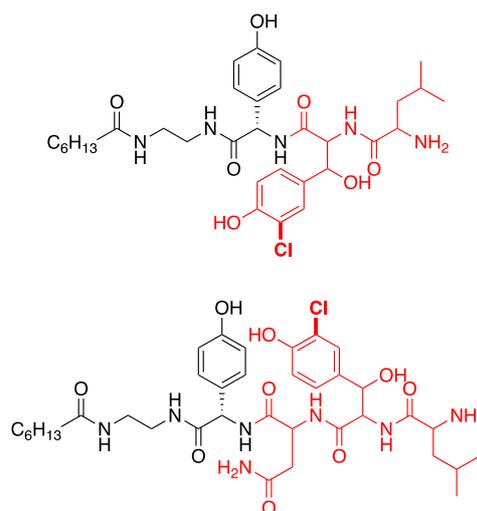


Table 4 - Vancomycin biosynthetic intermediates captured by chemical probes, and their relative intensities.

Both phenylalanine and phenylglycine substrates failed to capture any intermediates during the biosynthesis, suggesting that the phenol groups of the cognate amino acids may be important for substrate recognition.

Chemical probes based on β -alanine and β -hydroxytyrosine **180** and **175** were only able to offload leucine, the starter unit of the biosynthesis. The structure of the putative peptide formed, corresponding to the first condensation event and shown in Figure 139, was supported by MS² fragmentation. The interception of intermediates by the β -alanine substrate **180** has previously been proved in echinomycin and colibactin biosyntheses, with a number of different peptide species captured in both species. In the case of *A. orientalis*, no later stage intercepted intermediates could be observed.

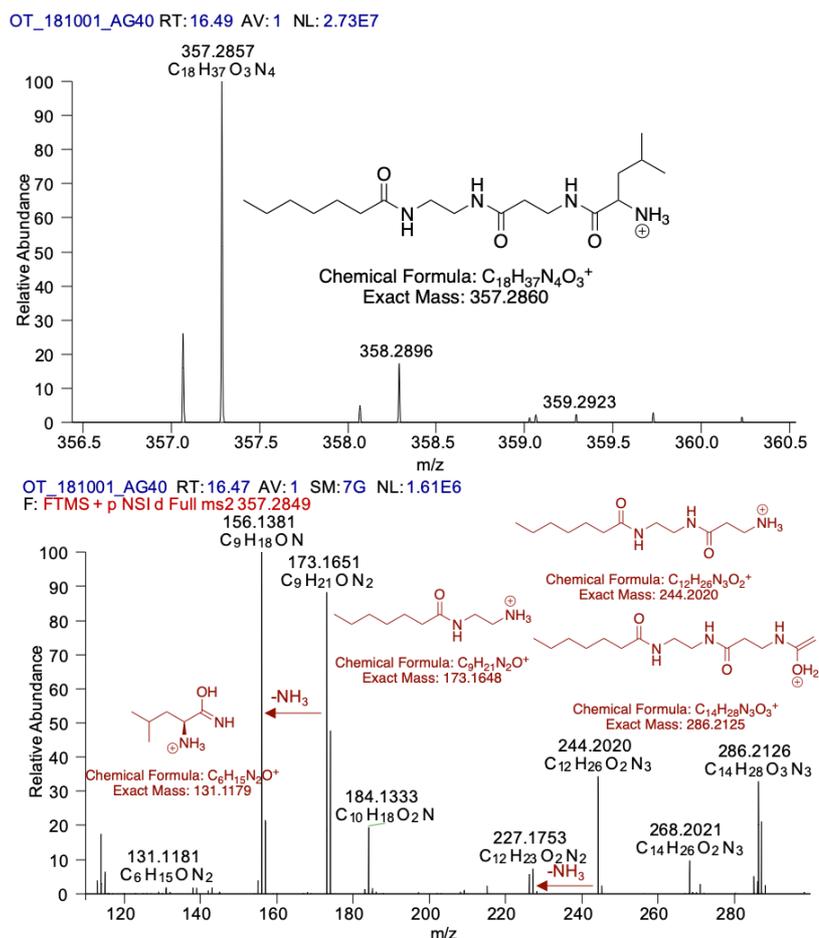


Figure 139 - Capture of leucine by the β -alanine based probe 180, with MS² fragmentation supporting the identity.

The β -hydroxytyrosine-based probe, as a cognate amino acid for vancomycin biosynthesis, was proposed to be well placed to intercept species in the condensation steps for which β HT is an acceptor amino acid. As β HT is the second residue in the vancomycin heptapeptide, the β HT-probe would be indeed expected to offload leucine, as seen in Figure 140. The MS² fragmentation of the peptide supports the identity of the structure proposed. No evidence could be found of any chlorinated species produced from halogenation of the probe, suggesting that all chlorinated intermediates intercepted by chemical probes are halogenated while bound to the NRPS before offloading.

The probe would also be expected to offload a later stage peptide intermediate corresponding to the condensation product of module 6. However, no such species could be detected in the culture extracts.

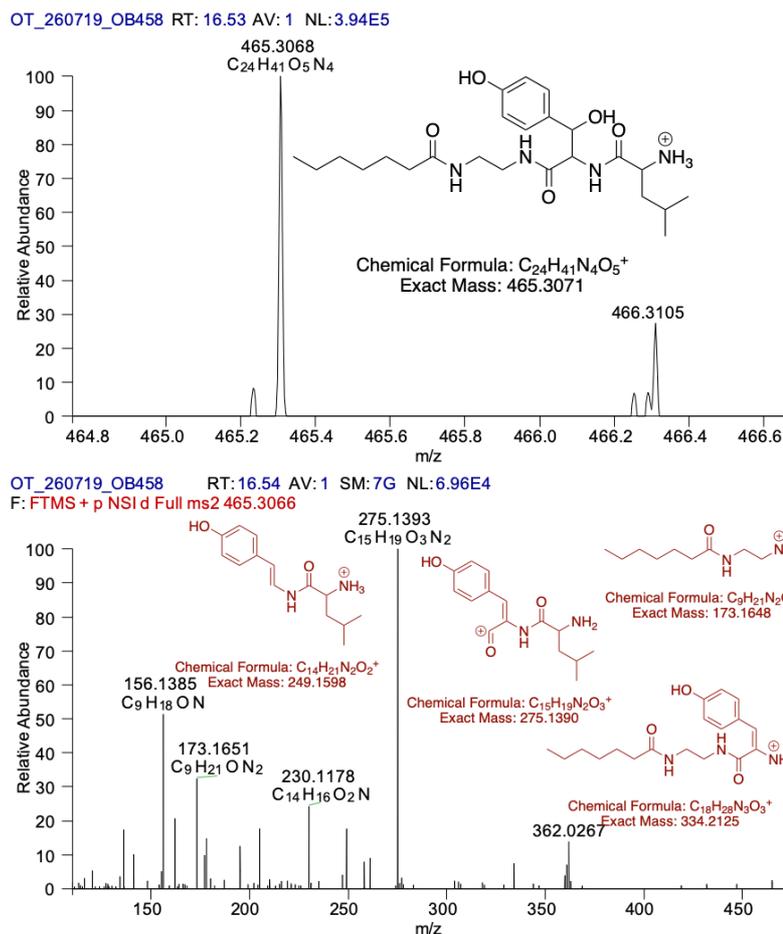


Figure 140 - Leucine (starter unit of vancomycin biosynthesis) offloaded by a chemical probe based on β -hydroxytyrosine 178; MS² fragmentation support the proposed structure, with diagnostic fragments showing the β HT-Leu dipeptide.

These preliminary results demonstrated the effectiveness of the chemical probes for the investigation of chain extension in vancomycin biosynthesis in the first instance, but also some limitations in their use. This will be further discussed in chapter 6.

The tyrosine-based probe, **174**, was also capable of offloading leucine. In addition, it also proved capable of offloading a putative enzyme-bound dipeptide species. The offloaded leucine species with high-resolution fragmentation to support its structure can be seen in Figure 141. The dipeptide species, shown in Figure 142, is the product of offloading from the second condensation domain and contains a chlorine atom. This is confirmed by both the high-resolution mass and the characteristic chlorine isotopic distribution. The proposed structure is also supported

by the structural fragments, which assign the amino acid order expected from the vancomycin structure.

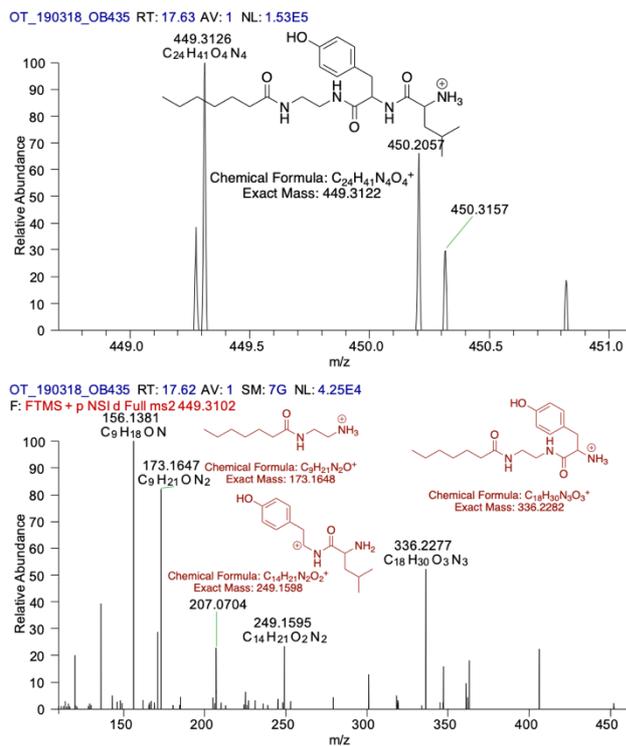


Figure 141 - Offloading of leucine starter unit by the tyrosine-based probe 174, with supporting MS² fragmentation.

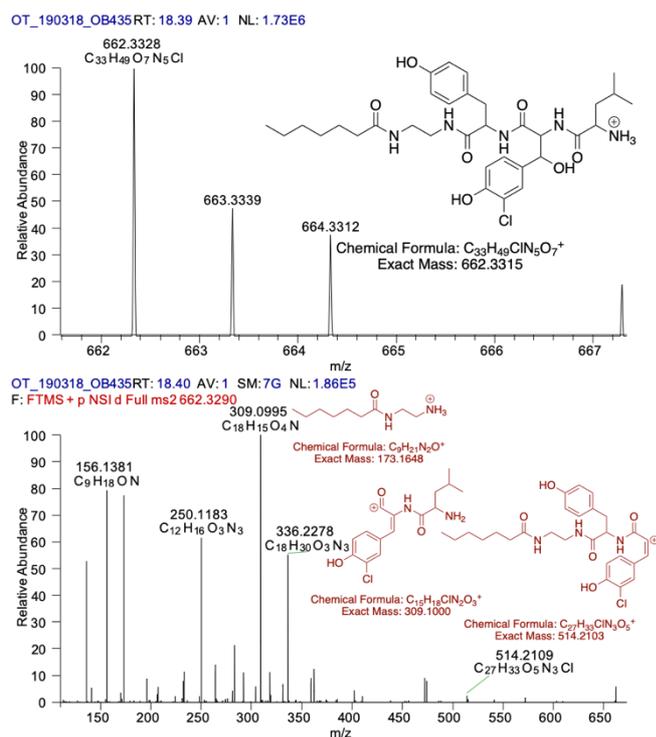


Figure 142 - Offloading of a chlorinated dipeptide intermediate from the vancomycin NRPS by the Tyr probe 174, with MS² fragments supporting the putative structure.

Similar intercepted intermediates were also be identified in extracts obtained by growing *A. orientalis* in the presence of the glycine-based probe **81**, as shown in Figures 143 and 144.

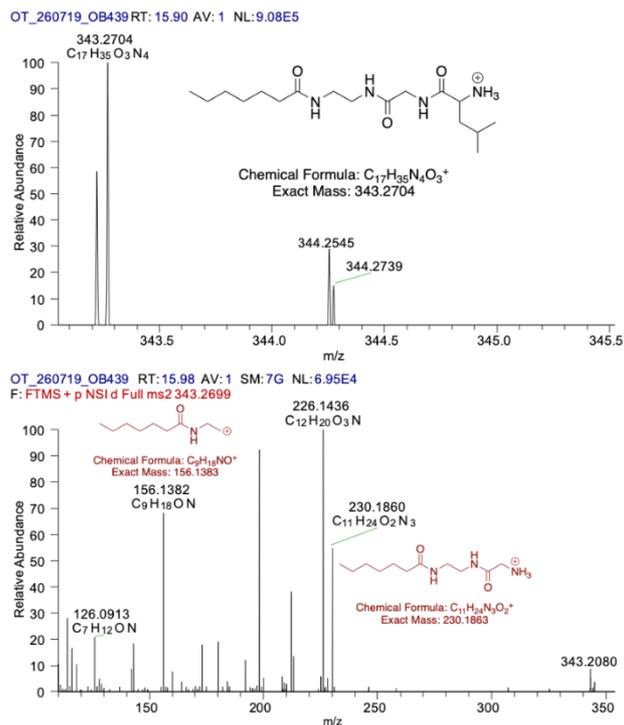


Figure 143 - Capture of the leucine starter unit from the vancomycin NRPS by the Gly probe 81, with MS² fragmentation.

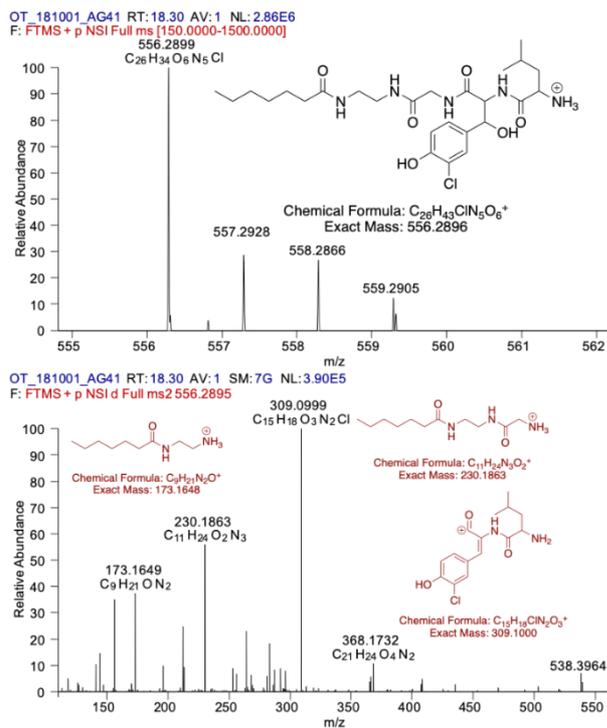


Figure 144 - Capture of a chlorinated dipeptide intermediate by the Gly probe 81, with MS² data to support the proposed structure.

This suggests that the first halogenation event in vancomycin biosynthesis occurs on enzyme-bound substrates, either on the individual T-domain bound β HT amino acid or on the NRPS-bound dipeptide itself before further peptide extension. In addition, tyrosine and glycine are not cognate substrates for tripeptide formation. This suggests that there is a degree of inherent substrate flexibility in this acceptor site. In a number of other glycopeptides, including ristomycin A and actinoidin A, which incorporate aromatic amino acids in the 3rd position. This will be discussed further in the chapter 6.

The Hpg-based probe, **178**, utilised next was the only substrate which was able to offload later stage intermediates. Offloaded leucine starter unit could be identified as usual. While a similar offloaded chlorinated dipeptide could also be detected (Figure 146), a dipeptide lacking chlorination was also detected in the organic extracts, as shown in Figure 145. All the proposed structures were supported by exact mass values and MS² fragmentation patterns. Indeed, analogous fragments can be identified from each dipeptide, at m/z 275.1391 (Figure 145) and m/z 309.1002 (Figure 146), which differ only in one chlorine atom. The Hpg-based probe was also capable of offloading a chlorinated tripeptide species (shown in Figure 147) from module 4 where Hpg is the expected acceptor amino acid. A characteristic chlorine isotope pattern as well as exact mass and fragments supported the natures of the proposed species, with the Hpg-Asn- β HT-Leu order of amino acids expected from the vancomycin structure itself and the (colinear) nature of the NRPS VpsA. Despite the same acceptor substrate specificity for module 5, no further advanced intermediates beyond module 4 were intercepted.

The Dhpg-based probe, **179**, as the cognate substrate for the sixth condensation domain, was envisaged to possibly offload a hexapeptide intermediate, allowing the clarification of the timing of the second halogenation event. However, no intermediates could be detected in the extracts from feedings with the Dhpg-based probe. This will need to be further investigated under different growth conditions for *A. orientalis*, but could not be conducted as part of this work due to time limitations.

The partial lack of chlorine in the dipeptide intermediate offloaded by the 4-hydroxyphenylglycine probe strongly suggest that chlorination by VhaA occurs at enzyme-bound dipeptide level, rather than at amino acid level.

The identification of chlorinated dipeptide intermediates from the biosynthesis of vancomycin is also in agreement with the previously reported chlorinated dipeptide obtained from genetic manipulation of BpsA (the first NRPS module of the balhimycin biosynthetic pathway).²⁹¹ In that case, fully chlorinated dipeptides were produced, for which specific information on the relative timing and the true substrates of chlorination by the halogenase VhaA could not be deduced. Our experiments suggest that, in vancomycin biosynthesis, the first chlorination occurs on an NRPS-bound dipeptide; however this assumption will require further corroboration by *in vitro* assays with recombinant proteins, or by further *in vivo* work with engineered vancomycin-producing strains. This will be discussed in detail in the chapter 6.

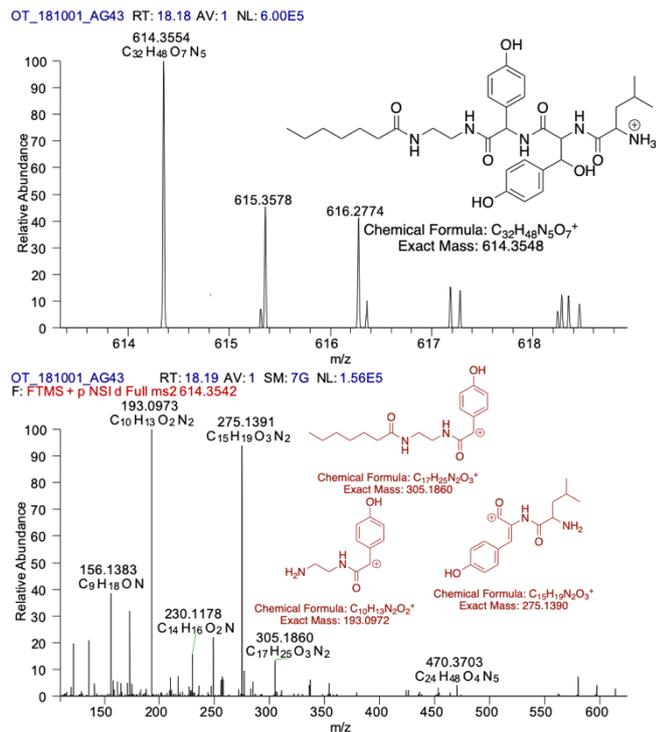


Figure 145 - Capture of a non-chlorinated dipeptide from the NRPS VpsA using a Hpg-based probe 178, with MS² fragments to support the assigned structure.

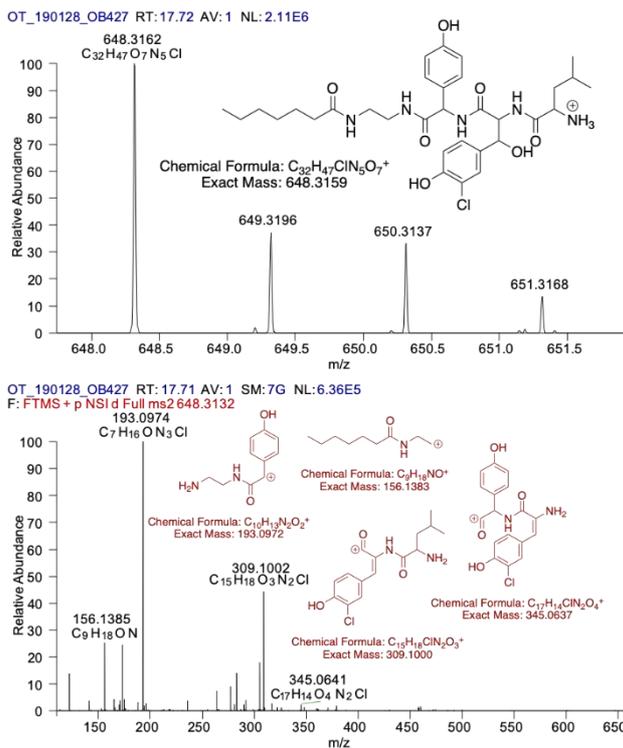


Figure 146 - Capture of a chlorinated dipeptide intermediate from the vancomycin NRP pathway using Hpg probe 178, with MS² fragments supporting the assigned structure.

The identification of the chlorinated intermediate in offloaded dipeptides from multiple chemical probes has also wider implications: it would suggest that chlorinated dipeptides are the cognate donor substrates for the condensation domain of module 3 in VpsB. This has implications for the rational redesign of glycopeptide biosynthetic pathways, particularly for the modules immediately downstream of halogenated intermediates. Again, this will be further discussed in the next chapter.

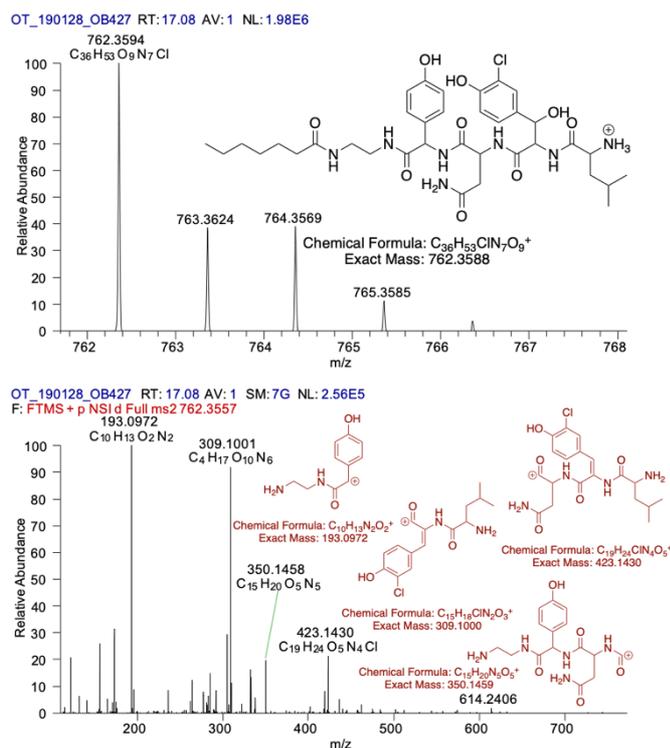
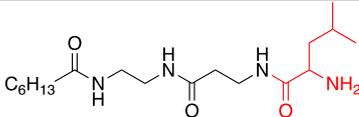
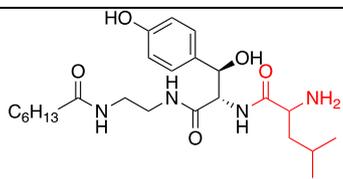
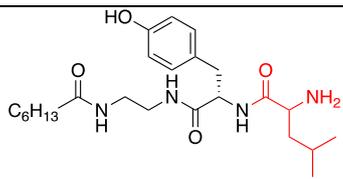
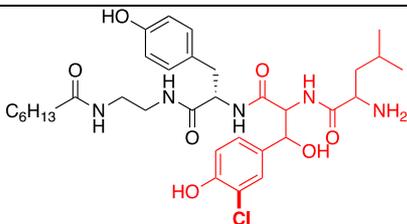
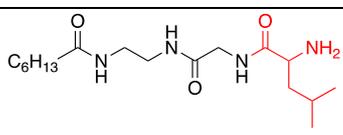
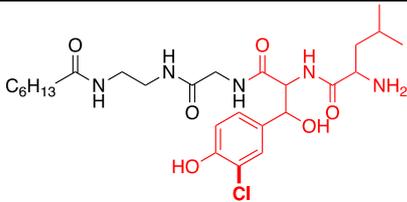


Figure 147 - Offloading of a chlorinated tripeptide intermediate from vancomycin biosynthesis using Hpg probe 178, with MS² fragments which support the structure.

So far it has not been possible to clarify the timing of the second chlorination event using chemical probes. Given the likely action of halogenase VhaA on NRPS bound intermediates and the apparent improvement on the Oxy enzymes rates of reaction upon chlorination, we expected to be able to intercept a linear hexapeptide intermediate, possibly dichlorinated.²⁸⁶ Possible reasons for the lack of evidence of this species will be further discussed in the chapter 6.

In terms of relative probe efficiency in offloading different intermediate species, the intensity of the species captured from offloading at the second condensation domain are consistently greater than the species captured from the first

condensation domain, shown in Table 5. This is clear in the case of the Tyr-, Gly-, and Hpg-probes which have offloaded both species. This may suggest that the substrate tolerance at position three in the heptapeptide backbone is greater than at other positions and this is reflected in less selective acceptor site in the second condensation domain, which will be discussed further in the chapter 6.

Probe	TIC	Intermediate	Normalised intermediate EIC
β -alanine 180	8.15 x 10^9		1.81 x 10^7
β -hydroxytyrosine 175	1.57 x 10^{10}		1.68 x 10^5
Tyrosine 174	7.13 x 10^9		1.57 x 10^5
			2.33 x 10^6
Glycine 81	1.54 x 10^{10}		3.52 x 10^5
			6.16 x 10^5

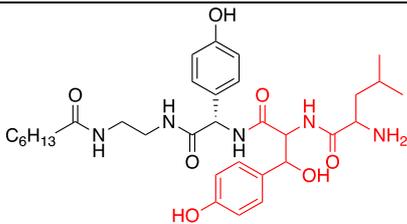
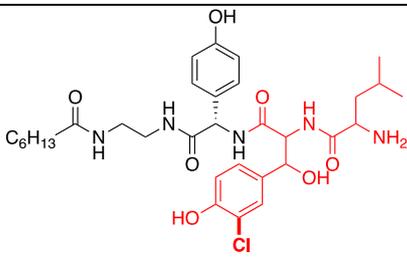
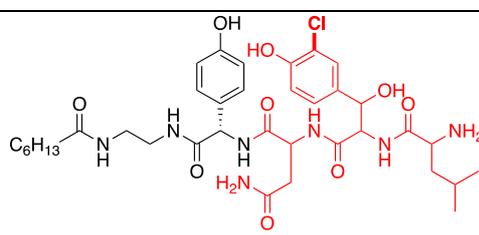
			2.48×10^5
Hydroxy-phenylglycine 178	1.29×10^{10}		2.58×10^6
			9.62×10^5

Table 5 - Intermediates captured from the vancomycin biosynthetic pathway and their intensities normalised relative to the total ion chromatogram (TIC) of the sample.

5. Investigation of scleric acid biosynthesis

5.1 Genome mining of Streptomyces

Actinobacteria, particularly those of the genus *Streptomyces*, are prolific producers of natural products. Screening efforts in culturable *Streptomyces* led to a vast number of antimicrobial compounds, with the result that between 1955 and 1962, roughly 80% of antibiotics originated from Actinobacteria.³⁰⁰ However, this screening of readily produced metabolites combined with the widespread nature of some gene clusters means that the rediscovery rate of natural products can be up to 99%.³⁰¹ The genome sequencing of *Streptomyces coelicolor* A3(2) brought about a paradigm shift in the search for novel natural products.³⁰² This revealed that in addition to the four studied natural products from *S. coelicolor*, the genome contained an additional 18 BGCs. This has been extended to a number of other actinomycetes, suggested there is an untapped wealth of compounds that are not detected under laboratory conditions. It is suggested that up to 90% of BGCs are cryptic under standard conditions, where they are silent or produced at very low levels.³⁰³

Part of the reason for the silent nature of these BGCs is the tight-control of these clusters at the transcriptional level. One abundant family of transcriptional repressors are the TetR family of proteins. Over 200,000 members have been identified, with 200 being partially characterized.³⁰⁴ The family is named after TetR, which controls the regulation of tetracycline resistance in *E. coli*. TetR functions by binding DNA as a pair of dimers between *tetR* and *tetA*, which encodes a tetracycline efflux pump. Tetracycline binds TetR in a conformation which prevents it binding DNA, allowing transcription of *tetA* and inducing resistance.³⁰⁵ These regulators have a diverse set of cellular functions, but a number have been associated with the regulation of antibiotic biosynthesis.³⁰⁶ It has been shown that deletion of TetR-like repressors within the genomes of *Streptomyces* can induce the production of novel secondary metabolites, as with methylenomycin, **194**, in *S. coelicolor* and the gaburedins in *S. venezuelae* shown in Figure 148.^{195,307}

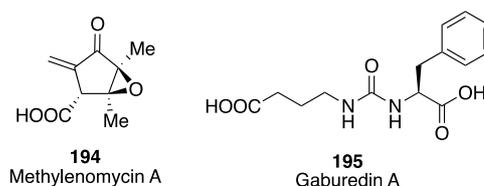


Figure 148 - Structures of methylenomycin A and gaburedin A, products discovered by the depression of biosynthetic gene clusters.

The five gene regulatory cassette in the BGC of methylenomycin in *S. coelicolor* A3(2) is well characterised. It has been shown that *mmvR* and *mmfR* code for TetR-like transcriptional repressors, while *mmfL*, *H*, and *P* are responsible for the biosynthesis of signalling molecules, the methylenomycin furans. Deletion of *mmvR* has been demonstrated to be effective for the depression of the cluster. This shows that manipulation of the regulatory cassettes of putative BGCs can be an effective strategy for the discovery of natural products usually silent under laboratory conditions.

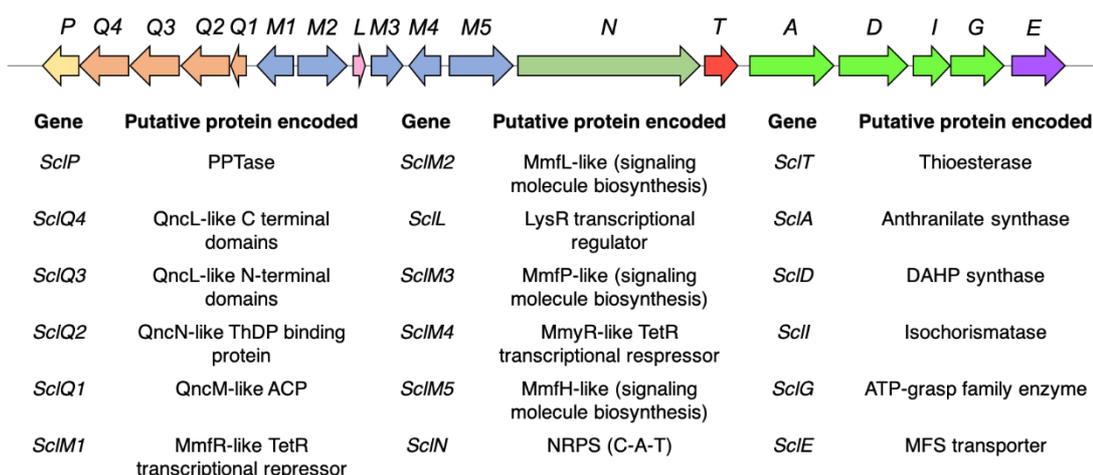


Figure 149 - The biosynthetic gene cluster of scleric acid, with the putative proteins encoded by the genes.

This approach was used for the discovery of the novel natural product scleric acid.¹⁹⁴ Initially, homologues of the 5 gene cassette were identified using a combination of MultiGene BLAST and clusterTools.^{308,309} This approach found 14 actinomycete which contained all 5 genes within a 50 kb region of the genome. A BGC from *Streptomyces sclerotialis* NRRL ISP-5269 was chosen from these, shown in Figure 149. Bioinformatic analysis of the cluster could not predict a putative structure of the natural product encoded on the basis of the enzymes involved. This potentially increased the likelihood of discovering a truly novel compound. The

cluster contained 2 operons adjacent to the regulatory cassette, encoding 18 genes in total. Of these 18, 11 were involved in biosynthesis, 6 in the regulation of the cluster, and 1 predicted to encode a membrane transporter protein.

5.2 Discovery of scleric acid

As *S. sclerotialis* proved to be genetically intractable, a 33 kb region of genomic DNA containing the 19.8 kb BGC was captured using TAR by Dr Fabrizio Alberti. This was integrated into the genomes of *S. coelicolor* M1152 and *S. albus* J1074. Analysis of ethyl acetate extracts of these heterologous hosts showed the absence of any new metabolites resulting from the introduction of the cluster, likely due to the action of the TetR-like transcriptional repressors.

Knockout of the *mmyR* homologue *sclM4* using pCRISPomyces-2 to achieve a 20 bp out-of-frame deletion was successful in the heterologous host *S. albus* and resulted in the production of a new metabolite. LC-MS identified the exact mass as 278.1020, from which the molecular formula $[C_{14}H_{15}NO_5]+H^+$ (m/z: 278.1023) could be assigned. This compound, named scleric acid, **196**, and shown in Figure 150, was isolated and characterised by Dr Lijiang Song. 1D and 2D NMR spectra of the purified compound were recorded, and the stereochemistry of the proline residue was determined by hydrolysis of the natural product and derivatisation with Marfey's reagent, which showed it to be L-proline.³¹⁰

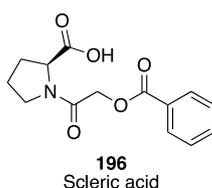


Figure 150 - Structure of scleric acid, a novel natural product from a derepressed BGC from *S. sclerotialis*.

5.3 Proposed biosynthetic route to scleric acid

The molecular formula of scleric acid, combined with homology of the enzymes of the BGC to those of known function allowed the proposal of a chemical structure **197** and a biosynthetic pathway shown in Figure 151.

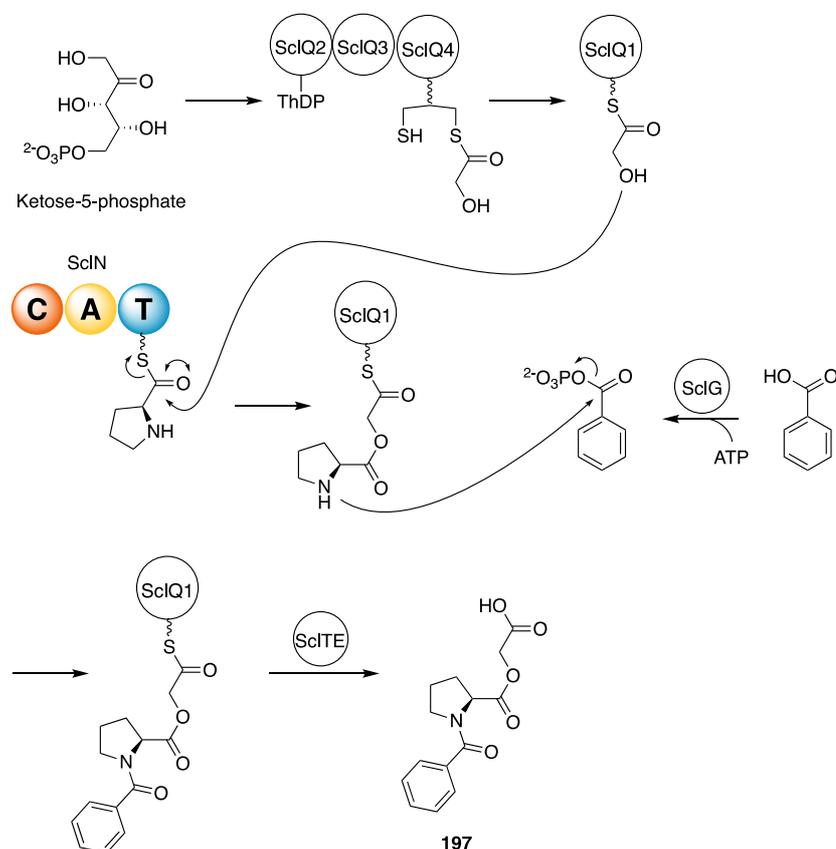


Figure 151 - Proposed structure of scleric acid, **197**, and its biosynthesis.

Crucially, three key building blocks: glycolic acid, benzoic acid, and L-proline, could be identified. The glycolic acid moiety is thought to be synthesized by SclQ1-4. *SclQ1-4* are homologous to *QncN*, *QncL*, and *QncM* in the *Streptomyces melanovinaceus* BGC encoding naphthyridinomycin biosynthesis.³¹¹ SclQ2 is homologous to QncN, a thiamine diphosphate binding protein. SclQ3 corresponds with the pyruvate dehydrogenase/transketolase pyrimidine binding domain and transketolase domain of QncL. SclQ4 is homologous to QncL, which contains a lipoyl attachment domain and an acyl-transferase domain, while SclQ1 is an acyl-carrier protein (ACP) like QncM. This are suggested to transform a ketose phosphate to a glycolicacyl unit, shown in Figure 152, which is incorporated into scleric acid.

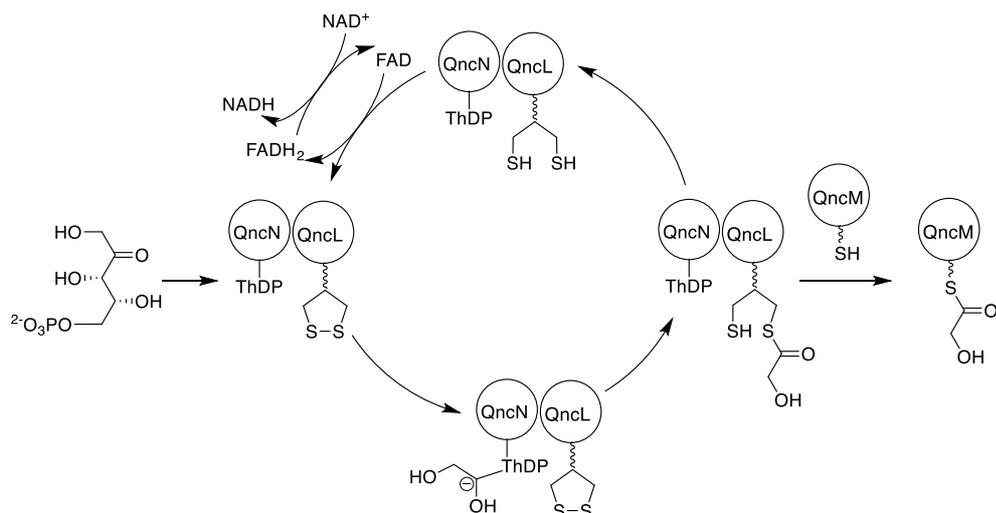


Figure 152 - The QncN, L, M pathway to convert xylose-5-phosphate to a glycolicacyl-S-ACP in naphthridinomycin biosynthesis, homologues of which are found in the scleric acid BGC.

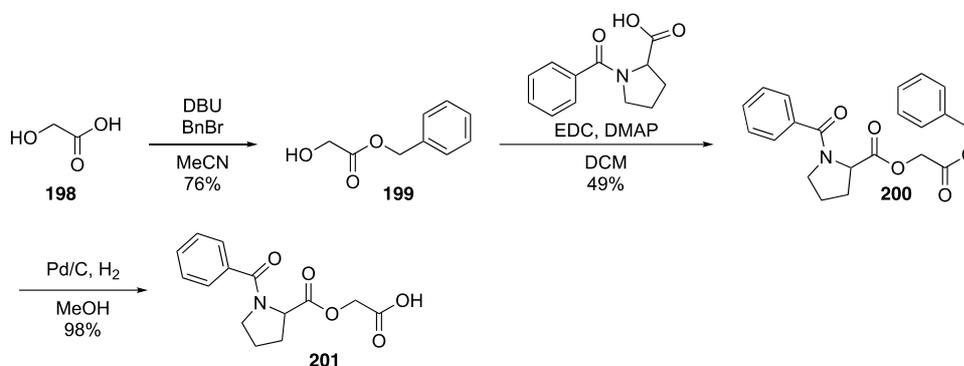
SclA, *sclD* and *sclI* have high homology to genes responsible for benzoic acid synthesis in *Streptomyces* sp. YN86.³¹² These are an anthranilate synthase PauY18, homologous to *SclA*, a 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase, PauY21, homologous to *SclD*. *SclI* is homologous to PauY19, an isochorismate synthase.

Finally, the proline is incorporated through the action of *SclN*. This is an NRPS which consists of a single minimal module containing a condensation domain, an adenylation domain predicted to incorporate L-proline, and a thiolation domain.¹⁹⁹ Additionally, a thioesterase *SclT* is thought to catalyse the hydrolysis of the condensation product of *SclN* from the T domain. *SclG* is an ATP-grasp family enzyme thought to bind and activate the benzoic acid unit and promote a condensation reaction with the product of *SclN*. *SclE* encodes a putative MFS transporter that may be responsible for the export of scleric acid once produced.

5.4 Synthesis and characterisation of a scleric acid isomer

In order to validate the proposed structure of scleric acid, an authentic standard was produced, as shown in Scheme 11. Initially, glycolic acid, **198**, was protected as a benzyl ester, **199** using 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) and benzyl

bromide. This was then coupled using EDC and DMAP to the racemic commercially available 1-benzoyl-pyrroldine-2-carboxylic acid to afford the benzyl ester of the proposed scleric acid, **200**. This was then deprotected by hydrogenolysis over a Pd/C catalyst to give the synthetic standard, **201**, of the proposed structure for scleric acid.



Scheme 11 - Synthesis of the initially proposed structure of scleric acid.

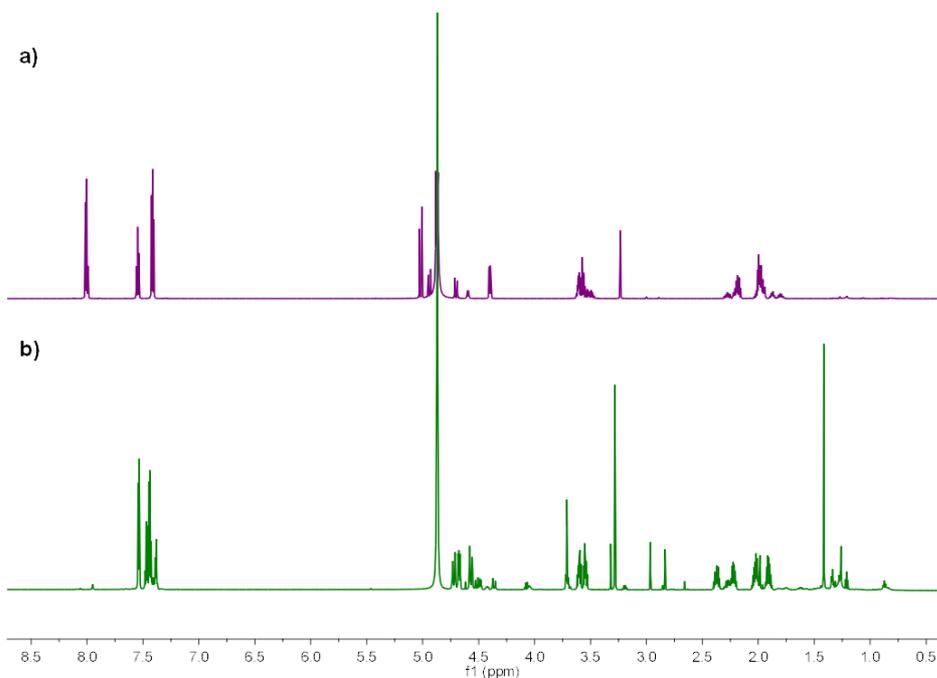


Figure 153 - Comparison of the ¹H NMR spectra of authentic scleric acid (a), with the synthetic standard of the proposed structure (b).

However, comparison of the HPLC retention time and NMR spectra of authentic scleric acid, **196**, produced by the heterologous host with the chemically synthesised standard, **201**, showed significant differences, notably in the aromatic region, as shown in Figure 153. Additionally, the signals of the standard between 5.0

– 4.5 ppm, those of the pyrrolidine CH group and the glycolic acid CH₂ group were clearly shifted from those of scleric acid. This suggests a different connectivity of the proline and glycolic acid moieties in the true natural product. The initially proposed structure and associated biosynthetic pathway were therefore not correct, and a new structure had to be proposed.

5.5 Biosynthesis of scleric acid

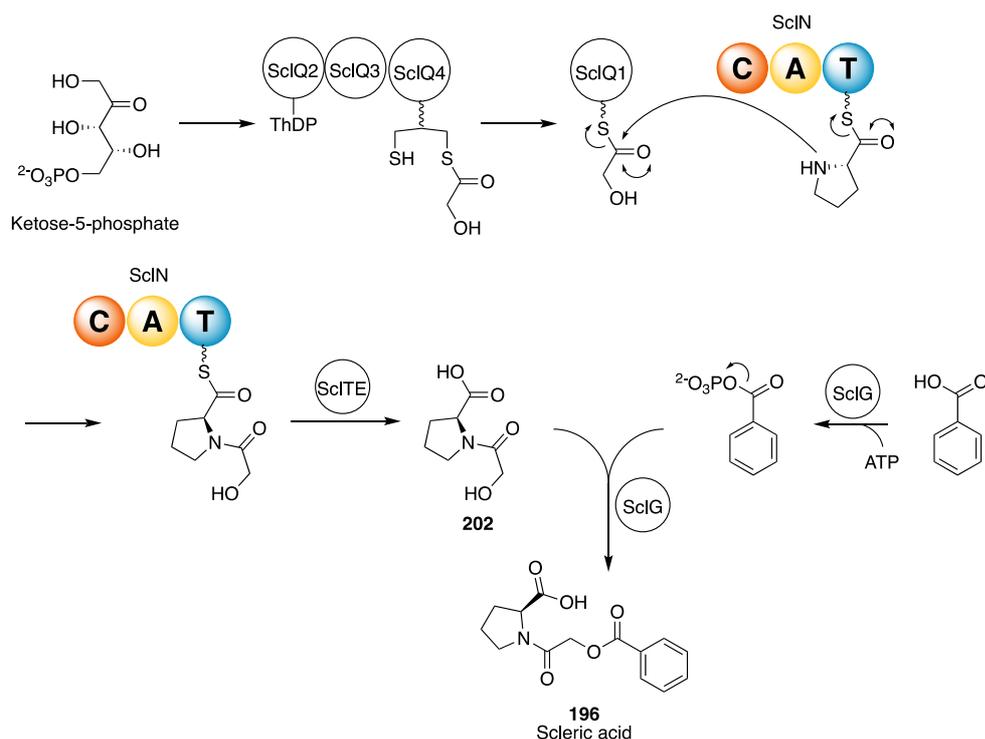


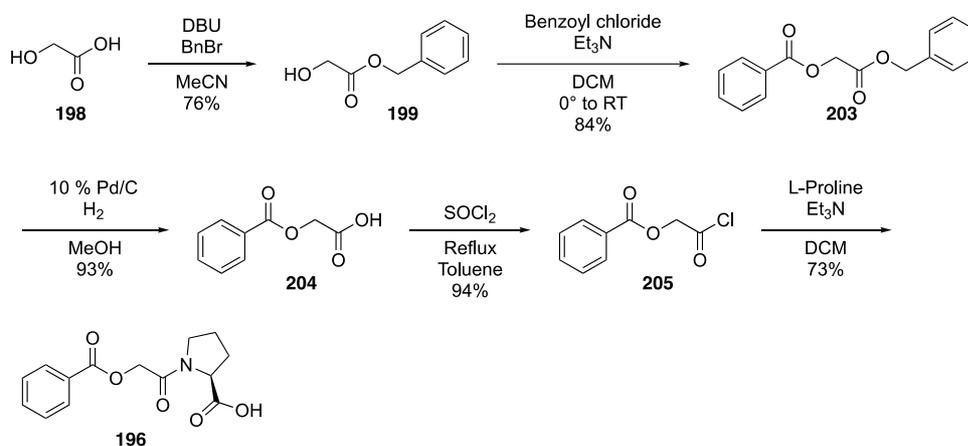
Figure 154 - Revised biosynthetic pathway to scleric acid accounting for the newly proposed chemical structure.

A revised biosynthetic scheme was proposed consistent with this structural revision, as shown in Figure 154. The condensation domain of SclN is proposed to catalyse an amide bond formation between L-proline and the SclQ1-bound 2-hydroxyacetyl thioester.

This NRPS-bound intermediate is then hydrolytically released, before ester formation with benzoic acid catalysed by SclG.

A synthetic standard for the new proposed structure was also produced, according to Scheme 12. As for the previous standard, glycolic acid was protected as

the benzyl ester, **199**. The free alcohol could then be esterified using benzoyl chloride and Et₃N, before hydrogenolysis to deprotect the carboxylic acid. Thionyl chloride could then be used to generate an acyl chloride, **205**, *in situ*, before the addition of L-proline to give scleric acid, **196**.



Scheme 12 - Synthetic route to the synthetic standard of scleric acid, used to confirm the structure of the natural product.

Comparison of scleric acid with the new synthetic standard by 1D and 2D NMR were in excellent agreement as shown in Figure 155, confirming the correct structure of scleric acid. The retention times of the two compounds in HPLC were also identical.

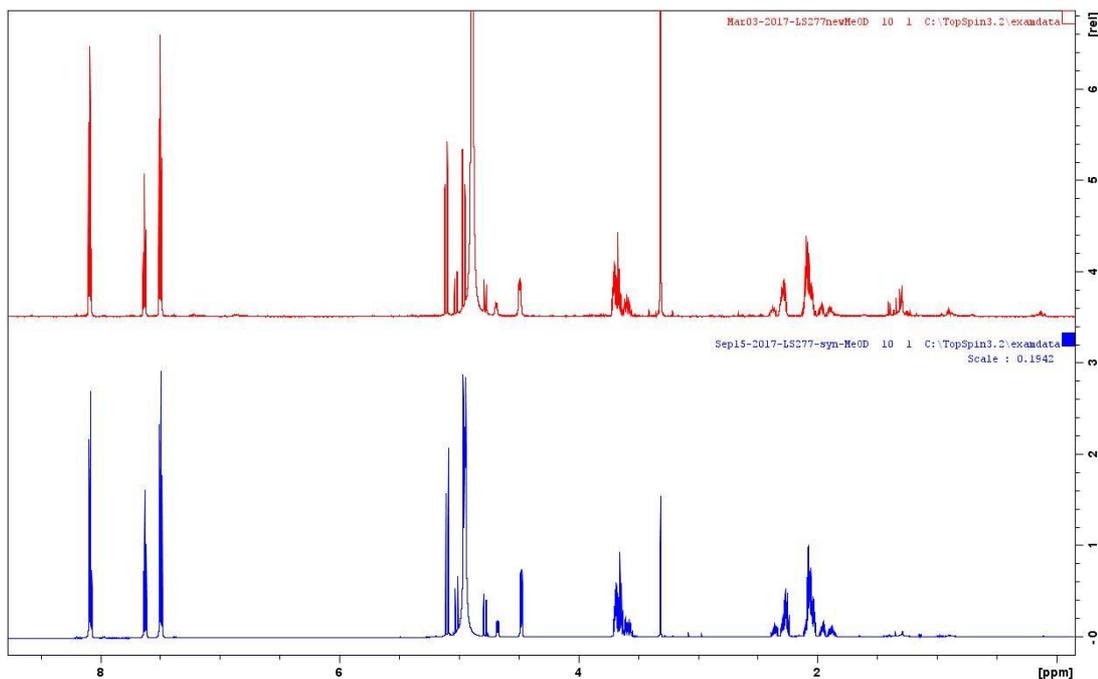
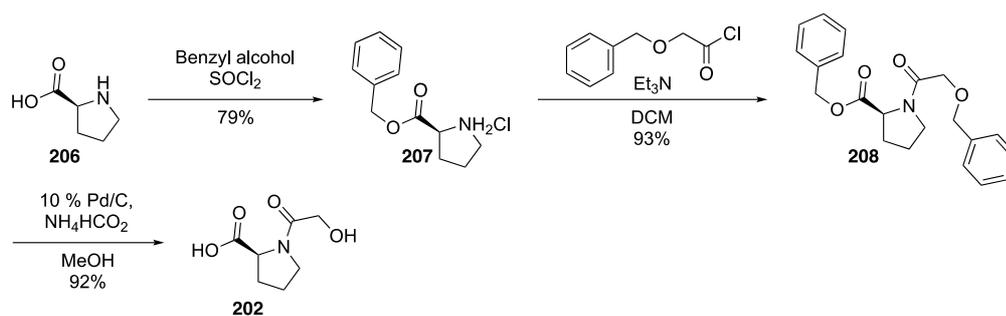


Figure 155 - Comparison of the ^1H NMR spectra of authentic scleric acid (top), and the synthetic standard (bottom).

To provide additional evidence for the proposed biosynthetic route, a number of double knockout mutant strains were constructed in the derepressed $\Delta sclM4$ strain by Dr Fabrizio Alberti. The $\Delta sclM4 \Delta sclN$ (NRPS knockout), and $\Delta sclM4 \Delta sclA$ (anthranilate synthase knockout) both exhibited abolished scleric acid production. Scleric acid was significantly reduced, but not abolished, in $\Delta sclM4 \Delta sclQ1-4$ strain. This is likely due to the fact that some glycolic acid is produced in *Streptomyces* as part of primary metabolism.³¹³ In the double knockouts $\Delta sclA$ and $\Delta sclQ1-4$ a new metabolite was detected, with a retention time of 3.0 minutes in HPLC and an exact mass of 174.0763. Based on the molecular formula of this mass ($[\text{C}_7\text{H}_{12}\text{NO}_4] + \text{H}^+ = 174.0761$) and the biosynthetic pathway, this was proposed to be (2-hydroxyacetyl)proline. Consistent with the condensation of glycolic acid and proline in the NRPS module, it was not present in the $\Delta sclN$ double mutant strain. However, this intermediate could not be isolated in sufficient quantities for purification and NMR analysis. It was therefore synthesised, as shown in Scheme 13.



Scheme 13 - Synthetic route to (2-hydroxyacetyl)proline, 202, an intermediate in scleric acid biosynthesis.

L-Proline was protected as the benzyl ester, **207**, using thionyl chloride and benzyl alcohol. Commercially available benzyloxyacetyl chloride could then be coupled to form the amide bond, with the alcohol protected as a benzyl ether to aid purification. The benzyl ester and ether could then be removed by hydrogenolysis to afford the intermediate. This compound, **202**, gave an identical retention time and mass spectrum to the species accumulated in the double knockout strains, confirming the identity of the product.

To confirm that (2-hydroxyacetyl)proline, **202**, was an intermediate in the biosynthesis of scleric acid, it was supplemented to the $\Delta sclM4 \Delta sclN$ strain at a 5 mM concentration, which was unable to produce scleric acid, **196**. In the acidified extracts of this feeding after growing for 5 days, scleric acid could be detected as shown in Figure 156. This shows that **202** can restore production and is a true pathway intermediate.

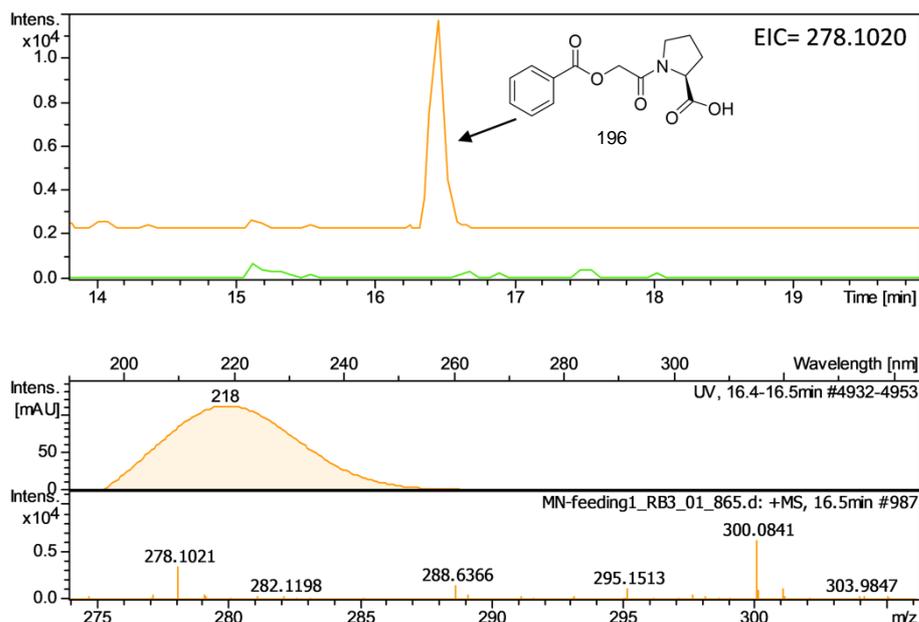
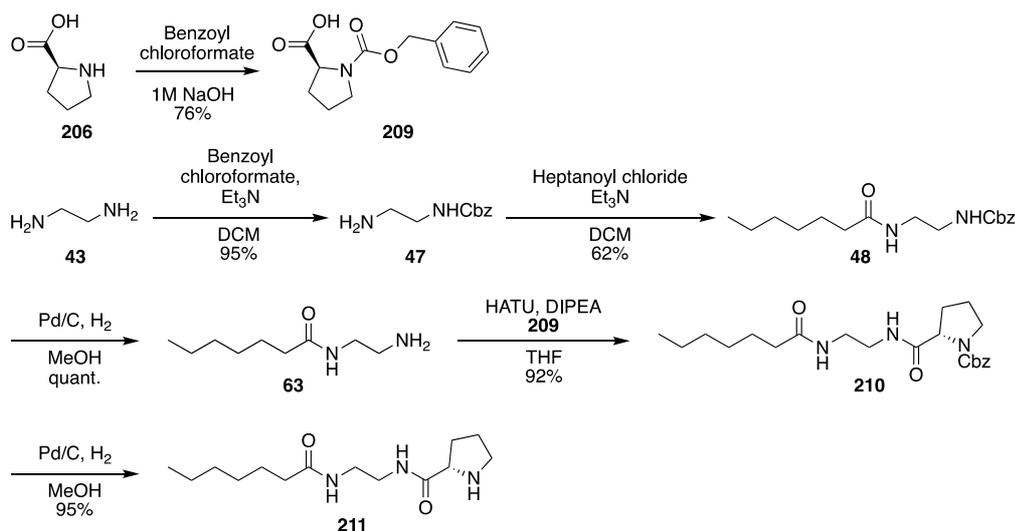


Figure 156 - Restoration of production of scleric acid, 196, in $\Delta sclM4 \Delta sclN$ double mutant when supplemented with 5 mM (2-hydroxyacetyl)proline, 202, (orange trace) compared to the unsupplemented double mutant control (green trace).

The biosynthesis of scleric acid was also investigated through the use of chemical probes to offload species involved in NRP biosynthesis. Within the biosynthetic pathway, the T-S-proline acts as the acceptor amino acid in the condensation step, while the SclQ1-bound glycolic acid thioester would act as the donor. Using chemical probes designed to act as a mimic for the acceptor amino acid, it was aimed to offload the product of condensation with glycolic acid, supporting the proposed pathway.

Probes based on glycine, *N*-methylglycine, and proline were chosen to examine this NRPS. Glycine has proved effective in a number of systems, likely due to its lack of steric bulk, while *N*-methylglycine would potentially better mimic the secondary amine of the natural proline substrate. The proline probe, **211**, was synthesized as shown in Scheme 14, while the glycine, **81**, and *N*-methylglycine probe, **67**, synthesis was described in previous chapters.



Scheme 14 - Synthesis of an NRPS chemical probe based on L-proline.

Masses corresponding to the offloading of glycolic acid by the probes, with MS² to support the proposed structures, could be found. However, these masses could also be detected in strains containing the repressed *scl* cluster, as well as $\Delta sclM4 \Delta sclN$ double knockout, which lacks the NRPS to catalyse amide bond formation, shown in Figure 157. In these strains a secondary source of glycolic acid is present, as seen by the residual scleric acid production in the $\Delta sclM4 \Delta sclQ1-4$ mutant. It is likely that the probes can react with this source of glycolic acid like the proline in this strain, giving rise to these glycolic acid bearing products *via* a second pathway. However, due to this activity, the probes cannot show bond formation catalysed by the C domain of the NRPS ScIN, and so do not provide additional evidence for the biosynthetic pathway.

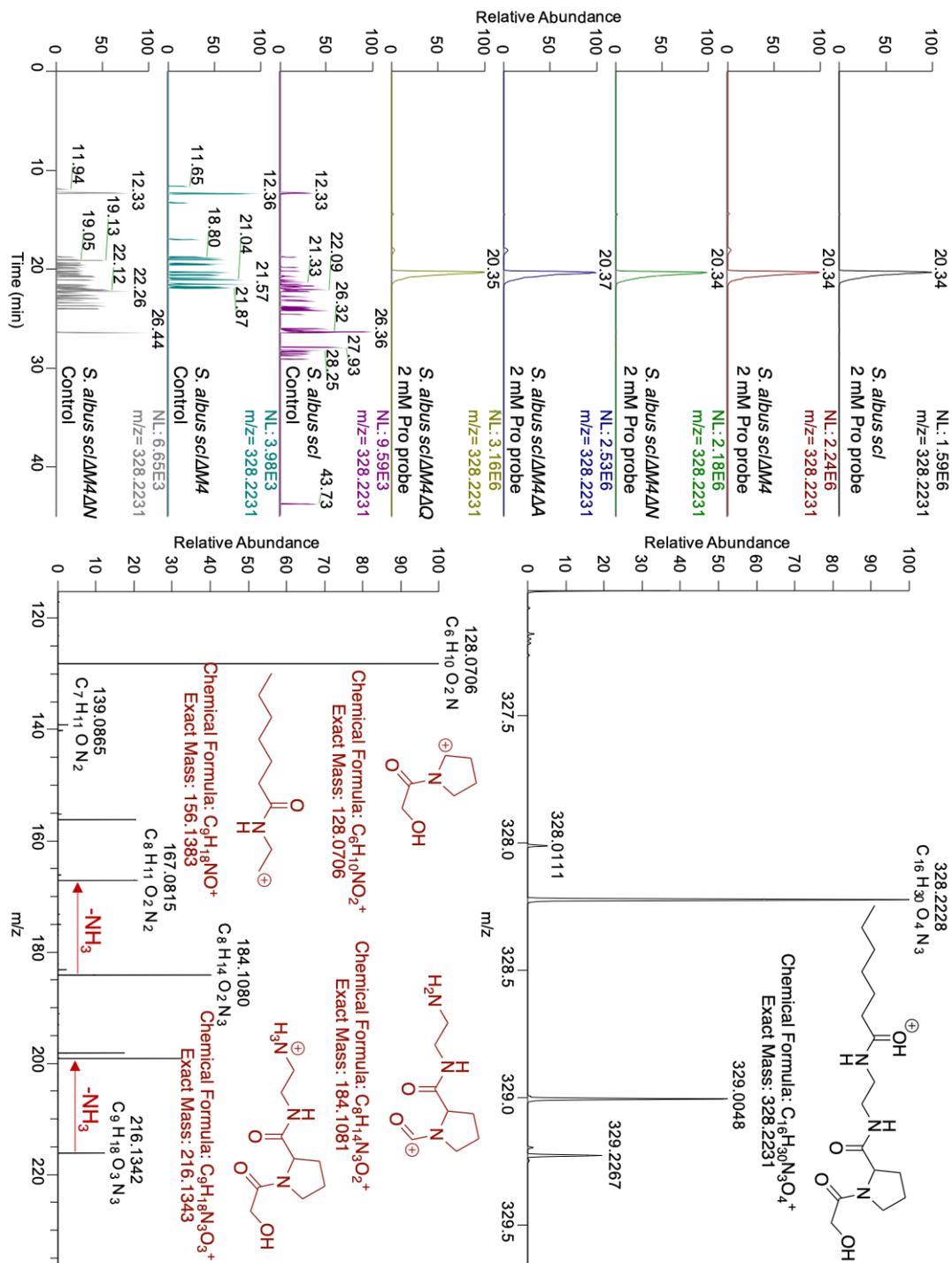


Figure 157 - - Investigating scleric acid biosynthesis using a Pro-based probe. Extracted ion chromatograms (EICs) of m/z: 328.3321 in *scl* containing *S. albus* strains with and without the probe. High resolution mass spectra of the detected species, and proposed structure with matching molecular formula, and MS² fragmentation of the species with proposed structures matching the molecular formula.

5.6 Biological activity of scleric acid

A facile synthetic route to scleric acid allowed the examination of a range of possible bioactivity assays. Antimicrobial testing against the ESKAPE pathogens: *Enterococcus faecium*, *Enterobacter cloacae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, and *Staphylococcus aureus* was tested through the Warwick Antimicrobial Screening Facility, but gave no observable MIC. Scleric acid was also tested for a range of biological activities by submission to Eli Lilly's Open Innovation Drug Discovery (OIDD) platform. This gave two notable results.

In a single point primary assay, at 20 μM , 32% growth inhibition of *Mycobacterium tuberculosis* H37Rv was observed. In assays against nicotinamide *N*-methyltransferase (NNMT) scleric acid showed an IC_{50} of 178.0 μM (NNMT MNAN) and 186.6 μM (NNMT SAH) on a concentration response assay. NNMT is a cancer-associated metabolic enzyme, with its overexpression linked to tumorigenesis.³¹⁴ It catalyses the transfer of a methyl group from SAM to nicotinamide, giving (*S*)-adenyl-L-homocysteine (SAH) and 1-methylnicotinamide (MNAN) as shown in Figure 158.

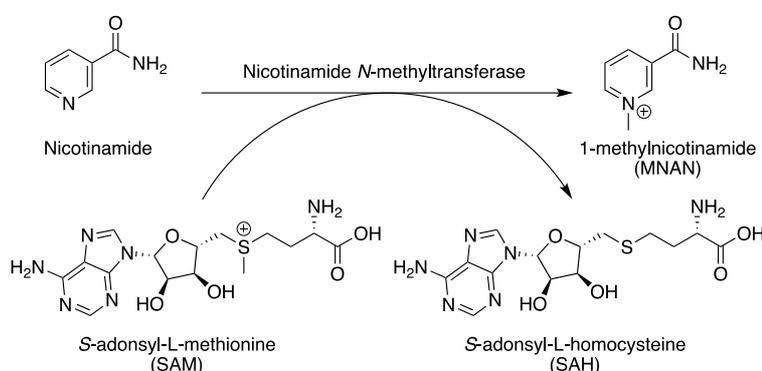


Figure 158 - The reaction catalysed by nicotinamide *N*-methyltransferase to produce 1-methylnicotinamide and (*S*)-adenyl-L-homocysteine from nicotinamide and SAM.

5.7 Generation of scleric acid derivatives

In light of this biological activity, a small number of derivatives of scleric acid were designed and synthesized in our group, **212** – **217**, as shown in Figure 159. These aimed to investigate the importance of some of the key structural elements of scleric acid, to identify essential moieties for bioactivity. These included the stereochemistry

and carboxylic acid of the proline residue and the ester linkage of glycolic acid and benzoic acid.

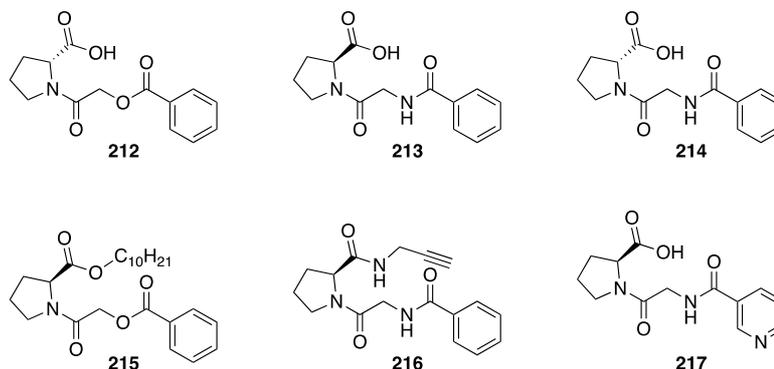
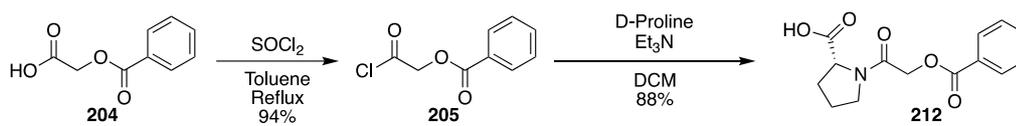
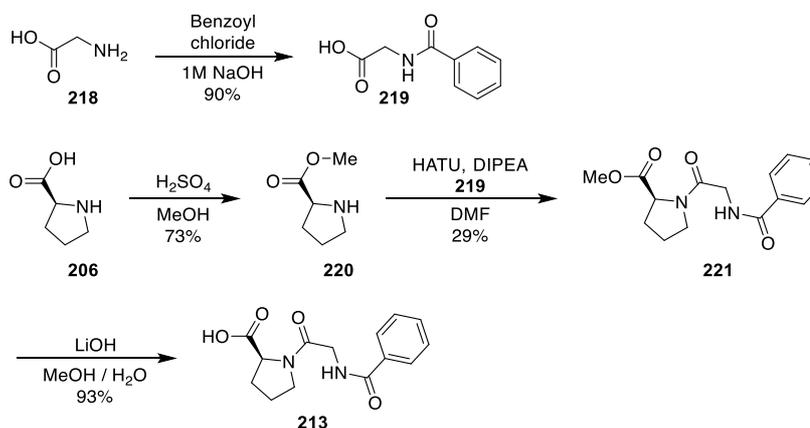


Figure 159 - Structures of the scleric acid derivatives synthesized to investigate the contribution of certain moieties to bioactivity.

The D analogue of scleric acid, **212**, could be prepared by an equivalent route to scleric acid shown in Scheme 15, with the substitution of D-proline in the final step to react with the formed acyl chloride, **205**, to form an amide bond.



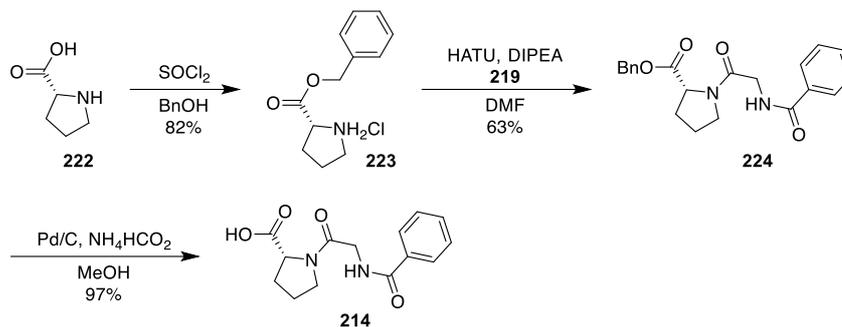
Scheme 15 - Synthetic route to the D-scleric acid analogue.



Scheme 16 - Synthetic route to the L-amide derivative of scleric acid, 213.

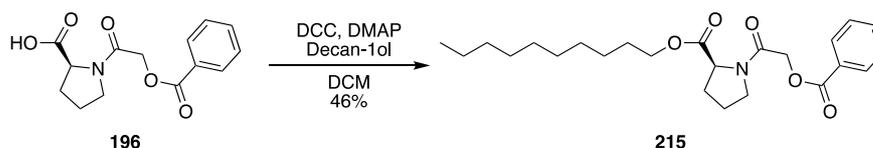
Analogues in which the ester is replaced by an amide bond were generated by reacting glycine with benzoic chloride to afford hippuric acid, **219**. This could then be

coupled to a protected proline, **220**, before methyl ester hydrolysis to give the desired derivative, **213**, as in Scheme 16.



Scheme 17 - Synthetic route to the D-amide derivative of scleric acid, **214**.

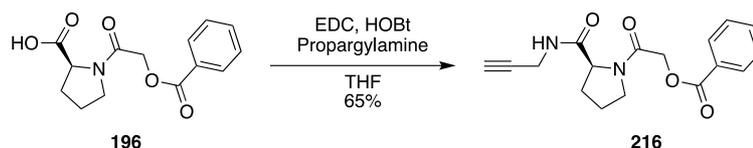
Due to the higher cost of D-proline, an alternate protection strategy was used as shown in scheme 17, with a benzyl ester, **223**, rather than the methyl ester used for the L-enantiomer. This would give a more efficient protection using thionyl chloride, and the deprotection would not require extraction of the small polar molecule from water.



Scheme 18 - Esterification of scleric acid with decan-1-ol.

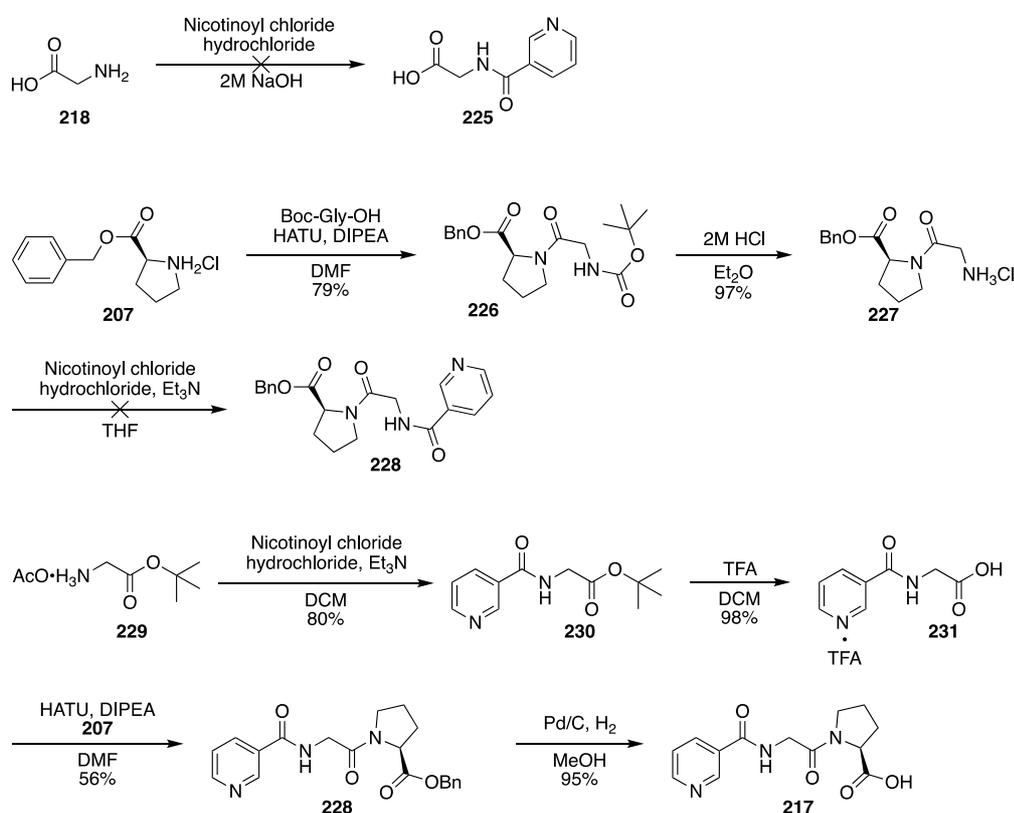
Scleric acid, **196**, could also be selectively derivatised at the carboxylic acid of the proline moiety. One approach was to couple decan-1-ol with scleric acid to give the decyl ester, **215**, shown in scheme 18. This was hoped to improve cellular uptake by masking the carboxylic acid, which would be charged at physiological pH, and improving lipophilicity.

A derivative containing an alkyne was also synthesised by Master student Rishi Chudasama, by coupling propargylamine to scleric acid, **196**, as shown in scheme 19. It was envisioned that this alkyne, **216**, could be used to investigate intracellular targets of scleric acid by coupling to a fluorophore using bio-orthogonal CuAAC click chemistry.



Scheme 19 - Synthetic route to the alkyne derivative, 216, by coupling with propargylamine, prepared by Rishi Chudasama.

Finally, in light of the inhibitory activity of scleric acid against NNMT, a compound containing a nicotinamide moiety in place of the benzoic acid ester was synthesized. Of the few NNMT inhibitors reported, a number incorporate this moiety or functionalised variants to bind to the active site. The incorporation of the nitrogen in the heterocycle should better mimic the nicotinamide substrate, and possess greater affinity to the substrate binding pocket of NNMT.



Scheme 20 - Synthetic route towards the nicotinamide derivative of scleric acid.

As shown in scheme 20, initially the coupling of glycine, **218**, with commercially available nicotinoyl chloride hydrochloride was attempted under Schotten-Baumann conditions. However, this produced a zwitterionic compound, **225**, which proved impossible to extract from the aqueous reaction mixture. To avoid this,

a later introduction of the nicotinamide moiety was tried. First the protected Gly-Pro dipeptide **226** was synthesised, before removing the *N*-terminal Boc ready for coupling. However, reaction with nicotinoyl chloride hydrochloride did not yield any of the desired product, possibly due to the ability of the benzyl-protected dipeptide to form a diketopiperazine product.

To avoid this possibility, the nicotinamide-glycine coupling was re-examined. Using the *tert*-butyl ester of glycine, **229**, the coupling proceeded in good yield, and could be facilely deprotected with TFA to afford compound **231**. This could then be coupled with the L-proline benzyl ester, which was deprotected with hydrogenolysis to afford the final derivative, **217**.

These compounds were tested against *M. tuberculosis* at the Institute of Infectious Disease and Molecular Medicine at the University of Cape Town. It was found that all compounds with the exception of the decanol ester **215** had MIC₉₀ values of >250 µg/ml. The decanol ester **215** had an MIC₉₀ of 125 µg/ml. While the activity is not comparable to known anti-TB drugs (e.g. isoniazid >4 µg/ml), this suggests that cell permeability may be a limiting factor for the compounds tested.³¹⁵ The addition of a highly lipophilic group may allow a greater intracellular concentration to be achieved. The lack of activity observed with the alkyne derivative **216** would suggest that the carboxylic acid moiety is necessary for activity and can only be masked temporarily in the form of a pro-drug. With this information, a second-generation of analogues could test the effect of modifications to other functionalities of scleric acid, such as the aromatic group. The analogues could also be assayed to examine the inhibition of NNMT, particularly the nicotinamide based analogue **217**. This will be discussed further in chapter 6.

6. Conclusions and Future Work

6.1 Investigation of echinomycin biosynthesis

The investigation of echinomycin biosynthesis described in chapter 2 consisted of chemical probing of the echinomycin NRPS, attempts to edit the echinomycin BGC *in vivo*, and reconstitution and study of module 1 *in vitro*.

6.1.1 Chemical probing of echinomycin biosynthesis

The attempts to investigate the flexibility of the echinomycin pathway using novel chemical probes has been successful with substrates closely mimicking the pantetheine moiety carrying both cognate and noncognate amino acid moieties. The use of a wide range of amino acid side chains within the chemical probes allowed the substrate tolerance of the C domains within the two NRPSs to be profiled simultaneously. This produced a number of interesting results; in particular I found that chemical probes based on the non-cognate amino acid glycine were effective at capturing intermediates from a number of modules, indicating the C domains possess a degree of substrate tolerance.

The species produced by the interception of biosynthetic intermediates with chemical probes are unnatural peptides produced by the NRPS, and potentially offer an insight into the potential to engineer the assembly line to create unnatural products.

Novel chemical probes containing a thioamide functional group instead of an amide at the *N*-acyl amide were synthesised in order to examine the tolerance of the C domains of the echinomycin pathway towards new functional groups, that while isosteric with an amide would expand the chemical space of the unnatural peptides that could be produced.

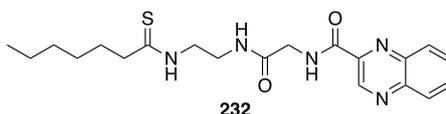


Figure 160 – Intermediate, 232, offloaded by the thioamide glycine probe, 81.

However, these molecules did not prove useful for investigations in *Streptomyces lasaliensis*: when fed to cultures, they proved largely toxic and only a

single intermediate, **232**, could be identified from cultures supplemented with the thioamide glycine-based probe **81**. This characterised species corresponded to the offloading of the quinoxaline-2-carboxylic acid starter unit of the pathway. This suggest that the carbonyl group is not strictly required from the acceptor amino acid during peptide bond formation and can be replaced by similar functional groups to some extent. Future investigation could involve the synthesis and the evaluation of probes bearing other isosteric substitutions at the same position or others. For example, bioisosteric replacement of the amide could be achieved moieties such as a ketomethylene group, a urea, or a sulfonamide, as shown in Figure 161.

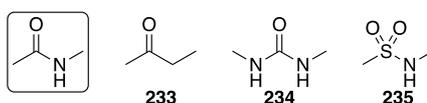


Figure 161 - Amide isosteres: ketomethylene **233**, urea **234**, sulfonamide **235**.

Indeed the same isosteric strategy could be applied to the amide bond between the diamine moiety of the probe and the amino acid, itself a replacement of a thioester.

Given the general toxicity of the thioamide moiety, this may be an interesting structural motif to include in antimicrobial compounds. It would be interesting to investigate the effect of the thioamide-containing probes on other microorganisms to understand if the toxicity of the probes is specific to *S. lasaliensis* or if it is general and less selective.

Substrates containing *N*-methylated amino acid moieties also proved poor probes of NRPS- catalysed amide bond formation *in vivo*, as no intercepted intermediates could be observed. This was somewhat unexpected as the presence of a methyl group on an N atom increases its nucleophilicity and activated *N*-methylated amino acids are expected to act as true acceptors in C-catalysed amide bond formation. However, the additional steric bulk can slow amide formation: as the probes are not acting in isolation, but competing with enzyme-bound biosynthetic substrates *in vivo*, we speculate that this rate reduction can greatly reduce the ability of the probes to efficiently intercept nascent biosynthetic intermediates. It has been shown that the same condensation domains can catalyse amide bond formation using either *N*-methyl or non-methylated amino acids *in vitro* in the NRPS of actinomycin from *Streptomyces*

*chrysomallus*¹⁸⁷ It would therefore be of interest to examine the use of non-methylated chemical probes in biosynthetic pathways which produce heavily *N*-methylated NRPs, where the expectation is they would react at a relatively faster rate compared to the *N*-methyl amino acids of the pathway and lead to an increased titre of offloaded species. Also *in vitro* experiments with recombinant C domains would shed light on whether *N*-methylated amino acid probes can be used to generate ‘unnatural’ peptides in the absence of a competing acceptor molecule.

The synthesis of a chemical probe incorporating cysteine **75** was also achieved. The preparation of this substrate was not trivial due to the extensive disulfide bond formation of the thiol moiety under a number of deprotection conditions, and the elimination of H₂S under harsh conditions. Unfortunately the use of this molecule did not provide any new insights into the relative substrate specificity of the echinomycin NRPS C domains as the probe was not found to intercept any intermediate species. The probe could be synthesised in a monomeric form but was observed to readily dimerise to the disulfide species under culture conditions, which would greatly limit the probe available to intercept intermediates. The number of NRPSs available to interact with the probe would also be limited due to the reduced growth of *S. lasaliensis* observed when grown in the presence of the cysteine-based probe at 1.0 or 2.0 mM concentrations. Additionally, the formation of mixed disulfide species greatly complicates the identification of captured intermediates using LC-MSⁿ. The issues concerning toxicity and disulfide formation before intermediate offloading could potentially be addressed by masking the thiol moiety, such as through the use of a thioester. This would slow disulfide formation upon *in vivo* hydrolysis to produce the desired probe and potentially reduce its toxicity. Disulfide formation after offloading could be dealt with through the addition of a reducing agent such as dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine (TCEP) before analysis to reduce any disulfide species to free thiols. Also, future *in vitro* work with recombinant enzymes and the cysteine substrate should lead to a much more clear-cut evaluation of its ability to act as chemical probe.

6.1.2 *In vivo* editing of the echinomycin BGC

Efforts to edit the biosynthetic gene cluster of the antitumor antibiotic echinomycin in *S. lasaliensis* or in an heterologous host were undertaken in order to inactivate the two condensation domains of the NRPS Ecm6. Inactivation of the first C domain would create a strain incapable of producing echinomycin or any biosynthetic intermediates, which would serve as a useful strain for control experiments. Inactivation of the second C domain would only allow formation of a dipeptide intermediate and be envisaged to increase the interception of intermediates at the first condensation domain which has displayed a broad acceptor tolerance.

Our efforts initially focused on the use of CRISPR/Cas9 systems to edit select genes, followed by attempts to capture the entire cluster for expression in a heterologous host. The generation of pCRISPomyces2 constructs, to introduce double strands breaks which would be repaired by homology directed repair using DNA from the plasmid, was successful. However, the successful conjugation of the plasmids into the native echinomycin producer, *S. lasaliensis*, could not be achieved. Attempts to introduce a second plasmid encoding a CRISPR/Cas9-CodA(sm) system, without any targeting sequence, was also unsuccessful. Therefore, it is likely that the expression of the Cas9 protein from *S. pyogenes* is toxic to *S. lasaliensis*, as it has been noted for other *Streptomyces* species.

The capture of the BGC to allow expression and genetic manipulation in a heterologous host was then undertaken using transformation associated recombination (TAR) cloning. This was partially successful in that a small part of the gene cluster was captured. The loss of the majority of the cluster did not allow the desired genetic manipulation of the biosynthetic genes. The cluster would be introduced to the capture vectors through homologous recombination with homology sequences on the vector. The absence of homology arms in isolated DNA, combined with the growth in the presence of counter-selection, suggests subsequent genetic rearrangement of the plasmid after cluster capture. The use of an *E. coli* strain (Stbl4) designed for unstable inserts would suggest the DNA isolated from *S. cerevisiae* was already rearranged, but changes in *E. coli* cannot be ruled out.

To avoid the rearrangement of DNA *in vivo*, an alternative method which could be explored in the future is the use of Direct Pathway Cloning (DiPaC).¹¹⁷ The cluster is captured in portions through long-amplicon PCR, followed by ligation of the fragments, which can be achieved *in vitro* by HiFi DNA assembly. This would circumvent *in vivo* genetic rearrangement. The technique has also been demonstrated to function to capture the erythromycin BGC, which is a similar composition (55 kb, 73% GC) to the echinomycin BGC (36 kb, 72% GC).³¹⁶

6.1.3 *In vitro* reconstitution of Ecm6 module 1

To enable *in vitro* study of Ecm6 module 1, the proteins FabC, Ecm1, and Ecm6 C1 have been successfully cloned and expressed in soluble forms. Ecm1 has been demonstrated to be functional in the activation of quinoxaline-2-carboxylic acid using an optimised Malachite green adenylation domain assay. The condensation domain assay showed no formation of the expected amide bond between the probes and FabC-bound quinoxaline-2-carboxylic acid. This could be due to the excised C domain not being functional outside a multidomain construct, or the strict requirement for the acceptor substrate to be enzyme-bound. In future, using SNAC amino acyl substrates as the acceptor amino acids, as reported in the literature could be potentially used to further investigate the activity of the current C domain construct. The replacement of a thioester with an amide linkage may also reduce the rate of amide bond formation to a level undetectable in *in vitro* assays but remain effective in *in vivo* systems. Such a disparity between *in vitro* activity and *in vivo* function has been observed in the site-directed mutagenesis of condensation domains in enterobactin NRPS EntF.¹¹¹ In the event of a non-functional domain, expanding the construct to a C-A didomain or a C-A-T tridomain could be tested to find a functional enzyme. Once a functional condensation domain construct has been identified, the chemical probes could then be tested to compare *in vitro* activity with that observed *in vivo*.

6.2 Chemical probing of the colibactin biosynthetic pathway

As shown in chapter 3, two new chemical probes, **136** and **138**, designed to mimic the chemically challenging aminomalonate moiety were successfully synthesised. The bacterial toxin colibactin is biosynthesised by a hybrid PKS-NRPS assembly line that incorporate two aminomalonate moieties. We envisaged aminomalonate-based substrates as possible dual probes for both PKS and NRPS enzymes (figure 105). My aim was to produce protected aminomalonate probes capable of being hydrolysed to an active form at a greater extent *in vivo*. This was specifically accomplished by preparing the trifluoroethanol ester aminomalonyl probe **136** and the butanol ester **138**. However, the introduction of an amide moiety in place of the β -ketone was also required in order to increase the stability of these molecules containing an α -amino group. This was shown in the inability to prepare β -keto- α -amino esters of pivaloyloxymethyl, **118**, trifluoro, **119**, or butanol, **120**, esters, which appeared to undergo decomposition readily under mildly acidic conditions. The testing of these substrate as PKS probes lead to no detection of intercepted intermediates. One likely explanation for this is a significant decrease in the rate of decarboxylative Claisen condensation required to form carbon-carbon bonds with potential polyketide intermediates, thus limit offloading of these species. Also, the stabilised nature of the newly generated carboxylate group would also slow the formation of an active glycine-based probe for peptide bond formation. Indeed, when tested on the colibactin system, no intermediates could be observed from interception of the NRPS components of the pathway.

In view of the recently elucidated role of ClbL in amide bond formation as an amidase, α -aminoketone species generated *in vivo* similar to our probes could also act as substrates for such enzymes. Specifically, the products of the probe condensation, **236**, from ClbL and ClbI would be identical as illustrated in Figure 162.

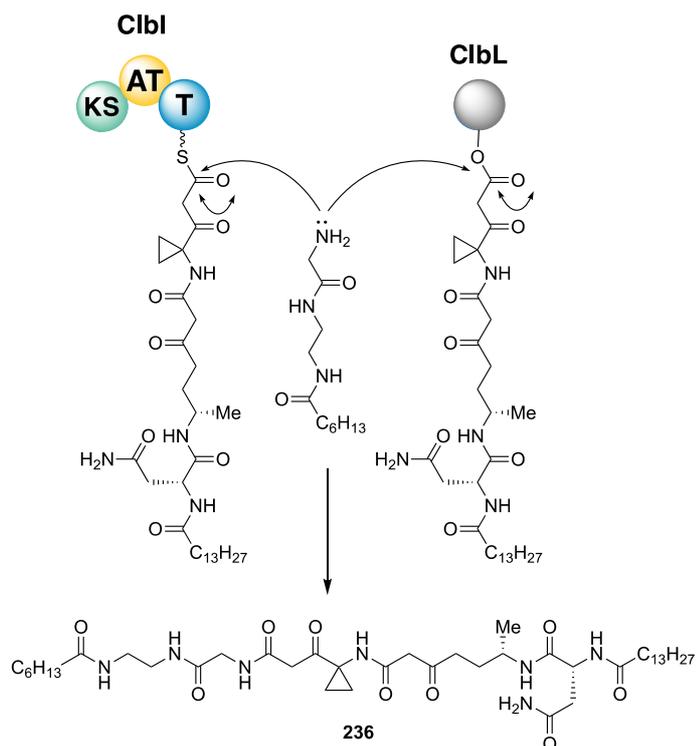


Figure 162 - The identical product that is formed from the offloading of either ClbI or ClbL by a chemical probe.

The NRPS probes developed in our lab mimic individual amino acids recruited for C-catalysed bond formation, whereas the amino substrate utilised for ClbL-catalysed reaction is a complex peptide. On this basis, we believe that condensation products such as **236** would result from offloading of biosynthetic species from ClbI rather than ClbL. Nonetheless this remains to be further confirmed by future *in vitro* experiments with a standalone recombinant ClbL. While ClbL has been shown to accept a range of α -aminoketone substrates, it has not been tested whether this promiscuity extends to an amide structure as found in the chemical probes.²³⁰

While some limitations in the use of chemical probes have emerged during our studies of the complex colibactin pathway, some interesting and useful observations have arisen. For instance, the dual feeding of a glycine-based probe **81** and a fluoromalonate-based probe **141** (kindly provided by fellow PhD student Rob Jenkins) showed the possibility to simultaneously interrogate different types of megasynthases within a single experiment. This offers an information-rich method to investigate other PKS-NRPS hybrid systems as an alternative to the proposed dual-action of the

problematic aminomalonate-based probes. The putative PKS intermediate shown in Figure 163, **237**, offloaded by the fluoromalonate probe would be formed from the interception of a bis-thiazole intermediate formed as a result of module skipping in the PKS of ClbK. This additional extension would be consistent with the action of ClbO, which was unknown at the time, but later shown to act after ClbK.

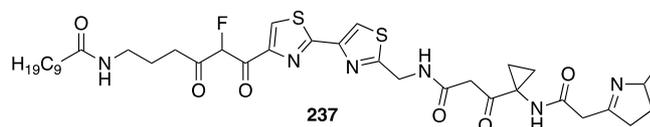


Figure 163 - Putative intermediate, 237, offloaded by the fluoromalonate probe, 141.

The previously synthesized cysteine-based probe could be exploited here as a cognate amino acid in the pathway and did lead to the identification of an offloaded intermediate. While this did not advance our understanding of the previously unknown portions of the biosynthetic pathway, it did show that the probe can be tolerated by some bacterial strains and offload intermediates, in contrast to the results obtained with *S. lasaliensis*.

The unknown nature of the intermediates on the colibactin pathway highlights a shortcoming in the current targeted use of chemical probes. While the probe-intermediate conjugates can be readily identified where they can be somehow predicted, examining enzymes of complete unknown function remains tricky. Further complications arise from biosynthetic intermediate modifications such as cyclisation, dehydration, oxidation and many others. On this basis we foresee the urgent need to develop a much more comprehensive untargeted mass spectrometry approach for the analysis of samples deriving from experiments with chemical probes in the future. Steps in this direction have already been taken within our research group.

While the structure of colibactin has been elucidated, and the biosynthetic route to the final product largely elucidated, it does offer inspiration and opportunity for future work. Many of the published mass spectrometry studies which have proved vital to the structure elucidation of colibactin has relied on the use of ¹³C-labelled amino acid precursors to verify the origin of proposed structures. The use of ¹³C-labelled amino acids in chemical probes would allow even more unambiguous

assignment of species as probe-intermediate conjugates by virtue of the characteristic mass shift.

It is also possible that as the fluoromalonate probe **141** utilised can compete with an aminomalonate extender unit, to form intermediate species **237** this could be used to understand the oxidation state of the thiazole/thioazoline ring during Claisen condensation. Deletion of the gene cassette responsible for aminomalonate biosynthesis, *clbDEF*, would eliminate competition and likely increase the titre of intermediates corresponding to the offloading at ClbO.

6.3 Chemical probing of vancomycin biosynthesis

In the investigation of the timing of biosynthetic events in the vancomycin NRPS, the ability of a range of chemical probes to offload intermediates has been demonstrated in chapter 4. The tolerance of the acceptor site of the first and second condensation domains was found to be the broadest, allowing three probes to offload intermediates at each of these stages. Notably the identification of a non-chlorinated intermediate offloaded using a Hpg-based probe strongly suggests the chlorination occurs at the level of an enzyme-bound dipeptide rather than at the enzyme-bound amino acid. The subsequent observation of a chlorinated tripeptide suggests that the chlorinated intermediates are the cognate substrate for downstream condensation domains and would rule out the late stage halogenation of a complete heptapeptide which had been proposed.

The flexibility of the first condensation domain towards non-cognate substrates such as tyrosine is interesting, as the related vancomycin-type glycopeptide pekiskomycin contains a tyrosine residue at position 2 and is a relatively recent evolutionary divergence from the vancomycin cluster.^{317,318} It is possible that the related NRPSs share some degree of substrate tolerance. On the other hand, glycine is capable of offloading from both modules 2 and 3, while no glycopeptides incorporate glycine in these positions.

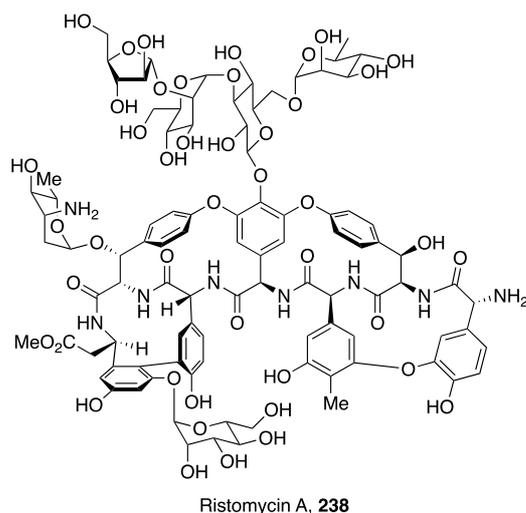


Figure 164 - Ristomycin A, a glycopeptide produced by a gene cluster related to vancomycin, containing an aromatic amino acid at residue three.

Glycopeptides produced by closely related gene clusters, such as ristomycin A shown in Figure 164, incorporate aromatic amino acids at position 3 which may partially explain the ability of tyrosine and Hpg-based probes to offload from the third module despite bearing little resemblance to the cognate asparagine residue. Such substrate promiscuity as a result of the evolutionary path of a C domain has been suggested to explain the tolerance for unnatural D-amino acids by a condensation domain in Tc_{p12}, the final NRPS in teicoplanin biosynthesis.⁵¹ Further exploration of this hypothesis could offer a basis for the rational reengineering of glycopeptide biosynthetic pathways. The timing of methylation of Dhpg at position three in ristomycin A would also present an interesting question to be addressed with chemical probes.

Indeed, the relative intensities of the offloaded intermediates provide a qualitative insight into the acceptor tolerance of the condensation domains. While the chemical probes based on aromatic amino acids generally had greater ion intensity, potentially due to improved ionisation efficiency or extraction, the lack of intermediates captured by phenylglycine and phenylalanine suggest that the C domains do exert a degree of selection and the intermediates detected from offloading by non-cognate amino acids are not an artefact resulting from improved detection. The interception of intermediate **240** from module 4 by only the Hpg-based probe may

indicate a stronger substrate specificity due to the crucial nature of the Hpg residue in the formation of the cross-linked heptapeptide scaffold.

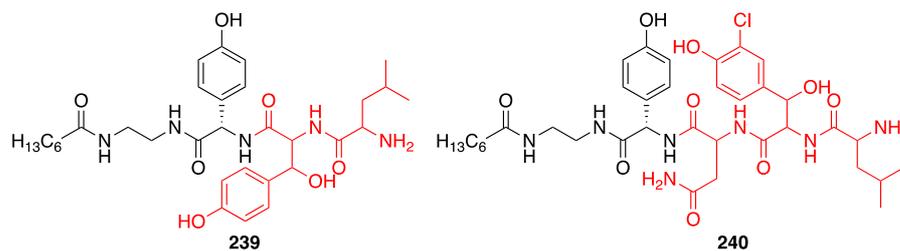


Figure 165 - Unchlorinated peptide offloaded using the Hpg-based probe, 178, and a tripeptide offloaded using the same chemical probe.

The expected offloading from later modules, beyond module four, was not observed. Cognate amino acids of modules five to seven are Hpg, β HHT, and Dhpg respectively. This prevented investigation of the second halogenation event catalysed by VhaA. This could possibly be attributed to a reduced flux of intermediates passing through the later modules compared to the earlier modules, but this is unlikely to account for such a decrease in offloading. As no probes displayed significant toxicity towards *A. orientalis*, feeding with an increased concentration to increase the likelihood of offloading could be pursued to examine late-stage biosynthesis. Additionally, an extended incubation could allow for a greater production of vancomycin and thus provide more opportunities for the interception of intermediates.

The identification of chlorinated dipeptides alongside the unchlorinated dipeptide intermediate **239**, strongly suggests halogenation occurring at enzyme-bound dipeptide level. However, further *in vitro* work would be required to definitively confirm this, using VhaA alongside an T-bound β HHT-Leu-NH₂ dipeptide. The incubation of VhaA with T-bound β HHT-NH₂ could rule out the halogenation at the amino acid level during biosynthesis. Alternatively, a truncation of the first NRPS, VpsA, to remove module one and prevent dipeptide formation could be used to examine amino acid level halogenation *in vivo*.

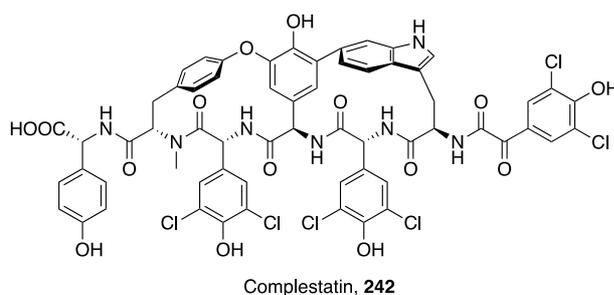
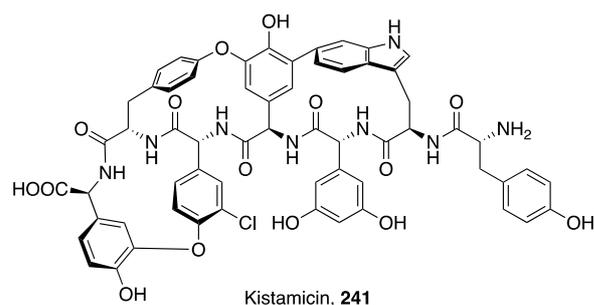


Figure 166 - Structures of kistamicin and complestatin, glycopeptide antibiotics which evolved separately to other GPAs, containing a tryptophan at residue two.

It would also be of interest to investigate the biosynthesis of other glycopeptide antibiotics using chemical probes based on several other cognate amino acids (yet to be prepared in the group). There are already notable differences between teicoplanin-type glycopeptides and vancomycin-type glycopeptides in biosynthesis, such as the timing of tyrosine hydroxylation. Study of the timing of halogenation in these systems to identify the likely substrates would aid in the rational design of novel glycopeptides. Of particular interest would be the type V glycopeptides, such as kistamicin, **241**, and complestatin, **242**, shown in Figure 166. The gene clusters of these compounds evolved independently from other glycopeptide antibiotics. These contain a tryptophan residue in position two and different patterns of chlorination, dichlorination at residues one, three, and five for complestatin, and halogenation of residue five in kistamicin.

6.4 Study of the biosynthesis of scleric acid

In the study of scleric acid biosynthesis described in chapter 5 I was able to unequivocally determine the structure of scleric acid from a silent, cryptic gene cluster from the intractable *Streptomyces sclerotialis*. This was achieved through the synthesis of a standard based on the proposed structure, **196**, as well as the synthesis

of a structural isomer, **201**, which was also possible based on the available spectroscopic data. Confirmation of the structure and predicted functions of the gene cluster allowed the proposal of a biosynthetic pathway. Synthesis of an intermediate, **202**, in this pathway allowed the chemical complementation of a knockout strain and restoration of scleric acid production. This provided evidence that the proposed pathway was correct.

Chemical probes were also used to interrogate the biosynthetic pathway, which contains a single NRPS. The probes were capable of forming an amide bond with glycolic acid as would occur in the proposed pathway, but residual production of glycolic acid meant this species was also detectable in control strains.

Activity of scleric acid against *Mycobacterium tuberculosis* and nicotinamide *N*-methyltransferase (NNMT) was used as a lead to generate a small library of scleric acid derivatives. Six compounds were synthesized and tested against *M. tuberculosis*, which revealed that the only compound with any antibacterial effect with the decanoyl ester of scleric acid, **215**. While the activity was modest with an MIC₉₀ of 125 µg/ml, this would suggest that the masking of the carboxylic acid moiety would be a beneficial starting point for further development. The negative charge of the carboxylate under physiological conditions could hinder cellular entry due to the negative charge present on the surface on the cell membrane. Additionally, the testing of the derivatives against NNMT, particularly the nicotinamide derivative, **217**, to study the activity against this underdeveloped target should also be pursued.

As scleric acid is not produced in either the native or heterologous host strains under laboratory conditions, the effect of scleric acid on *S. sclerotialis* is unknown. Understanding its role in the cell could potentially reveal details on the of method of its observed action against other bacteria. Further characterisation of the biosynthetic enzymes through *in vitro* study would allow the proposed steps to be verified, and the substrate scope to be investigated. Particularly the ATP-grasp enzyme SclG could be envisaged to be a useful biocatalyst.

7. Experimental

7.1 General chemical methods

7.1.1 Chemicals and solvents

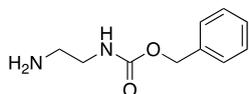
All chemicals were purchased from Sigma-Aldrich, VWR International, Fisher Scientific, Fluorochem or Carbosynth and used without further purification. Dry solvents were purchased from Fisher Scientific or dried using solvent towers. Reagent grade solvents were purchased from Fisher Scientific. Analytical TLC was performed on aluminium sheets precoated with silica gel 60 (F₂₅₄, Merck) and visualised under ultra-violet light (short wave) and using potassium permanganate or ninhydrin stains and heating with a heat gun. Silica gel for column chromatography was purchased from Sigma-Aldrich (Tech grade, pore size 60 Å, 230-400 mesh).

7.1.2 Compound characterisation

Infra-red spectra were recorded on a Bruker Alpha-T FTOR spectrometer using 16 scans. Absorption maxima (ν_{\max}) are quoted in wavenumbers (cm^{-1}). ¹H, ¹³C and ¹⁹F NMR spectra were recorded in *d*₄-MeOD or CDCl₃ unless stated otherwise on the following Bruker Avance instruments: DPX-300 MHz, DPX-400 MHz, DRX-500MHz or AV-600 MHz. Chemical shifts are reported in parts per million (ppm) and coupling constants (*J*) are reported in Hertz (Hz). High-resolution mass spectra (HRMS) of synthesized compounds were obtained using electrospray ionisation (ESI) on a MaXis UHR-TOF (Bruker Daltonics) or on a Bruker MaXis (ESI-HR-MS). Low-resolution mass spectra were recorded on an Agilent 6130B ESI-MS. Optical rotations were obtained using an AA-1000 Polarimeter from Optical Activity Ltd. HPLC purification was performed on an Agilent 1260 with a Phenomex Synergi Polar RP 80 Å (250 x 10.0 mm, 4 µm) column. The mobile phase was a gradient of H₂O and MeCN (HPLC grade, with 0.1 % formic acid) at a flow rate of 2.5 ml/min. UV absorbance was measured at 210 nm, 254 nm, and 280 nm.

7.1.3 Chemical synthesis

7.1.3.1 Synthesis of benzyl (2-aminoethyl)carbamate (**47**)

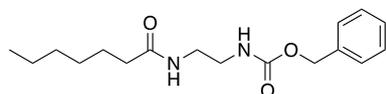


A solution of benzylchloroformate (1.3 mL, 9 mmol) in dry dichloromethane (25.0 mL) was added dropwise over 2 hours to a solution of ethylenediamine (6.0 mL, 90 mmol) in dichloromethane (90 mL) at 0 °C under an argon atmosphere. The reaction was stirred at 0 °C for 2 hours. The reaction mixture was then washed with sat. NaCl solution (3 x 15 mL), dried with MgSO₄ (s), filtered, and concentrated *in vacuo* to afford a white solid **47** (1.64 g, 95 %), which was used without further purification.

¹H NMR (300 MHz, MeOD): δ 7.40 – 7.30 (5H, m, ArH), 5.10 (2H, s, CH₂Ar), 3.24 (2H, dd, 4.9 Hz, 10.5 Hz, NHCH₂), 2.81 (2H, t, 5.5 Hz, CH₂NH₂); LR-MS: calculated for C₁₀H₁₅N₂O₂ [M+H]⁺: 195.2, found: 195.1.

The data are in accordance with those reported in the literature.¹⁵³

7.1.3.2 Synthesis of benzyl (2-heptanamidoethyl)carbamate (**48**)

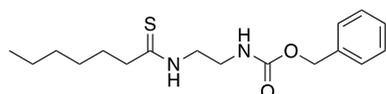


To a solution of (benzyl(2-aminoethyl)carbamate) (1.748 g, 9.0 mmol) in dry dichloromethane (60.00 mL) triethylamine (1.81 mL, 11.7 mmol) was added at 0 °C under an argon atmosphere. Heptanoyl chloride (1.63 mL, 11.7 mmol) was added dropwise, and stirred for a further 30 minutes at 0 °C. The reaction was allowed to return to room temperature overnight. The organic phase was washed with 1M HCl (20 mL), sat. NaCl (aq) (20 mL) and sat. NaHCO₃ (aq) (20 mL). It was then dried over MgSO₄, filtered and concentrated. The crude product was purified by column chromatography with a gradient of cyclohexane: EtOAc (9:1) to cyclohexane: EtOAc (3:7) to afford **48** as a white solid (1.703 g, 62 %). R_f = 0.27 (cyclohexane:EtOAc 1:1); ¹H NMR (300 MHz, CDCl₃): δ 7.37 – 7.31 (5H, m, ArH), 5.98 (1H, br s, CONHCH₂), 5.19 (1H, br s, CO₂NH), 5.10 (2H, s, CH₂Ar), 3.43 – 3.29 (4H, m, NHCH₂CH₂NH),

2.14 (2H, t, 7.6 Hz, COCH₂), 1.64 – 1.54 (2H, m, COCH₂CH₂), 1.33 – 1.24 (6H, m, CH₂), 0.88 (3H, t, 5.8 Hz, CH₂CH₂CH₃); **LR-MS**: calculated for C₁₇H₂₅N₂O₃ [M-H]⁻: 305.2, found: 305.1.

The data are in accordance with those reported in the literature.¹⁵³

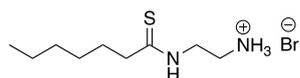
7.1.3.3 Synthesis of benzyl (2-heptanethioamidoethyl)carbamate (**49**)



A solution of benzyl (2-heptanamidoethyl)carbamate (1.200 g, 3.92 mmol) and Lawesson's reagent (1.109 g, 2.74 mmol) in anhydrous tetrahydrofuran (50 mL) under an inert atmosphere, was heated to reflux for 3.5 hours. The solvent was removed *in vacuo* and the residue dissolved in EtOAc, washed with sat. NaHCO₃ (aq) (40 mL), and sat. NaCl (aq) (3 x 40 mL). The organic layer was dried with MgSO₄, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography using EtOAc:cyclohexane (1:3) to afford **49** as a white solid (859 mg, 68 %). **R_f** = 0.35 (3:1 petroleum ether:EtOAc).

¹H NMR (500 MHz, CDCl₃): δ 8.35 (1H, s, CSNHCH₂), 7.32-7.23 (5H, m, ArH), 5.19 (1H, s, CH₂NHCOO), 5.04 (2H, s, CH₂Ar), 3.67 (2H, dd, 4.9 Hz, 10.3 Hz, CSNHCH₂), 3.41 (2H, dd, 5.9 Hz, 11.0 Hz, CH₂CH₂NH), 2.55 (2H, t, 7.7 Hz, CH₂CS), 1.70 – 1.62 (2H, m, CH₂CH₂CS), 1.29-1.16 (6H, m, CH₂), 0.81 (3H, t, 6.8 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, CDCl₃): δ 206.6 (CSNH), 158.3 (NHCO), 136.0 (Ar), 128.6 (Ar), 128.4 (Ar), 128.1 (Ar), 67.3 (CH₂Ar), 48.1 (CSNHCH₂), 47.1 (CSCH₂), 39.5 (CH₂NH), 31.5 (CH₂), 29.1 (CH₂CH₂CS), 28.6 (CH₂), 22.5 (CH₂), 14.1 (CH₃); **HRMS (ESI)**: calculated for C₁₇H₂₆N₂NaO₂S [M+Na]⁺: 345.1607, found: 345.1608.

7.1.3.4 Synthesis of 2-heptanethioamidoethan-1-aminium bromide (**50**)

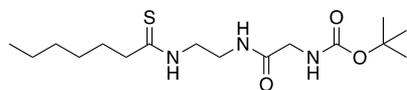


To a flask containing benzyl (2-heptanethioamidoethyl)carbamate (500 mg, 1.55 mmol) under an inert atmosphere, HBr (33% in AcOH, 10 mL) was added and stirred

overnight. The solution was concentrated *in vacuo* and the product triturated with diethyl ether (100 mL) to afford **50** as an orange solid (337 mg, 91 %).

¹H NMR (500 MHz, MeOD): δ 3.94 (2H, t, 6.5 Hz, CSNHCH₂), 3.25 (2H, t, 6.5 Hz, CH₂CH₂NH), 2.66 (2H, t, 7.6 Hz, CH₂CS), 1.80 – 1.72 (2H, m, CH₂CH₂CS), 1.36–1.28 (6H, m, CH₂), 0.89 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, MeOD): δ 208.3 (CSNH), 45.6 (CH₂CS), 42.1 (CSNHCH₂), 37.6 (CH₂CH₂NH₃), 31.2 (CH₂), 29.2 (CH₂CH₂CS), 28.3 (CH₂), 21.8 (CH₂), 12.7 (CH₃); **HRMS (ESI)**: calculated for C₉H₂₁N₂S [M+H]⁺: 189.1420, found: 189.1421.

7.1.3.5 Synthesis of *tert*-butyl (2-((2-heptanethioamidoethyl)amino)-2-oxoethyl)carbamate (**51**)

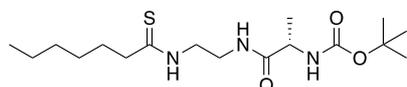


To a solution of *N,N*-diisopropylethylamine (DIPEA, 1.42 mL, 8.13 mmol), 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU, 927 mg, 2.44 mmol), and *N*-Boc glycine (427 mg, 2.44 mmol) in dry dichloromethane at 0 °C *N*-(2-aminoethyl)heptanethioamide (545 mg, 2.03 mmol) was added and the mixture allowed to warm to room temperature overnight. The organic phase was washed with 1M HCl (10 mL), sat. NaCl (aq) (10 mL), and sat. NaHCO₃ (aq) (10 mL). The organic phase was dried with MgSO₄, filtered and concentrated. The crude product was purified by column chromatography with a gradient of EtOAc: cyclohexane (2: 3) to afford **51** as a colourless oil (108 mg, 15 %). **R_f** = 0.18 (1:1 EtOAc: petroleum ether); **IR** (thin film) ν_{\max} : 3280 (N-H stretch), 2956 (C-H stretch), 1654 (C=O stretch), 1162 (C=S stretch);

¹H NMR (500 MHz, CDCl₃): δ 8.77 (1H, s, CSNH), 7.12 (1H, s, CH₂CH₂NH), 5.32 (1H, s, NHCOO), 3.72 (2H, d, 5.9 Hz, COCH₂NH), 3.78 – 3.64 (2H, m, CSNHCH₂), 3.48 (2H, dt, CH₂CH₂NH), 2.57 (2H, t, CH₂CS), 1.57 (2H, quint, CH₂CH₂CS), 1.38 (9H, s, C(CH₃)), 1.22 (6H, m, CH₂), 0.81 (3H, t, CH₃CH₂); **¹³C NMR** (150 MHz, CDCl₃): δ 206.7 (CSNH), 171.6 (NHCOCH₂), 156.2 (COOR), 80.6 (C(CH₃)), 47.0 (CH₂CH₂CS), 46.8 (CSNHCH₂), 44.3 (COCH₂NH), 38.6 (CH₂CH₂NH), 31.5 (CH₂),

29.3 (CH₂CS), 28.8 (CH₂), 28.3 (C(CH₃)), 22.5 (CH₂), 14.0 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₆H₃₁N₃NaO₃S [M+Na]⁺: 368.1978, found: 368.1977.

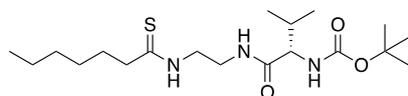
7.1.3.6 Synthesis of *tert*-butyl (S)-1-((2-heptanethioamidoethyl)amino)-1-oxopropan-2-yl)carbamate (**52**)



To a solution of DIPEA (1.42 mL, 8.13 mmol), HATU (927 mg, 2.44 mmol), and *N*-Boc-L-alanine (461 mg, 2.44 mmol) in anhydrous dichloromethane at 0 °C *N*-(2-aminoethyl)heptanethioamide (543 mg, 2.03 mmol) was added and the mixture allowed to return to room temperature overnight. The organic phase was washed with 1M HCl (10 mL), sat. NaCl (aq) (10 mL) and sat. NaHCO₃ (aq) (10 mL). The organic phase was dried with MgSO₄, filtered and concentrated. The crude product was purified by column chromatography with a gradient of EtOAc: cyclohexane (1:1) to afford the product as a thick, colourless oil (586 mg, 80 %). **R_f** = 0.37 (1:1 petroleum ether: EtOAc); **IR** (thin film) ν_{max} : 3278 (N-H stretch), 2956 (C-H stretch), 1655 (C=O stretch), 1161 (C=S stretch);

¹H NMR (500 MHz, CDCl₃): δ 8.56 (1H, s, CSNH), 6.72 (1H, s, CH₂NHCO), 4.84 (1H, s, NHCOO), 4.13 – 4.08 (1H, m, COCH(CH₃)NH), 3.86 – 3.66 (2H, m, CSNHCH₂), 3.59 – 3.49 (2H, m, CH₂CH₂NH), 2.64 (2H, t, 7.6 Hz, CH₂CH₂CS), 1.79 – 1.70 (2H, m, CH₂CH₂CS), 1.45 (9H, s, C(CH₃)₃), 1.36 (3H, d, 7.1 Hz, COCH(CH₃)), 0.88 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, CDCl₃): δ 206.5 (CSNH), 175.0 (NHCOCH(CH₃)), 155.6 (NHCOO), 80.5 (C(CH₃)₃), 50.4 (COCH(CH₃)), 47.0 (CH₂CH₂CS), 47.0 (CSNHCH₂), 38.4 (NHCH₂CH₂NH), 31.5 (CH₂), 29.3 (CH₂CH₂CS), 28.6 (CH₂), 28.3 (C(CH₃)₃), 22.5 (CH₂), 18.4 (CH(CH₃)), 14.0 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₇H₃₃N₃NaO₃S [M+Na]⁺: 382.2135, found: 382.2138; **[α]_D²⁸**: +6.0 (0.050, MeOH);

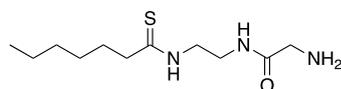
7.1.3.7 Synthesis of *tert*-butyl (S)-1-((2-heptanethioamidoethyl)amino)-3-methyl-1-oxobutan-2-yl)carbamate (53)



To a solution of DIPEA (1.31 mL, 7.52 mmol), HATU (859 mg, 2.26 mmol), and *N*-Boc-L-valine (490 mg, 2.26 mmol) in anhydrous dichloromethane at 0 °C, *N*-(2-aminoethyl)heptanethioamide (503 mg, 1.88 mmol) was added and the mixture allowed to return to room temperature overnight. The organic phase was washed with 1M HCl (10 mL), sat. NaCl_(aq) (10 mL) and sat. NaHCO_{3(aq)} (10 mL). The organic phase was dried with MgSO₄, filtered and concentrated. The crude product was purified by column chromatography with a gradient of EtOAc: cyclohexane (1:1) to afford the product as a white solid (572 mg, 78 %). **R_f** = 0.09 (3:1 petroleum ether: EtOAc), **IR** (thin film) ν_{max} : 3341 (N-H stretch), 2957 (C-H stretch), 1642 (C=O stretch), 1169 (C=S stretch);

¹H NMR (500 MHz, CDCl₃): δ 8.86 (1H, s, CSNH), 7.30 (1H, s, CH₂NHCO), 5.20 (1H, d, 7.1 Hz, NHCOO), 3.88 (1H, dd, 6.3 Hz, 8.0 Hz, COCH(CH(CH₃)₂)NH), 3.76–3.63 (2H, m, CSNHCH₂), 3.50 – 3.43 (2H, m, NHCH₂CH₂NH), 2.56 (2H, t, 7.7 Hz, CH₂CS), 2.07 – 1.97 (1H, m, CH(CH₃)₂), 1.74 (2H, quin, 7.9 Hz, CH₂CH₂CS), 1.45 (9H, s, C(CH₃)₃), 1.35 – 1.25 (6H, m, CH₂), 0.96 (3H, d, 6.8 Hz, CH(CH₃)₂), 0.91 (3H, d, 6.8 Hz, CH(CH₃)₂), 0.87 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, CDCl₃): δ 205.5 (CSNH), 173.1 (NHCO), 155.1 (NHCOO), 79.2 (C(CH₃)₃), 59.1 (CH(CH(CH₃)₂)), 46.0 (CSNHCH₂), 45.9 (CH₂CH₂CS), 37.4 (NHCH₂CH₂NH), 30.5 (CH₂), 29.8 (CH(CH₃)₂), 28.3 (CH₂CH₂CS), 27.6 (CH₂), 27.3 (C(CH₃)₃), 21.5 (CH₂), 18.3 (CH(CH₃)₂), 17.0 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₉H₃₇N₃NaO₃S [M+Na]⁺: 410.2448, found: 410.2449; [α]_D²⁸: +1.4 (0.070, MeOH).

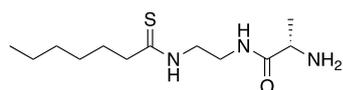
7.1.3.8 Synthesis of 2-amino-*N*-(2-heptanethioamidoethyl)acetamide (54)



To a solution of *tert*-butyl (2-((2-heptanethioamidoethyl)amino)-2-oxoethyl)carbamate (49 mg, 0.14 mmol) in dry dichloromethane (3 mL), TFA (0.11 mL, 1.4 mmol) was added dropwise, and stirred at room temperature for 3.5 hours. TFA and DCM were removed *in vacuo*, and the residue dissolved in MeOH. Amberlyst A-21 resin was prepared by washing commercially available resin with MeOH (10 mL/g), THF (10 mL/g), and DCM (10 mL/g) before drying under a high vacuum. Amberlyst A-21 resin (840 mg) was added and the solution was stirred for 2 days. The resin was removed by filtration and the filtrate concentrated *in vacuo* to afford **54** as a pale yellow solid (19 mg, 55%). $R_f = 0.30$ (1:1 DCM: MeOH); **IR** (thin film) ν_{\max} : 3471 (N-H stretch), 2926 (C-H stretch), 1641 (C=O stretch), 1174 (C=S stretch), 1112 (C-N stretch);

¹H NMR (500 MHz, MeOD): δ 3.78 (2H, t, 5.4 Hz, CSNH₂CH₂), 3.67 (2H, s, COCH₂NH₃), 3.52 (2H, t, 5.4 Hz, CH₂CH₂NH), 2.62 (2H, t, 7.3 Hz, CH₂CH₂CS), 1.81 – 1.68 (2H, m, CH₂CH₂CS), 1.41 – 1.29 (6H, m, CH₂), 0.92 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, MeOD): δ 206.7 (CSNH), 166.3 (NHCO), 45.8 (CH₂CH₂CS), 44.4 (CSNHCH₂), 40.1 (COCH₂NH₃), 37.4 (CH₂CH₂NH), 31.3 (CH₂), 29.3 (CH₂CH₂CS), 28.3 (CH₂), 22.2 (CH₂), 13.0 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₁H₂₄N₃OS [M+H]⁺: 246.1635, found: 246.1637.

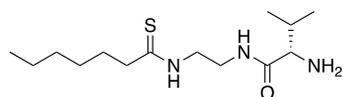
7.1.3.9 Synthesis of (S)-2-amino-N-(2-heptanethioamidoethyl)propenamide (**55**)



To a solution of *tert*-butyl (S)-1-((2-heptanethioamidoethyl)amino)-1-oxopropan-2-yl)carbamate (59 mg, 0.16 mmol) in anhydrous dichloromethane (3.00 mL), TFA (0.13 mL, 1.6 mmol) was added dropwise, and stirred at room temperature for 4 hours. The residue was dissolved in MeOH, and Amberlyst A-21 resin (520 mg) was added and stirred for 2 days. The resin was removed by filtration and the filtrate concentrated *in vacuo* to afford **55** as a white solid (22 mg, 54%). $R_f = 0.46$ (1:1 DCM: MeOH), **IR** (thin film) ν_{\max} : 3236 (N-H stretch), 2925 (C-H stretch), 1641 (C=O stretch), 1174 (C=S stretch), 1089 (C-N stretch);

¹H NMR (500 MHz, CDCl₃): δ 3.79 (2H, td, 1.7 Hz, 6.1 Hz, 5.9 Hz, CSNHCH₂), 3.51 (2H, t, 6.2 Hz, CH₂CH₂NH), 3.40 (1H, q, 6.9 Hz, COCH(CH₃)) 2.62 (2H, t, 7.8 Hz, CH₂CS), 1.79 – 1.70 (2H, m, CH₂CH₂CS), 1.40 – 1.31 (6H, m, CH₂), 1.28 (3H, d, 6.9 Hz, COCH(CH₃)), 0.92 (3H, t, 6.8 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, CDCl₃): δ 206.7 (CSNH), 169.9 (CONH), 48.9 (COCH(CH₃)), 45.8 (CH₂CS), 44.6 (CSNHCH₂), 37.5 (CH₂CH₂NH), 31.3 (CH₂), 29.3 (CH₂CH₂CS), 28.3 (CH₂), 22.2 (CH₂), 16.2 (COCH(CH₃)), 13.0 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₂H₂₅N₃NaOS [M+Na]⁺: 282.1611, found: 282.1613; [α]_D²⁸: -40.9 (0.036, MeOH).

7.1.3.10 Synthesis of (S)-2-amino-N-(2-heptanethioamidoethyl)-3-methylbutanamide (56)

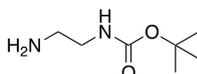


To a solution of *tert*-butyl (S)-1-((2-heptanethioamidoethyl)amino)-3-methyl-1-oxobutan-2-yl)carbamate (21 mg, 0.05 mmol) in dichloromethane (5 mL) was added Amberlyst 15 resin (160 mg) (Prepared by soaking commercially available resin in MeOH (5 mL/g) for 24 hours, before washing with MeOH and neutralising with 7N ammonia in MeOH. The neutralised resin was acidified with 3M HCl in 50% MeOH and washed with MeOH, THF, and DCM before drying under a high vacuum) and the solution was stirred overnight. The resin was removed by filtration, and washed with petroleum ether, EtOAc and MeOH. The resin was mixed with 7N NH₃ in MeOH (3 mL) for 1 hour. The resin was removed by filtration, and the filtrate concentrated *in vacuo* to afford **56** as a white solid (14 mg, 91 %). **R_f**: 0.06 (1:4 MeOH: DCM), **IR** (thin film) ν_{max}: 3281 (N-H stretch), 2928 (C-H stretch), 1652 (C=O stretch), 1161 (C=S stretch);

¹H NMR (500 MHz, CDCl₃): δ 3.77 (2H, t, 6.4 Hz, CSNHCH₂), 3.49 (2H, t, 6.2 Hz, CH₂CH₂NH), 3.08 (1H, d, 5.7 Hz, COCH(CH(CH₃)₂)), 2.65 – 2.59 (2H, m, CH₂CS), 1.96 (1H, td, 6.9 Hz, 13.6 Hz, CH(CH(CH₃)₂)), 1.80 – 1.71 (2H, m, CH₂CH₂CS), 1.42 – 1.30 (6H, m, CH₂), 0.98 (3H, d, 6.8 Hz, CH(CH₃)₂), 0.96 – 0.90 (6H, m, CH₃CH₂, CH(CH₃)₂); **¹³C NMR** (150 MHz, CDCl₃): δ 206.6 (CSNH), 168.6 (NHCO), 58.5 (COCH(CH(CH₃)₂)), 45.8 (CH₂CS), 44.5 (CSNHCH₂), 37.3 (CH₂CH₂NH), 31.3

(CH₂), 30.0 (CH(CH₃)₂), 29.3 (CH₂CH₂CS), 28.3 (CH₂), 22.2 (CH₂), 17.5 (CH(CH₃)₂), 16.5 (CH(CH₃)₂), 13.0 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₄H₃₀N₃OS [M+H]⁺: 288.2104, found: 288.2107; [α]_D²⁸: +2.2 (0.069, MeOH).

7.1.3.11 Synthesis of *tert*-butyl (2-aminoethyl)carbamate (**44**)

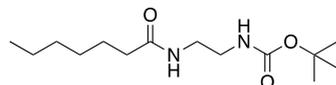


Di-*tert*-butyl dicarbonate (Boc₂O, 5.74 mL, 25 mmol) was dissolved in CHCl₃ (150.0 mL) and added dropwise to ethylene diamine (16.7 mL, 250 mmol) in CHCl₃ (100.0 mL) and stirred at 0 °C for 3 hours, then for 14 hours at room temperature. The organic phase was washed with H₂O (4 x 50 mL), dried with MgSO₄ (s), filtered and concentrated *in vacuo* to afford *tert*-butyl (2-aminoethyl)carbamate as a pale brown oil (3.81 g, 95 %).

¹H NMR (300 MHz, CDCl₃): δ 5.11 (1H, br s, NHCOO), 3.10 (2H, q, 6.9 Hz, CONHCH₂), 2.73 (2H, t, 6.3 Hz, CH₂NH₂), 1.58 (2H, s, CH₂NH₂), 1.38 (9H, s, C(CH₃)₃); **LRMS(ESI)**: calculated for [M+H]⁺: 160.1, found: 160.2.

The data are in accordance with those reported in the literature.³¹⁹

7.1.3.12 Synthesis of *tert*-butyl (2-heptanamidoethyl)carbamate (**45**)

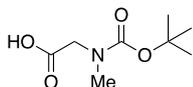


tert-Butyl (2-aminoethyl)carbamate (1.00 g, 6.25 mmol) under argon was dissolved in anhydrous DCM; triethylamine (Et₃N, 0.98 mL, 6.87 mmol) was added and the solution cooled to 0 °C. Heptanoyl chloride (1.06 mL, 6.87 mmol) was added dropwise, and stirred at room temperature for 16 hours. The organic layer was washed with 0.2M HCl (20 mL), sat. NaHCO₃ (20 mL) and brine (20 mL), dried with MgSO₄, filtered and concentrated *in vacuo* to afford **45** as a fluffy white solid (1.64 g, 97 %).

¹H NMR (400 MHz, CDCl₃): δ 6.14 (1H, s, CONH), 4.91 (1H, s, NHCOO), 3.39 – 3.22 (4H, m, NHCH₂CH₂NH), 2.16 (2H, t, 7.7 Hz, CH₂CO), 1.44 (9H, s, C(CH₃)₃), 1.31 – 1.25 (6H, m, CH₂), 0.88 (3H, t, 6.7 Hz, CH₃CH₂); **LRMS (ESI)**: calculated for [M+H]⁺: 272.2, found: 272.1.

The data are in accordance with that reported in the literature.³²⁰

7.1.3.13 Synthesis of *N*-(*tert*-butoxycarbonyl)-*N*-methylglycine (**60**)



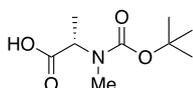
Sarcosine (445 mg, 5 mmol) was dissolved in H₂O (50 mL) with Et₃N (2.09 mL, 15 mmol) and di-*tert*-butyl dicarbonate (1.091 g, 5 mmol) and stirred at room temperature for 3 hours. The reaction mixture was acidified to pH 3 with 1M HCl and extracted with EtOAc (3 x 10 mL). The organic phase was washed with H₂O (10 mL), dried with MgSO₄(s), filtered, and concentrated to afford **60** as a white solid (310 mg, 33%).

¹H NMR (400 MHz, MeOD, a 1:1 mixture of E: Z isomers): δ 4.02 (s) and 3.98 (s) (together COCH₂), 2.93 (3H, s, CONCH₃), 1.46 (s) and 1.43 (s) (together C(CH₃)₃).

LRMS (ESI): calculated for C₈H₁₄NO₄ [M-H]⁻: 188.1, found: 188.1.

The data are in accordance with those reported in the literature.³²¹

7.1.3.14 Synthesis of *N*-(*tert*-butoxycarbonyl)-*N*-methyl-L-alanine (**61**)



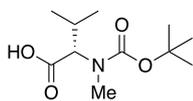
N-Boc-L-alanine (757 mg, 4 mmol) and methyl iodide (MeI, 2.49 mL, 40 mmol) were dissolved in anhydrous THF (15 mL) under an inert atmosphere, and cooled to 0 °C. NaH (60% in oil, 1.60 g, 40 mmol) was added and stirred for 1.5 hours, after which the reaction was allowed to return to room temperature overnight. The reaction mixture was quenched with H₂O (5 mL) and THF removed *in vacuo*. The aqueous layer was diluted with H₂O (5 mL) and washed with EtOAc (10 mL). The aqueous layer was acidified to pH 3 with 1M HCl and extracted with EtOAc (3 x 15 mL). The organic layer was dried with MgSO₄(s), filtered and concentrated. The crude product was purified by column chromatography, using a gradient of EtOAc to EtOAc: MeOH (19:1) to afford **61** as a white powder (579 mg, 71%).

¹H NMR (400 MHz, MeOD a 1:1 mixture of E: Z isomers): δ 4.76 – 4.63 (m) and 4.45 – 4.34 (m) (together 1H, CHCOOH), 2.85 (3H, s, NCH₃), 1.45 (9H, s, C(CH₃)₃),

1.42 – 1.37 (3H, m, CHCH₃); **LRMS (ESI)**: calculated for C₉H₁₆NO₄ [M-H]⁻: 202.1, found: 202.2.

The data are in accordance with those reported in the literature.³²²

7.1.3.15 Synthesis of *N*-(*tert*-butoxycarbonyl)-*N*-methyl-L-valine (**62**)



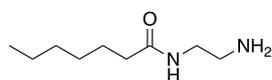
N-Boc-L-valine (757 mg, 3.48 mmol) was dissolved in anhydrous THF (10 mL) under an inert atmosphere, and cooled to 0 °C, and MeI (2.49 mL, 40 mmol) was added. NaH (60% in oil, 1.60 g, 40 mmol) and stirred at 0 °C for 1.5 hours and allowed to return to room temperature overnight. The reaction was quenched with H₂O (1.2 mL), and THF removed *in vacuo*. The aqueous layer was diluted with H₂O (20 mL) and washed with EtOAc (10 mL). The aqueous layer was acidified to pH 3 with 1M HCl and extracted with EtOAc (3 x 15 mL). The organic layer was dried with MgSO₄ (s), filtered, and concentrated to give a yellow oil. The crude product was purified by column chromatography, using a gradient of EtOAc to EtOAc: MeOH (19:1) to afford **62** as a thick colourless oil (791 mg, 86%).

¹H NMR (400 MHz, MeOD, a 1:1 mixture of E: Z isomers): δ 4.29 (d, 10.3 Hz) and 4.04 (d, 10.1 Hz) (together 1H, HO₂CCH), 2.28 – 2.10 (1H, m, CHCH(CH₃)₂), 1.46 (9H, s, C(CH₃)₃), 1.02 (3H, d, 6.6 Hz, CH(CH₃)₂) 0.89 (3H, t, 5.7 Hz, CH(CH₃)₂).

LRMS (ESI): calculated for C₁₁H₂₀NO₄ [M-H]⁻: 230.1, found: 230.2.

The data are in accordance with those reported in the literature.³²²

7.1.3.16 Synthesis of *N*-(2-aminoethyl)heptanamide (**63**)



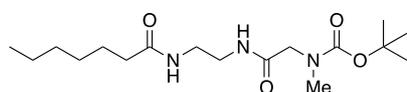
To benzyl (2-heptanamidoethyl)carbamate (600 mg, 1.96 mmol) and Pd/C (10%) (167 mg, 1.57 mmol) under an argon atmosphere, anhydrous MeOH (20 ml) was added. The mixture was degassed by bubbling argon gas through, before bubbling hydrogen gas through the solution for 3 hours. The reaction mixture was filtered

through Celite and concentrated *in vacuo* to afford *N*-(2-aminoethyl)heptanamide (338 mg, quant.) as a white solid.

¹H NMR (300 MHz, MeOD): δ 3.27 (2H, t, 6.2 Hz, CONHCH₂), 2.77 (2H, t, 5.8 Hz, CH₂NH₂), 2.21 (2H, t, 7.3 Hz, CH₂CO), 1.68 – 1.57 (2H, m, CH₂CH₂CO), 1.40 – 1.28 (6H, m, CH₂), 0.92 (3H, t, 5.2 Hz, CH₃CH₂); **LRMS (ESI)**: calculated for C₉H₂₁N₂O [M+H]⁺: 173.2, found: 173.1.

The data are in accordance with those reported in the literature.¹⁵³

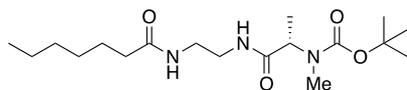
7.1.3.17 Synthesis of *tert*-butyl (2-((2-heptanamidoethyl)amino)-2-oxoethyl)(methyl)carbamate (**64**)



To a solution of *N*-(2-aminoethyl)heptanamide (232 mg, 1.35 mmol), Boc-*N*-methyl glycine (280 mg, 1.48 mmol) and DIPEA (0.52 mL, 2.97 mmol) in anhydrous THF (10 mL) at 0 °C under an inert atmosphere, was added HATU (677 mg, 1.78 mmol). The reaction was warmed and stirred at room temperature overnight, then the solvent was removed *in vacuo*. The residue was dissolved in DCM (20 mL), and the organic layer washed with 1M HCl (20 mL), sat. NaCl (aq) (20 mL) and sat. NaHCO₃ (aq) (20 mL). The organic phase was dried with MgSO₄ (s), filtered and concentrated *in vacuo*. The crude product was purified by column chromatography, eluting isocratically with EtOAc: MeOH (19:1) to afford **64** as a beige solid (267 mg, 58%). **R_f**: 0.26 (1:24 MeOH: EtOAc); **IR** (thin film) ν_{\max} : 3283 (N-H stretch), 2926 (C-H stretch), 1697 (C=O stretch), 1662 (C=O stretch), 1147 (C-N stretch);

¹H NMR (500 MHz, CDCl₃): δ 6.66 (1H, s, NH), 6.18 (1H, s, NH), 3.84 (2H, s, COCH₂N), 3.43 – 3.35 (4H, m, NHCH₂CH₂NH), 2.93 (3H, s, NCH₃), 2.16 (2H, t, 7.8 Hz, CH₂CO), 1.60 (2H, quin, 7.2 Hz, CH₂CH₂CO), 1.46 (9H, s, C(CH₃)₃), 1.34 – 1.24 (6H, m, CH₂), 0.87 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, CDCl₃): δ 174.2 (CONH), 170.6 (CONH), 156.5 (NCOO), 80.8 (C(CH₃)₃), 53.1 (COCH₂NH), 39.9 (NHCH₂CH₂NH), 36.7 (CH₂CH₂CO), 31.5 (CH₂), 28.9 (CH₂), 28.33 (C(CH₃)₃), 25.6 (CH₂CH₂CO), 22.5 (CH₂), 14.0 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₇H₃₃N₃NaO₄ [M+Na]⁺: 366.2365, found: 366.2363.

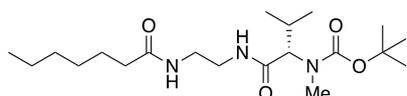
7.1.3.18 Synthesis of *tert*-butyl (S)-(1-((2-heptanamidoethyl)amino)-1-oxopropan-2-yl)(methyl)carbamate (65)



To a solution of *N*-(2-aminoethyl)heptanamide (169 mg, 0.98 mmol), Boc-*N*-methyl alanine (219 mg, 1.08 mmol), and DIPEA (0.38 mL, 2.16 mmol) in anhydrous THF (10 mL) at 0 °C under an inert atmosphere, was added HATU (423 mg, 1.27 mmol). The reaction was allowed to return to room temperature overnight, then the solvent was removed *in vacuo*. The residue was dissolved in DCM (20 mL), and the organic layer washed with 1M HCl (15 mL), sat. NaCl_(aq) (15 mL) and sat. NaHCO_{3(aq)} (15 mL). The organic phase was dried with MgSO_{4(s)}, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography, eluting with a gradient of EtOAc: MeOH (32:1) to (19:1) to afford **65** as a colourless oil (155 mg, 44%). **R_f**: 0.48 (1:19 MeOH: EtOAc); **IR** (thin film) ν_{\max} : 3306 (N-H stretch), 2928 (C-H stretch), 1690 (C=O stretch), 1644 (C=O stretch), 1149 (C-N stretch);

¹H NMR (500 MHz, CDCl₃): δ 6.67 (1H, s, NHCOCH), 6.31 (1H, s, CH₂CONH), 4.61 (1H, br s, NHCOO), 3.41 – 3.30 (4H, m, NHCH₂CH₂NH), 2.16 – 2.12 (2H, m, CH₂CO), 2.04 (1H, t, 7.7 Hz, COCHCH₃), 1.62 – 1.52 (2H, m, CH₂CH₂CO), 1.46 (9H, s, C(CH₃)₃), 1.33 (3H, d, 7.2 Hz, COCHCH₃), 1.31 – 1.22 (6H, m, CH₂), 0.86 (3H, t, 6.8 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, CDCl₃): δ 174.2 (CH₂CONH), 173.1 (NHCOCH), 80.8 (C(CH₃)₃), 40.2 (NHCH₂CH₂NH), 39.9 (NHCH₂CH₂NH), 36.8 (CH₂CO), 31.6 (CH₂), 30.4 (COCHNH), 29.1 (CH₂), 28.5 (C(CH₃)₃), 25.8 (CH₂CH₂CO), 22.6 (CH₂), 14.1 (CH₃CH₂), 13.9 (CH₃CH); **HRMS (ESI)**: calculated for C₁₈H₃₅N₃O₄Na [M+Na]⁺: 380.2520, found: 380.2510; [α]_D²⁸: -20.8 (0.060, MeOH).

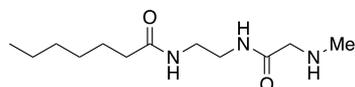
7.1.3.19 Synthesis of *tert*-butyl (S)-(1-((2-heptanamidoethyl)amino)-3-methyl-1-oxobutan-2-yl)(methyl)carbamate (66)



To a solution of *N*-(2-aminoethyl)heptanamide (169 mg, 0.98 mmol), Boc-*N*-methyl-L-alanine (250 mg, 1.08 mmol), and DIPEA (0.38 mL, 2.16 mmol) in anhydrous THF (10 mL) at 0 °C under an inert atmosphere, was added HATU (423 mg, 1.27 mmol). The reaction was warmed and stirred at room temperature overnight, then the solvent removed *in vacuo*. The residue was dissolved in DCM (20 mL), and the organic layer washed with 1M HCl (15 mL), sat. NaCl (aq) (15 mL) and sat. NaHCO₃ (aq) (15 mL). The organic phase was dried with MgSO₄ (s), filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography, eluting isocratically with pure EtOAc to afford **66** as a white solid (240 mg, 63%). **R_f**: 0.47 (EtOAc), **IR** (thin film) ν_{\max} : 3295 (N-H stretch), 2928 (C-H stretch), 1696 (C=O stretch), 1641 (C=O stretch), 1148 (C-N stretch);

¹H NMR (500 MHz, CDCl₃): δ 6.63 (1H, s, CH₂CONH), 6.30 (1H, s, NHCOCH), 3.98 (1H, d, 10.9 Hz COCHNH), 3.40 – 3.31 (4H, m, NHCH₂CH₂NH), 2.31 – 2.19 (1H, m, CHCH(CH₃)₂), 2.13 (2H, t, 7.7 Hz, CH₂CO), 1.63 – 1.53 (2H, m, CH₂CH₂CO), 1.45 (9H, s, C(CH₃)₃), 1.32 – 1.23 (6H, m, CH₂), 0.92 (3H, d, 6.5 Hz, CH(CH₃)₂), 0.89 – 0.84 (6H, CH₃CH₂, CH(CH₃)₂); **¹³C NMR** (150 MHz, CDCl₃): δ 173.8 (CH₂CONH), 172.4 (NHCOCH), 157.2 (NHCOO), 80.6 (C(CH₃)₃), 65.0 (COCHNH), 41.0 (NHCH₂CH₂NH), 39.0 (NHCH₂CH₂NH), 36.9 (CH₂CO), 31.6 (CH₂), 29.1 (CH₂), 28.5 (C(CH₃)₃), 26.2 (COCHCH(CH₃)₂), 25.8 (CH₂CH₂CO), 22.6 (CH₂), 19.7 (CH(CH₃)₂), 18.7 (CH(CH₃)₂), 14.2 (CH₃CH₂); **HRMS (ESI)**: calculated for C₂₀H₃₉N₃O₄Na [M+Na]⁺: 408.2833, found: 408.2830; [α]_D²⁸: -69.3 (0.070, MeOH).

7.1.3.20 Synthesis of *N*-(2-(2-(methylamino)acetamido)ethyl)heptanamide (**67**)

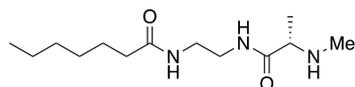


tert-Butyl (2-((2-heptanamidoethyl)amino)-2-oxoethyl)(methyl)carbamate (30 mg, 0.09 mmol) was dissolved in DCM (3 mL), and Amberlyst 15 resin (249 mg) added. After 24 hours, when TLC confirmed absence of starting material, the resin was removed by filtration, and washed with petroleum ether, EtOAc and MeOH. The resin

was then stirred with 7M NH₃ in MeOH (3 mL) for 1 hour. The resin was then removed by filtration, and the filtrate concentrated to afford **67** as a white solid (20 mg, 95%). **R_f**: 0.25 (1:1 MeOH: DCM), **IR** (thin film) ν_{\max} : 3293 (N-H stretch), 2925 (C-H stretch), 1637 (C=O stretch), 1242 (C-N stretch);

¹H NMR (500 MHz, MeOD): δ 3.36 – 3.29 (4H, m, NHCH₂CH₂NH), 3.21 (2H, s, COCH₂NH), 2.39 (3H, s, NHCH₃), 2.19 (2H, t, 7.6 Hz, CH₂CO), 1.61 (2H, quin, 7.1 Hz, CH₂CH₂CO), 1.38 – 1.30 (6H, m, CH₂), 0.93 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, MeOD): δ 175.3 (CONH), 172.5 (CONH), 53.4 (COCH₂NH), 38.5 (NHCH₂CH₂NH), 35.8 (CH₂CO), 34.7 (CH₃NH), 31.3 (CH₂), 28.7 (CH₂), 25.6 (CH₂CH₂CO), 22.2 (CH₂), 13.0 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₂H₂₆N₃O₂ [M+H]⁺: 244.2020, found: 244.2016.

7.1.3.21 Synthesis of (S)-N-(2-(2-(methylamino)propanamido)ethyl)heptanamide (**68**)

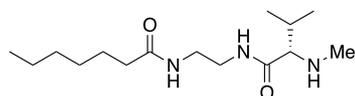


tert-Butyl (S)-1-((2-heptanamidoethyl)amino)-1-oxopropan-2-yl(methyl)carbamate (30 mg, 0.08 mmol) was dissolved in anhydrous DCM (3 mL), and Amberlyst 15 resin (240 mg) added. Stirred at room temperature for 30 hours. The resin was removed by filtration and washed with petroleum ether, EtOAc and finally MeOH, then stirred for 1 hour in 7M NH₃ in MeOH (3 mL). The resin was removed by filtration, and the filtrate concentrated to afford **68** as a white solid (20 mg, 90%). **R_f**: 0.37 (1:1 MeOH:DCM); **IR** (thin film) ν_{\max} : 3291 (N-H stretch), 2925 (C-H stretch), 1641 (C=O stretch), 1243 (C-N stretch);

¹H NMR (500 MHz, MeOD): δ 3.36 – 3.29 (4H, m, NHCH₂CH₂), 3.08 (1H, q, 6.9 Hz, CH(CH₃)), 2.32 (3H, s, NHCH₃), 2.19 (2H, t, 7.6 Hz, CH₂CO), 1.61 (2H, quin, 7.3 Hz, CH₂CH₂CO), 1.39 – 1.30 (6H, m, CH₂), 1.25 (3H, d, 6.9 Hz, CH(CH₃)), 0.92 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, CDCl₃): δ 176.1 (COC(CH₃)), 175.3 (CH₂CO), 59.0 (CH(CH₃)), 38.6 (NHCH₂CH₂), 35.8 (CH₂CO), 33.0 (NHCH₃), 31.3 (CH₂), 28.6 (CH₂), 25.5 (CH₂CH₂CO), 22.1 (CH₂), 17.8 (CH(CH₃)), 13.0 (CH₃CH₂);

HRMS (ESI): calculated for $C_{13}H_{27}N_3NaO_2$ $[M+Na]^+$: 280.1995, found: 280.1999; $[\alpha]_D^{28}$: -13.4 (0.056, MeOH).

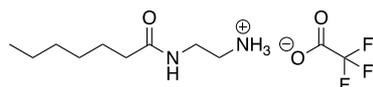
7.1.3.22 Synthesis of (S)-N-(2-(3-methyl-2-(methylamino)butanamido)ethyl)heptanamide (69)



tert-Butyl (S)-(1-((2-heptanamidoethyl)amino)-3-methyl-1-oxobutan-2-yl)(methyl)carbamate (30 mg, 0.08 mmol) was dissolved in DCM (3 mL), and Amberlyst 15 resin (223 mg) added. Stirred at room temperature for 30 hours. The resin was removed by filtration and washed with petroleum ether, EtOAc and MeOH, then stirred for 1 hour in 7M NH_3 in MeOH (3 mL). The resin was removed by filtration, and the filtrate concentrated to afford **69** as a white solid (22 mg, 100%). R_f = 0.63 (1:1 DCM: MeOH); **IR** (thin film) ν_{max} : 3289 (N-H stretch), 2927 (C-H stretch), 1640 (C=O stretch), 1237 (C-N stretch);

1H NMR (500 MHz, $CDCl_3$): δ 3.38 – 3.31 (4H, m, $NHCH_2CH_2$), 2.73 (1H, d, 6.0 Hz, $COCHNHCH_3$), 2.31 (3H, s, $NHCH_3$), 2.19 (2H, t, 7.6 Hz, CH_2CO), 1.90 (1H, dq, 6.8 Hz, 13.4 Hz, $CH(CH_3)_2$), 1.61 (2H, quin, 7.4 Hz, CH_2CH_2CO), 1.39 – 1.30 (6H, m, CH_2), 0.96 (6H, dd, 3.3 Hz, 6.8 Hz, $CH(CH_3)_2$), 0.93 (3H, t, 6.8 Hz, CH_3CH_2); **^{13}C NMR** (150 MHz, $CDCl_3$): δ 175.3 (CH_2CO), 175.0 (CO), 70.2 (COCH), 38.7 ($NHCH_2CH_2$), 38.3 ($NHCH_2CH_2$), 35.8 (CH_2CO), 33.9 ($NHCH_3$), 31.1 (CH_2), 31.1 ($CH(CH_3)_2$), 28.6 (CH_2), 25.5 (CH_2CH_2CO), 22.2 (CH_2), 18.2 ($CH(CH_3)_2$), 17.8 ($CH(CH_3)_2$), 13.0 (CH_3CH_2); **HRMS (ESI):** calculated for $C_{15}H_{31}N_3NaO_2$ $[M+Na]^+$: 308.2308, found: 308.2307; $[\alpha]_D^{28}$: +6.1 (0.074, MeOH).

7.1.3.23 Synthesis of 2-heptanamidoethan-1-aminium trifluoroacetate (73)

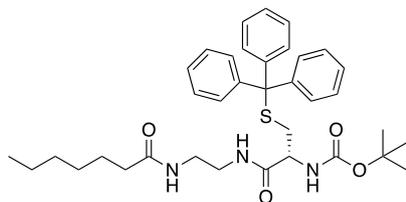


tert-Butyl (2-heptanamidoethyl)carbamate (1.02 g, 3.74 mmol) under an argon atmosphere was dissolved in anhydrous DCM (20 mL) and trifluoroacetic acid (TFA)

(2.87 mL, 37.40 mmol) added slowly. The mixture was stirred at room temperature for 2 hours. DCM and TFA were removed *in vacuo* to afford 2-heptanamidoethan-1-aminium trifluoroacetate as a brown oil (1.07 g, 100 %).

¹H NMR (300 MHz, MeOD): δ 3.43 (2H, t, 5.6 Hz, CONHCH₂), 3.03 (2H, t, 5.3 Hz, CH₂CH₂NH₃), 2.31 – 2.19 (2H, m, CH₂CO) 1.67 – 1.51 (2H, m, CH₂CH₂CO), 1.31 – 1.25 (6H, m, CH₂), 0.95 – 0.87 (3H, m, CH₃CH₂); **LRMS (ESI)**: calculated for C₁₈H₄₁N₄O₂ [2M+H]⁺: 345.3, found: 345.9.

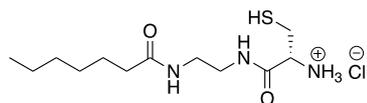
7.1.3.24 Synthesis of *tert*-butyl (*R*)-(1-((2-heptanamidoethyl)amino)-1-oxo-3-(tritylthio)propan-2-yl)carbamate (**74**)



2-Heptanamidoethan-1-aminium 2,2,2-trifluoroacetate (393 mg, 1.37 mmol), *N*-Boc-L-Cys(Trt)-OH (700 mg, 1.51 mmol) and DIPEA (0.52 mL, 3.01 mmol) were dissolved in dry THF (15 mL) and cooled to 0 °C. HATU (678 mg, 1.78 mmol) was added, and the reaction allowed to return to room temperature overnight. The solvent was removed *in vacuo* and the residue dissolved in DCM (25 mL). The organic layer was washed with 1M HCl (10 ml), sat. NaCl (aq) (10 mL) and sat. NaHCO₃ (aq) (10 mL). The organic layer was dried with MgSO₄ (s), filtered, and concentrated. The crude product was purified by column chromatography using petroleum ether: EtOAc (2:3) to afford **74** as a colourless oil (378 mg, 45%). **R_f**: 0.31 (1:1 EtOAc: petroleum ether); **IR** (thin film) ν_{max} : 2927 (C-H stretch), 1650 (C=O stretch), 698 (Ar C-H bend);

¹H NMR (500 MHz, CDCl₃): δ 7.45 – 7.22 (15H, m, ArH), 6.47 (1H, s, CONH), 6.26 – 6.22 (1H, m, CONH), 4.88 (1H, d, 7.1 Hz, NHCOO), 3.83 (1H, q, 6.4 Hz, COCHNH), 3.43 – 3.25 (4H, m, NHCH₂CH₂NH), 2.78 – 2.70 (1H, m, CHCH₂S), 2.55 (1H, dd, 5.4 Hz, 12.8 Hz, CHCH₂S), 2.09 (2H, t, 7.6 Hz CH₂CO), 1.57 (2H, quin, 7.4 Hz, CH₂CH₂CO), 1.44 (9H, s, C(CH₃)₃), 1.32 – 1.22 (6H, m, CH₂), 0.89 (3H, t, 6.9 Hz, CH₃CH₂); **¹³C NMR** (125 MHz, CDCl₃): δ 174.1 (CH₂CONH), 171.6 (NHCOCH), 155.5 (NHCOO), 144.4 (CHCCH), 129.6 (Ar), 128.2 (Ar), 127.1 (Ar), 80.6 (C(CH₃)₃), 67.4 (C(C₆H₅)₃), 53.9 (COCHNH), 39.7 (NHCH₂CH₂NH), 39.6 (NHCH₂CH₂NH), 36.7 (CH₂CO), 24.0 (CH₂S), 31.6 (CH₂), 29.1 (CH₂), 28.4 (C(CH₃)₃), 25.7 (CH₂CH₂CO), 22.6 (CH₂), 14.2 (CH₃); **HRMS (ESI)**: calculated for C₃₆H₄₇N₃NaO₄S [M+Na]⁺: 640.3179, found: 640.3175; **[α]_D²⁸**: +9.3 (0.086, MeOH).

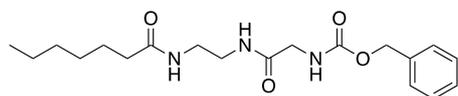
7.1.3.25 Synthesis of (*R*)-1-((2-heptanamidoethyl)amino)-3-mercapto-1-oxopropan-2-aminium chloride (75)



tert-Butyl (*R*)-1-((2-heptanamidoethyl)amino)-1-oxo-3-(tritylthio)propan-2-yl)carbamate (100 mg, 0.16 mmol) under an argon atmosphere was dissolved in trifluoroethanol (5 ml), and triisopropylsilane (TIPS, 100 μ l, 0.82 mmol) was added. Concentrated hydrochloric acid (50 μ l, 0.58 mmol) was added, and the reaction stirred at room temperature for 4 hours. The solvent was removed *in vacuo*, and the residue washed with diethyl ether (3 x 5 ml) to afford (*R*)-1-((2-heptanamidoethyl)amino)-3-mercapto-1-oxopropan-2-aminium chloride as a white solid (48 mg, 96%). **R_r**: 0.05 in MeOH; [α]_D²⁸: +2.8 (0.054, MeOH); **IR** (thin film) ν_{max} : 3267 (N-H stretch), 2926 (C-H stretch), 1636 (C=O stretch), 1238 (C-N stretch);

¹H NMR (500 MHz, MeOD): δ 3.99 (1H, t, 5.9 Hz, COCHNH₂), 3.34 – 3.26 (4H, m, NHCH₂CH₂NH), 3.04 (1H, dd, 5.0 Hz, 14.6 Hz, CH₂SH), 2.93 (1H, dd, 6.9 Hz, 14.7 Hz, CH₂SH), 2.19 (2H, t, 7.6 Hz, CH₂CO), 1.58 (2H, m, CH₂CH₂CO), 1.33 – 1.27 (6H, m, CH₂), 0.89 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (125 MHz, MeOD): δ 176.9 (CH₂CONH), 168.6 (NHCOCH), 56.2 (COCHNH₂), 40.5 (NHCH₂CH₂NH), 39.6 (NHCH₂CH₂NH), 37.2 (CH₂CO), 32.7 (CH₂), 30.0 (CH₂), 26.9 (CH₂CH₂CO), 26.3 (CH₂SH), 25.6 (CH₂), 14.4 (CH₃CH₂), **HRMS (ESI)**: calculated for C₁₂H₂₆N₃O₂S [M+H]⁺: 276.1740, found: 276.1739.

7.1.3.26 Synthesis of benzyl (2-((2-heptanamidoethyl)amino)-2-oxoethyl)carbamate (82)



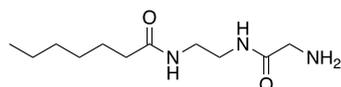
N-(2-aminoethyl)heptanamide (226 mg, 1.31 mmol) and *Z*-Gly-OH (301 mg, 1.44 mmol) under an argon atmosphere was dissolved in anhydrous DMF (15 ml), cooled to 0 °C and DIPEA (0.68 ml, 3.93 mmol) was added. After stirring for 10 minutes, HATU (1.05 g, 2.77 mmol) was added and the reaction allowed to return to room

temperature. DMF was largely removed *in vacuo*, and the product precipitated with the additional of H₂O (200 ml) and collected by filtration. The crude product was purified by column chromatography with a stepwise gradient (from 97:3 EtOAc: MeOH to 9:1 EtOAc: MeOH), affording benzyl (2-((2-heptanamidoethyl)amino)-2-oxoethyl)carbamate as a white solid (326 mg, 68%).

¹H NMR (300 MHz, CDCl₃): δ 7.36 (5H, s, ArH), 6.74 (1H, s, CONH), 6.02 (1H, s, CONH), 5.39 (1H, s, CONH), 5.14 (2H, s, CH₂Ar), 3.85 (2H, d, 5.5 Hz COCH₂NH), 3.43 – 3.36 (4H, m, NHCH₂CH₂NH), 2.16 (2H, t, 4.5 Hz, CH₂CO), 1.65 – 1.57 (2H, m, CH₂CH₂CO), 1.32 – 1.26 (6H, m, CH₂), 0.86 (3H, t, 6.0 Hz, CH₃CH₂), LRMS (ESI): calculated for C₁₉H₃₀N₃O₄ [M+H]⁺: 363.2, found: 364.0.

The data is in accordance with those reported in the literature.¹⁵³

7.1.3.27 Synthesis of *N*-(2-(2-aminoacetamido)ethyl)heptanamide (81)

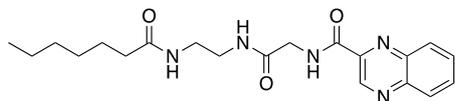


To benzyl (2-((2-heptanamidoethyl)amino)-2-oxoethyl)carbamate (326 mg, 0.90 mmol) and Pd/C (77 mg, 0.72 mmol) under argon, anhydrous MeOH (18 mL) was added and the solution was degassed by bubbling argon through the solution. H₂ (g) was then bubbled through the solution. After 4 hours, the mixture was filtered through Celite and concentrated *in vacuo* to afford *N*-(2-(2-aminoacetamido)ethyl)heptanamide as a white solid (203 mg, 99%).

¹H NMR (300 MHz, MeOD): δ 3.34 – 3.31 (4H, m, NHCH₂CH₂NH), 3.26 (2H, s, COCH₂NH), 2.19 (2H, t, 7.5 Hz, CH₂CO), 1.64 – 1.56 (2H, m, CH₂CH₂CO), 1.37 – 1.28 (6H, m, CH₂), 0.92 (3H, t, 6.0 Hz, CH₃CH₂); LRMS (ESI): calculated for C₁₁H₂₄N₃O₂ [M+H]⁺: 229.2, found: 229.4.

The data is in accordance with those reported in the literature.¹⁵³

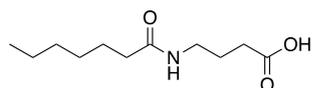
7.1.3.28 Synthesis of *N*-(2-((2-heptanamidoethyl)amino)-2-oxoethyl)quinoxaline-2-carboxamide (80)



N-(2-(2-aminoacetamido)ethyl)heptanamide (30 mg, 0.13 mmol) and quinoxaline-2-carboxylic acid (30 mg, 0.17 mmol) under an argon atmosphere were dissolved in anhydrous THF (5 ml) and cooled to 0°C. DIPEA (68 μ l, 0.39 mmol) was added, and after stirring for 15 minutes, HATU (64 mg, 0.17 mmol) was added. The reaction was warmed and stirred at room temperature. After 24 hours, the solvent was removed *in vacuo* and the residue redissolved in EtOAc. The organic phase was washed with 1M HCl (2 x 5 mL), sat. NaHCO₃ (aq) (2 x 5 mL), and sat. NaCl (aq) (5 mL). The organic phase was dried with MgSO₄ (s) and concentrated *in vacuo* to afford *N*-(2-((2-heptanamidoethyl)amino)-2-oxoethyl)quinoxaline-2-carboxamide as a fawn solid (45 mg, 90%). **R_f**: 0.07 in EtOAc; **IR** (thin film) ν_{max} : 3402 (N-H stretch), 2926 (C-H stretch), 1657 (C=O stretch), 1640 (C=O stretch), 775 (C-H bend);

¹H NMR (500 MHz, CDCl₃): δ 9.61 (1H, s, NCHC(CO)N), 8.59 (1H, t, 5.7 Hz, COCH₂NH), 8.14-8.17 (1H, m, NCCH), 8.11 (1H, dd, 1.2 Hz, 8.0 Hz, NCCH), 7.80 – 7.85 (2H, m, CHCHCHCH), 7.18 (1H, s, CONH), 6.21 (1H, s, CONH), 4.21 (2H, d, 5.9 Hz, COCH₂NH), 3.39 – 3.47 (4H, m, NHCH₂CH₂NH), 2.10 – 2.15 (2H, m, CH₂CH₂CO), 1.46 – 1.53 (2H, m, CH₂CH₂CO), 1.17 – 1.26 (6H, m, CH₂), 0.87 (3H, t, 6.9 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, CDCl₃): δ 175.0 (COCH₂NH), 169.6 (CH₂CONHCH₂), 164.2 (NHCOC), 144.1 (NCCH), 143.7 (COCCH), 142.9 (NHCOC), 140.4 (NCCH), 132.0 (CHCHCHCH), 131.1 (CHCHCHCH), 129.9 (NCCH), 129.6 (NCCH), 43.4 (COCH₂NH), 40.8 (NHCH₂CH₂NH), 39.6 (NHCH₂CH₂NH), 36.7 (CH₂CH₂CO), 31.6 (CH₂), 29.0 (CH₂), 25.7 (CH₂CH₂CO), 22.6 (CH₂), 14.1 (CH₃CH₂); **HRMS(ESI)**: calculated for C₂₀H₂₇N₅O₃Na [M+Na]⁺:408.2006, found: 408.2008.

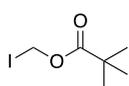
7.1.3.29 Synthesis of 4-heptanamidobutanoic acid (**112**)



γ -Aminobutyric acid (928 mg, 9 mmol) and Et₃N (3.78 mL, 27 mmol) were dissolved in anhydrous MeOH under an inert atmosphere and cooled to 0 °C. Heptanoyl chloride (1.39 mL, 9 mmol) was added dropwise, and the reaction warmed and stirred at room temperature overnight. The solvent was removed *in vacuo*, and redissolved in CHCl₃ (30 mL). H₂O (20 mL) was added and the solution was acidified to pH1 with 1M HCl. The aqueous layer was extracted with CHCl₃ (20 mL); the combined organic layers were dried with MgSO₄ (s), filtered and concentrated to afford **112** as a white solid (1.727 g, 89%). **R_f**: 0.75 (1:4 MeOH: DCM); **IR** (thin film) ν_{max} : 3308 (O-H stretch), 2927 (C-H stretch), 1710 (C=O stretch), 1629 (C=O stretch), 1253 (C-O stretch);

¹H NMR (500 MHz, MeOD): δ 3.20 (2H, t, 6.9 Hz, CONHCH₂), 2.32 (2H, t, 7.5 Hz, CH₂COOH), 2.17 (2H, t, 7.5 Hz, CH₂CONH), 1.78 (2H, quin, 7.2 Hz, CH₂CH₂COOH), 1.60 (2H, quin, 7.3 Hz, CH₂CH₂CONH), 1.36 – 1.28 (6H, m, CH₂), 0.90 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, MeOD): δ 176.9 (COOH), 176.4 (CONH), 39.7 (CONHCH₂), 37.1 (CH₂CONH), 32.7 (CH₂), 32.2 (CH₂COOH), 30.0 (CH₂), 27.0 (CH₂CH₂CO), 25.8 (CH₂CH₂COOH), 23.6 (CH₂), 14.4 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₁H₂₁NO₃Na [M+H]⁺: 238.1415, found: 238.1414.

7.1.3.30 Synthesis of iodomethyl pivalate (**114**)

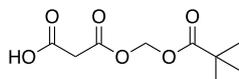


To a solution of chloromethylpivalate (1.44 mL, 10 mmol) in acetone (15 mL) was added NaI (2.70 g, 18 mmol) and stirred under an inert atmosphere with light excluded by aluminium foil for 4 hours. The reaction mixture was concentrated *in vacuo*. The mixture was resuspended in petroleum ether and filtered to remove NaCl. The crude compound was purified by column chromatography using 9:1 petroleum ether: EtOAc to afford **114** as an oil (2.15 g, 89%). **R_f**: 0.8 in 9:1 pet. ether: EtOAc.

¹H NMR (400 MHz, CDCl₃): δ 5.92 (2H, s, CH₂I), 1.19 (9H, s, C(CH₃)₃).

Characterisation data are in accordance with those reported in the literature.³²³

7.1.3.31 Synthesis of 3-oxo-3-((pivaloyloxy)methoxy)propanoic acid (115)

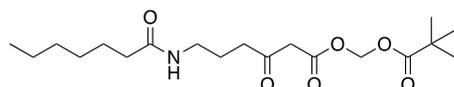


Under light-exclusion using aluminium foil, malonic acid (2.22 g, 21.3 mmol) was added to iodomethylpivalate (2.15 g, 8.88 mmol) in anhydrous THF (20 mL), and DIPEA was added (1.55 mL, 8.88 mmol). The mixture was stirred overnight at 50 °C. Added EtOAc (30 mL) and sat. NaCl (30 mL). The organic phase was washed with sat. NaCl (aq) (30 mL) and 5% NaHCO₃ (aq) (3 x 30 mL). The aqueous layer was acidified to pH 2 with 3 M HCl and extracted with EtOAc (3 x 25 mL) to give **115** as an orange oil (1.337 g, 69%).

¹H NMR (300 MHz, CDCl₃): δ 5.84 (2H, s, CH₂COOH), 3.51 (2H, s, OCH₂O), 1.24 (9H, s, C(CH₃)₃); LRMS (ESI): calculated for C₁₈H₂₇O₁₂ [2M-H]⁻: 350.2, found: 350.0.

Characterisation data are in accordance with those reported in the literature.³²⁴

7.1.3.32 Synthesis of (pivaloyloxy)methyl 6-heptanamido-3-oxohexanoate (116)

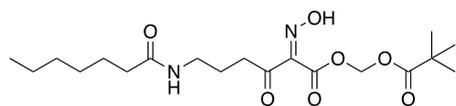


For solution A: 4-heptanamidobutanoic acid (825 mg, 3.83 mmol) was dissolved in acetonitrile (40 mL) at 0 °C under an inert atmosphere, and *N,N*-carbonyldiimidazole (CDI, 639 mg, 3.94 mmol) was added and stirred vigorously for 1.5 hours, while allowing to return to room temperature. For solution B: to a solution of monopivaloyloxymethyl malonate (1.337 g, 6.13 mmol) in anhydrous acetonitrile (50 mL) at 0 °C under an inert atmosphere was added anhydrous MgCl₂ (875 mg, 9.19 mmol), and stirred dropwise for 15 minutes. Et₃N (1.29 mL, 9.19 mmol) was added dropwise over 10 minutes at 0 °C, then the suspension was stirred at room temperature for 15 minutes. Solution A was added dropwise to solution B, and the mixture was

stirred at 30 °C for 16 hours. The reaction mixture was concentrated and the residue redissolved in EtOAc (40 mL). The slurry was washed with 1M HCl (2 x 10 mL), sat. NaHCO₃ (aq) (2 x 10 mL) and sat. NaCl (aq) (20 mL). The organic phase was dried with MgSO₄ (s), filtered and concentrated to afford **116** as a pure yellow oil (1.239 g, 87%). **IR** (thin film) ν_{max} : 2935 (C-H stretch), 1758 (C=O stretch), 1712 (C=O stretch), 1631 (C=O stretch), 1168 (C-O stretch), 1111 (C-O stretch);

¹H NMR (500 MHz, CDCl₃): δ 5.80 (2H, s, OCH₂O), 5.67 (1H, s, CONH), 3.52 (2H, s, COCH₂CO), 3.27 (2H, dd, 6.6 Hz, 12.9 Hz, NHCH₂), 2.62 (2H, t, 6.9 Hz, CH₂COCH₂), 2.17 (2H, t, 7.7 Hz, CH₂CONH), 1.83 (2H, quin, 6.8 Hz, NHCH₂CH₂), 1.63 (2H, quin, 7.1 Hz, CH₂CH₂CO), 1.34 – 1.28 (6H, m, CH₂), 1.25 (9H, s, C(CH₃)₃), 0.90 (3H, t, 6.6 Hz, CH₃CH₂); **¹³C NMR** (150 MHz, CDCl₃): δ 201.7 (CH₂COCH₂), 177.2 (C(CH₃)₃CO), 173.5 (CONH), 165.9 (CH₂COO), 79.7 (OCH₂O), 48.8 (COCH₂CO), 40.3 (CH₂COCH₂), 38.5 (NHCH₂), 36.8 (CH₂CONH), 32.6 (C(CH₃)₃), 31.5 (CH₂), 29.0 (CH₂), 26.9 (C(CH₃)₃), 25.7 (CH₂CH₂CONH), 23.3 (NHCH₂CH₂), 22.5 (CH₂), 14.0 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₉H₃₃NO₆Na [M+Na]⁺: 394.2200, found: 394.2203.

7.1.3.33 Synthesis of (pivaloyloxy)methyl-6-heptanamido-2-(hydroxyimino)-3-oxohexanoate (**117**)

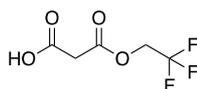


(Pivaloyloxy)methyl 6-heptanamido-3-oxohexanoate (50 mg, 0.135 mmol) was dissolved in AcOH (2 mL) and cooled to 0 °C. NaNO₂ (19 mg, 0.27 mmol) was dissolved in H₂O (1 mL) and added dropwise. The solution was stirred for 165 minutes, after which EtOAc (5 mL) and sat. NaCl (aq) (10 mL) were added. The aqueous layer was extracted with EtOAc (2 x 5 mL) and the organic layers combined. The organic phase was washed with sat. NaHCO₃ (aq) (2 x 10 mL), dried with MgSO₄ (s), filtered and concentrated. The crude product was purified by column chromatography eluting isocratically with petroleum ether: EtOAc (20:13) to afford **117** as a pale yellow solid (20 mg, 36%). **R_f** = 0.29 (1:1 EtOAc: petroleum ether); **IR**

(thin film) ν_{max} : 2929 (C-H stretch), 1728 (C=O stretch), 1627 (C=O stretch), 1551 (C=N stretch), 1117 (N-O stretch);

$^1\text{H NMR}$ (500 MHz, CDCl_3): δ 12.46 (1H, s, NOH), 6.04 (1H, t, 5.6 Hz, NH), 5.92 (2H, s, OCH_2O), 3.31 (2H, dd, 6.7 Hz, 13.3 Hz, NHCH_2), 2.85 (2H, t, 7.1 Hz, $\text{CH}_2\text{COC}(\text{NOH})$), 2.21 (2H, t, 7.7 Hz, CH_2CO), 1.88 (2H, quin, 7.1 Hz, NHCH_2CH_2), 1.61 (2H, quin, 7.1 Hz, $\text{CH}_2\text{CH}_2\text{CO}$), 1.33 – 1.25 (6H, m, CH_2), 1.25 (9H, s, $\text{C}(\text{CH}_3)_3$), 0.89 (3H, t, 6.8 Hz, CH_3CH_2); $^{13}\text{C NMR}$ (150 MHz, CDCl_3): δ 195.2 ($\text{COC}(\text{NOH})$), 177.1 ($\text{COC}(\text{CH}_3)_3$), 175.0 (CONH), 160.8 ($\text{C}(\text{NOH})\text{COO}$), 149.4 ($\text{C}(\text{NOH})$), 79.2 (OCH_2O), 39.2 (NHCH_2), 38.8 ($\text{C}(\text{CH}_3)_3$), 36.7 (CH_2CONH), 34.8 (CH_2CO), 31.5 (CH_2), 28.9 (CH_2), 26.8 ($\text{C}(\text{CH}_3)_3$), 25.6 ($\text{CH}_2\text{CH}_2\text{CONH}$), 23.2 (NHCH_2CH_2), 22.5 (CH_2), 14.0 (CH_3CH_2); **HRMS (ESI)**: calculated for $\text{C}_{19}\text{H}_{32}\text{N}_2\text{O}_7\text{Na}$ $[\text{M}+\text{Na}]^+$: 423.2102, found: 423.2101.

7.1.3.34 Synthesis of 3-oxo-3-(2,2,2-trifluoroethoxy)propanoic acid (**122**)

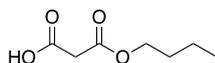


To Meldrum's acid (1.5 g, 10.5 mmol) under argon, 2,2,2-trifluoroethanol (1.5 mL, 20.8 mmol) and toluene (15 mL) were added. The mixture was heated to reflux for 22 hours and the solvent was removed *in vacuo* to afford **122** as a yellow-white solid (1.73 g, 89 %).

$^1\text{H NMR}$ (300 MHz, CDCl_3): δ 4.65 (2H, q, 8.7 Hz), 3.50 (2H, s), **LRMS (ESI)**: calculated $[\text{M}-\text{H}]^+$: 185.0, found: 184.9.

Characterisation data is in accordance with reported in the literature.³²⁵

7.1.3.35 Synthesis of 3-butoxy-3-oxopropanoic acid (**125**)



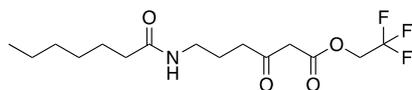
To Meldrum's acid (2 g, 13.9 mmol) under argon, was added *n*-butanol dried over molecular sieves (1.27 mL, 13.9 mmol) and anhydrous toluene (15 mL). The mixture was heated to reflux for 26 hours and the solvent was removed *in vacuo*. The crude

residue was purified by silica chromatography, eluted with 9:1 petroleum ether: EtOAc, followed by 4:1 DCM: MeOH to afford **125** as a colourless oil (1.54 g, 69%).

$^1\text{H NMR}$ (400 MHz, CDCl_3): δ 4.19 (2H, t, 6.6 Hz, COOCH_2), 3.44 (2H, s, OCCH_2CO), 1.74 – 1.62 (2H, m, CH_2), 1.48 – 1.35 (2H, m, CH_2), 0.94 (3H, t, 7.3 Hz, CH_3CH_2), **LRMS (ESI)**: calculated for $[\text{M}-\text{H}]^+$: 159.1, found: 159.1.

Characterization data is in accordance with that reported in the literature.³²⁶

7.1.3.36 Synthesis of 2,2,2-trifluoroethyl 6-heptanamido-3-oxohexanoate (**123**)

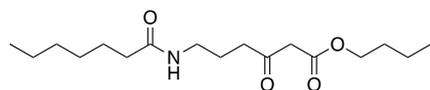


Compound **112** (824 mg, 3.83 mmol) was dissolved in anhydrous MeCN (30 mL) and CDI (640 mg, 3.95 mmol) added at 0 °C under argon. The solution was left to warm to room temperature and stirred for 90 minutes. In a separate flask, **122** (1.139 g, 6.12 mmol) under argon was dissolved in anhydrous MeCN (30 mL); anhydrous MgCl_2 (875 mg, 9.19 mmol) was added at 0 °C and stirred for 15 minutes before adding Et_3N (1.28 mL, 9.19 mmol) dropwise at 0 °C and then stirring the suspension at room temperature for further 15 minutes. The solution containing **112** was added dropwise to the solution of **125**, and the mixture was stirred overnight at 30 °C. The solvent was removed *in vacuo* and the residue redissolved in EtOAc (40 mL). The layers were separated and the organic phase was further washed with 1M HCl (2 x 10 mL), sat. NaHCO_3 (aq) (2 x 10 mL), sat. NaCl (aq) (20 mL), then dried with MgSO_4 (s), filtered and concentrated to afford **123** as a white solid (1.20 g, 92 %). ; **IR** (thin film) ν_{max} : 3324 (N-H stretch), 2935 (C-H stretch), 1758 (C=O stretch), 1712 (C=O stretch), 1631 (C=O stretch), 1168 (C-F stretch);

$^1\text{H NMR}$ (500 MHz, CDCl_3): δ 11.64 (1H, s, COHCH_2CO), 5.58 (1H, s, CONH), 4.52 (2H, q, CH_2CF_3 , 8.3 Hz), 3.57 (2H, s, COCH_2CO), 3.26 (2H, q, 6.5 Hz, CONHCH_2), 2.61 (2H, t, 6.8 Hz, $\text{NHCH}_2\text{CH}_2\text{CH}_2$), 2.14 (2H, t, 7.6 Hz, CH_2CO), 1.82 (2H, quin, 6.7 Hz, NHCH_2CH_2), 1.63 – 1.56 (2H, m, $\text{CH}_2\text{CH}_2\text{CO}$), 1.33 – 1.26 (6H, m, CH_2), 0.88 (3H, t, 6.7 Hz, CH_3CH_2); $^{13}\text{C NMR}$ (125 MHz, CDCl_3): δ 201.5 (CH_2COCH_2), 173.6 (CONH), 165.7 (CH_2COO), 122.9 (CH_2CF_3), 60.8 (CH_2CF_3),

48.6 (COCH₂CO), 40.5 (NHCH₂CH₂CH₂), 38.7 (CONHCH₂), 37.0 (CH₂CO), 31.7 (CH₂), 29.1 (CH₂), 25.8 (CH₂CH₂CO), 23.5 (NHCH₂CH₂), 22.6 (CH₂), 14.2 (CH₃CH₂); ¹⁹F NMR (375 MHz, CDCl₃): δ -73.7 (3F, t, 8.3 Hz, CF₃); HRMS(ESI): calculated for C₁₅H₂₄F₃NNaO₄ [M+Na]⁺: 362.1550, found: 362.1551.

7.1.3.37 Synthesis of butyl 6-heptanamido-3-oxohexanoate (**126**)

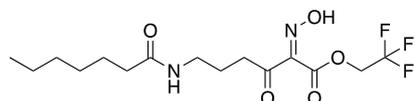


Compound **112** (420 mg, 1.95 mmol) was dissolved in anhydrous MeCN (15 mL) and CDI added (326 mg, 2.01 mmol) at 0 °C under argon; the reaction was left to warm to room temperature and then stirred for 90 minutes. In a separate flask, **125** (0.5 g, 3.12 mmol) under argon was dissolved in anhydrous MeCN (20 mL); anhydrous MgCl₂ (446 mg, 4.68 mmol) was added at 0 °C and the suspension stirred for 15 minutes before adding Et₃N (652 μL, 4.68 mmol) dropwise at 0 °C, then stirring further at room temperature for 15 minutes. The solution containing **112** was added dropwise to the solution of **125**; this mixture was stirred overnight at 30 °C. The solvent was removed *in vacuo* and the remaining residue redissolved in EtOAc and H₂O. The layers were separated, and the organic phase was further washed with 1M HCl (2 x 15 mL), sat. NaHCO₃ (aq) (2 x 15 mL) and sat. NaCl (aq) (20 mL), then dried with MgSO₄ (s), filtered and concentrated to afford **126** as a white solid (538 mg, 88 %). IR (thin film) ν_{max}: 3317 (N-H stretch), 2929 (C-H stretch), 1736 (C=O stretch), 1706 (C=O stretch), 1637 (C=O stretch), 1153 (C-O stretch);

¹H NMR (500 MHz, CDCl₃): δ 5.70 (1H, s, CONH), 4.13 (2H, t, 6.6 Hz, COOCH₂), 3.44 (2H, s, COCH₂CO), 3.25 (2H, q, 6.4 Hz, CONHCH₂), 2.61 (2H, t, 6.7 Hz, CH₂COCH₂CO), 2.14 (2H, t, 7.6 Hz, CH₂CONH), 1.66 – 1.57 (4H, m, COOCH₂CH₂, CH₂CH₂CONH), 1.42 – 1.34 (2H, m, COOCH₂CH₂CH₂), 1.33 – 1.27 (6H, m, CH₂), 0.93 (3H, t, 7.3 Hz, COOCH₂CH₂CH₂CH₃), 0.87 (3H, t, 6.1 Hz, CH₃CH₂); ¹³C NMR (125 MHz, CDCl₃): δ 202.9 (COCH₂COO), 173.5 (COOCH₂CH₂CH₂), 167.5 (COO), 65.5 (COOCH₂), 49.4 (COCH₂COO), 40.5 (CH₂COCH₂COO), 38.8 (CONHCH₂), 37.0 (CH₂CONH), 31.7 (CH₂), 30.6 (COOCH₂CH₂), 29.1 (CH₂), 25.8 (CH₂CH₂CONH), 23.4 (CONHCH₂CH₂), 22.7 (CH₂), 19.2 (COOCH₂CH₂CH₂), 14.2

(CH₃CH₂), 13.8 (CH₂CH₃); **HRMS(ESI)**: calculated for C₁₇H₃₁NO₄Na [M+Na]⁺: 336.2145, found: 336.2146.

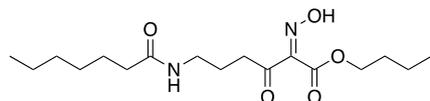
7.1.3.38 Synthesis of 2,2,2-trifluoroethyl-6-heptanamido-2-(hydroxyimino)-3-oxohexanoate (**124**)



To **123** (250 mg, 0.74 mmol) in AcOH (3 mL) and H₂O (2.5 mL) at 0 °C was added NaNO₂ (508 mg, 7.4 mmol) in H₂O (8 mL); the solution was returned to room temperature and stirred for 16 hours. EtOAc (20 mL) was added, the layers separated and the organic phase washed with sat. NaCl_(aq) (2 x 25 mL); the brine layer was also extracted with EtOAc (2 x 15 mL) and the organic layers combined and dried with MgSO_{4(s)}, filtered and concentrated *in vacuo*. The crude residue was purified by silica chromatography, eluting with 1:1 petroleum ether: Et₂O to Et₂O to afford **124** as a white solid (166 mg, 61 %). **IR** (thin film) ν_{\max} : 2926 (C-H stretch), 1761 (C=O stretch), 1694 (C=O stretch), 1613 (C=O stretch), 1568 (C=N stretch), 1153 (C-F stretch);

¹H NMR (500 MHz, CDCl₃): δ 12.38 (1H, s, NOH), 5.82 (1H, s, CONH), 4.65 (2H, q, 8.3 Hz, COOCH₂), 3.33 (2H, dd, 6.7 Hz, CONHCH₂), 2.88 (2H, t, 7.0 Hz, CH₂COCN), 2.19 (2H, t, 7.6 Hz, CH₂CONH), 1.91 (2H, quin, 7.1 Hz, CONHCH₂CH₂), 1.62 – 1.56 (2H, m, CH₂CH₂CONH), 1.33 – 1.25 (6H, m, CH₂), 0.87 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (125 MHz, CDCl₃): δ 195.2 (CH₂COCN), 175.2 (CONH), 160.4 (COO), 149.1 (CNOH), 122.7 (CF₃, q, 277 Hz), 61.1 (COOCH₂, q, 37 Hz), 39.4 (CONHCH₂), 36.9 (CH₂CONH), 35.0 (CH₂COCN), 31.6 (CH₂), 29.0 (CH₂), 25.7 (CH₂CH₂CONH), 23.3 (CONHCH₂CH₂), 22.6 (CH₂), 14.1 (CH₃CH₂); **¹⁹F NMR** (300 MHz, CDCl₃): δ -73.5 (CF₃); **HRMS(ESI)**: calculated for C₁₅H₂₃F₃N₂O₅Na [M+Na]⁺: 391.1451, found: 391.1449.

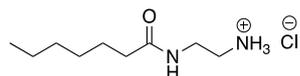
7.1.3.39 Synthesis of butyl-6-heptanamido-2-(hydroxyimino)-3-oxohexanoate (**127**)



Compound **126** (120 mg, 0.38 mmol) was dissolved in AcOH (1 mL) and added to NaNO₂ (106 mg, 1.53 mmol) in ice-cold H₂O (2 mL); the reaction was stirred at 0 °C for 2 hours, then at room temperature for 2.5 hours. The reaction was diluted with H₂O and extracted with EtOAc (3 x 10 mL); the organic layer was isolated and washed with sat. NaHCO₃ (aq) until gas ceased to evolve, then dried with MgSO₄ (s), filtered and concentrated *in vacuo*. The crude residue was purified by silica chromatography, eluting with petroleum ether to Et₂O, to afford **127** as a white solid (78 mg, 58 %). **IR** (thin film) ν_{\max} : 3397 (N-H stretch), 2957 (C-H stretch), 1713 (C=O stretch), 1686 (C=O stretch), 1622 (C=O stretch), 1543 (C=N stretch);

¹H NMR (500 MHz, CDCl₃): δ 12.07 (1H, s, NOH), 5.95 (1H, t, CONH, 5.4 Hz), 4.31 (2H, t, 6.7 Hz, COOCH₂), 3.31 (2H, q, 6.7 Hz, CONHCH₂), 2.85 (2H, t, 7.1 Hz, CH₂COC(NO₂H)), 2.19 (2H, t, 7.7 Hz, CH₂CONH), 1.88 (2H, quin, 7.1 Hz, CONHCH₂CH₂), 1.72 – 1.66 (2H, m, COOCH₂CH₂), 1.64 – 1.57 (2H, m, CH₂CH₂CONH), 1.45 – 1.37 (2H, m, COOCH₂CH₂CH₂), 1.33 – 1.25 (6H, m, CH₂), 0.93 (3H, t, 7.4 Hz, COOCH₂CH₂CH₂CH₃), 0.87 (3H, t, 6.7 Hz, CH₃CH₂); **¹³C NMR** (125 MHz, CDCl₃): δ 195.9 (COCN), 175.0 (CONH), 162.4 (COO), 150.6 (COCNCOO), 66.0 (COOCH₂), 39.4 (CONHCH₂), 36.9 (CH₂CONH), 35.1 (CH₂COCN), 30.6 (COOCH₂CH₂), 29.0 (CH₂), 25.8 (CH₂CH₂CONH), 23.4 (CONHCH₂CH₂), 22.6 (CH₂), 19.1 (COOCH₂CH₂CH₂), 14.2 (CH₃CH₂), 13.8 (CH₃CH₂); **HRMS(ESI)**: calculated for C₁₇H₃₀N₂O₅Na [M+Na]⁺: 365.2047, found: 365.2041.

7.1.3.40 Synthesis of 2-heptanamidoethan-1-aminium chloride (**131**)



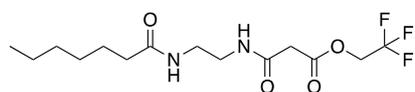
tert-Butyl (2-heptanamidoethyl)carbamate (503 mg, 1.85 mmol) under an argon atmosphere was dissolved in anhydrous Et₂O (5 mL) and 2M HCl in Et₂O (5 mL) was

added. The reaction mixture was stirred at room temperature for 4 hours, and the solvent removed *in vacuo* to afford 2-heptanamidoethan-1-aminium chloride as a white solid (385 mg, quant.).

$^1\text{H NMR}$ (300 MHz, MeOD): δ 3.35 – 3.30 (4H, m, $\text{NHCH}_2\text{CH}_2\text{NH}$), 2.19 (2H, t, 7.5 Hz, CH_2CO), 1.66 – 1.55 (2H, m, $\text{CH}_2\text{CH}_2\text{CO}$), 1.37 – 1.29 (6H, m, CH_2), 0.92 (3H, t, 6.0 Hz, CH_3CH_2); **LRMS (ESI)**: calculated for $\text{C}_9\text{H}_{21}\text{N}_2\text{O}$ $[\text{M}+\text{H}]^+$: 173.2, found: 173.2.

The data are in accordance with those reported in the literature.¹⁵³

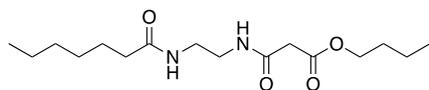
7.1.3.41 Synthesis of 2,2,2-trifluoroethyl 3-((2-heptanamidoethyl)amino)-3-oxopropanoate (**132**)



Compound **131** (363 mg, 1.74 mmol) and **122** (421 mg, 2.26 mmol) under argon were dissolved in anhydrous THF, cooled to 0 °C and DIPEA (909 μL , 5.22 mmol) was added and stirred for 30 minutes. HATU (859 mg, 2.26 mmol) was added and the reaction was warmed to room temperature and stirred for 4 hours. Solvent was removed *in vacuo* and the residue redissolved in EtOAc (20 mL) and 1M HCl (20 mL); the organic layer was isolated and further washed with sat. NaHCO_3 (aq) (20 mL) and sat. NaCl (aq) (20 mL). The crude product was purified by silica chromatography, eluting with EtOAc to 9: 1 EtOAc: Acetone to afford **132** as a white solid (427 mg, 72 %).

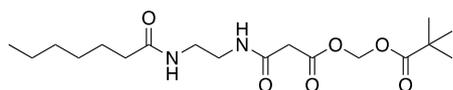
$^1\text{H NMR}$ (500 MHz, CDCl_3): δ 7.20 (1H, s, CONH), 6.00 (1H, s, CONH), 4.54 (2H, q, 8.3 Hz, CH_2CF_3); 3.45 – 3.41 (4H, m, $\text{NHCH}_2\text{CH}_2\text{NH}$), 2.80 (2H, s, COCH_2CO), 2.18 (2H, t, 7.7 Hz, CH_2CONH), 1.64 – 1.57 (2H, m, $\text{CH}_2\text{CH}_2\text{CO}$), 1.34 – 1.26 (6H, m, CH_2), 0.88 (3H, t, 6.6 Hz, CH_3CH_2); $^{13}\text{C NMR}$ (125 MHz, CDCl_3): δ 174.7 (CH_2CONH), 167.4 (NHCOCCH_2), 165.1 (COO), 122.8 (q, 277 Hz, CF_3), 61.1 (q, 37 Hz, CH_2CF_3), 39.8 ($\text{NHCH}_2\text{CH}_2\text{NH}$), 38.8 (COCH_2CO), 36.8 (CH_2CONH), 31.6 (CH_2), 29.1 (CH_2), 25.7 ($\text{CH}_2\text{CH}_2\text{CONH}$), 22.6 (CH_2), 14.2 (CH_3); $^{19}\text{F NMR}$ (300 MHz, CDCl_3): δ -73.7 (CF_3); **HRMS(ESI)**: calculated for $\text{C}_{14}\text{H}_{23}\text{F}_3\text{N}_2\text{O}_4\text{Na}$ $[\text{M}+\text{Na}]^+$: 363.1502, found: 363.1500.

7.1.3.42 Synthesis of butyl 3-((2-heptanamidoethyl)amino)-3-oxopropanoate (133)



Compound **131** (279 mg, 1.34 mmol) and **125** (279 mg, 1.74 mmol) under argon were dissolved in THF (8 mL) and cooled for 0 °C; DIPEA (0.7 mL, 4.01 mmol) was added in and stirred for 30 minutes. HATU (661 mg, 1.74 mmol) was added in one portion, the solution left to warm to room temperature and stirred for 48 hours. The solvent was removed *in vacuo*, and the residue redissolved in EtOAc (20 mL) and 1M HCl (15 mL). The organic phase was separated, washed with sat. NaHCO₃ (aq) (15 mL) and sat. NaCl (aq) (15 mL), then dried with MgSO₄ (s), filtered and concentrated *in vacuo*. The crude product was purified by silica chromatography 1:3 petroleum ether: EtOAc to 9:1 EtOAc: Acetone to afford **133** as a white solid (270 mg, 64 %). **IR** (thin film) ν_{\max} : 3289 (N-H stretch), 2929 (C-H stretch), 1742 (C=O stretch), 1636 (C=O stretch); **¹H NMR** (500 MHz, CDCl₃): δ 7.45 (1H, s, CONH), 6.18 (1H, s, CONH), 4.14 (2H, t, 6.7 Hz, COOCH₂), 3.47 – 3.36 (4H, m, NHCH₂CH₂NH), 3.30 (2H, s, COCH₂CO), 2.17 (2H, t, 7.7 Hz, CH₂CONH), 1.68 – 1.55 (4H, m, CH₂CH₂CONH, COOCH₂CH₂), 1.45 – 1.34 (2H, m, COOCH₂CH₂CH₂), 1.33 – 1.25 (6H, m, CH₂), 0.93 (3H, t, 7.4 Hz, CH₃), 0.87 (3H, t, 6.6 Hz, CH₃); **¹³C NMR** (125 MHz, CDCl₃): δ 174.3 (CH₂CH₂CONH), 169.5 (COO), 166.6 (NHCOCH₂CO), 65.7 (COOCH₂), 41.3 (COCH₂CO), 40.1 (NHCH₂CH₂NH), 40.0 (NHCH₂CH₂NH), 36.9 (CH₂CONH), 31.7 (CH₂), 30.6 (COOCH₂CH₂), 29.1 (CH₂), 25.8 (CH₂CH₂CONH), 22.6 (CH₂), 19.2 (COOCH₂CH₂CH₂), 14.2 (CH₃), 13.8 (COOCH₂CH₂CH₂CH₃); **HRMS(ESI)**: calculated for C₁₆H₃₀N₂O₄Na [M+Na]⁺: 337.2098, found: 337.2089.

7.1.3.43 Synthesis of ((3-((2-heptanamidoethyl)amino)-3-oxopropanoyl)oxy)methyl pivalate (134)



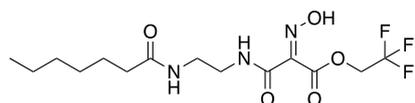
Compound **131** (429 mg, 2.06 mmol) and **115** (583 mg, 2.67 mmol) under argon were dissolved in anhydrous THF (10 mL), cooled to 0 °C; DIPEA (1.07 mL, 6.17 mmol)

was added in and stirred for 30 minutes. HATU (1.02 g, 2.67 mmol) was added and stirred at room temperature for 16 hours. The solvent was removed *in vacuo*, and the residue redissolved in EtOAc (25 mL). The organic layer was washed with 1M HCl (2 x 15 mL), sat. NaHCO₃ (aq) (25 mL) and sat. NaCl (aq) (2 x 25 mL), then dried with MgSO₄ (s), filtered and concentrated *in vacuo*. The crude product was purified by silica chromatography, eluting with 1:1 petroleum ether: EtOAc to 19:1 EtOAc: Acetone, to afford **134** as a white powder (395 mg, 40 %).

¹H NMR (500 MHz, CDCl₃): δ 7.22 (1H, s, CONH), 6.16 (1H, s, CONH), 5.79 (2H, s, OCH₂O), 3.45 – 3.37 (4H, m, NHCH₂CH₂NH), 3.35 (2H, s, COCH₂CO), 2.20 – 2.16 (2H, m, CH₂CONH), 1.64 – 1.57 (2H, m, CH₂CH₂CONH), 1.33 – 1.25 (6H, m, CH₂), 1.22 (9H, s, C(CH₃)₃), 0.87 (3H, t, 6.6 Hz, CH₃); ¹³C NMR (125 MHz, CDCl₃): δ 177.4 (COC(CH₃)₃), 174.5 (CH₂CONH), 167.8 (CH₂COOCH₂), 165.5 (NHCOCH₂CO), 80.2 (OCH₂O), 41.6 (COCH₂CO), 40.3, 39.9 (NHCH₂CH₂NH), 38.9 (C(CH₃)₃), 36.8 (CH₂CONH), 31.7 (CH₂), 29.1 (CH₂), 27.0 (C(CH₃)₃), 25.8 (CH₂CH₂CONH), 22.6 (CH₂), 14.2 (CH₃);

HRMS(ESI): calculated for C₁₈H₃₂N₂O₆Na [M+Na]⁺: 395.2153, found: 395.2156.

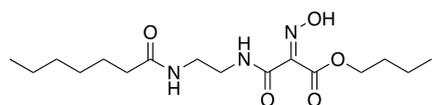
7.1.3.44 Synthesis of 2,2,2-trifluoroethyl 3-((2-heptanamidoethyl)amino)-2-(hydroxyimino)-3-oxopropanoate (**135**)



Compound **132** (126 mg, 0.37 mmol) was dissolved in EtOH (5 mL) and concentrated aqueous hydrochloric acid (12 M, 2.5 mL) was added. NaNO₂ (204 mg, 2.96 mmol) in H₂O (1 mL) was added dropwise and stirred at room temperature for 45 minutes. The solvent was removed *in vacuo* and the residue redissolved in H₂O (15 mL). The aqueous layer was extracted with EtOAc (3 x 20 mL) and back washed with sat. NaCl (aq) (20 mL). The organic phase was dried with MgSO₄ (s), filtered and concentrated. The crude product was purified by silica chromatography, eluting with EtOAc to afford **135** as a white solid (129 mg, 82 %) as a mixture of *cis* and *trans* isomers. IR (thin film) ν_{max}: 3256 (N-H stretch), 2927 (C-H stretch), 1745 (C=O stretch), 1609 (C=O stretch), 1553 (C=N stretch), 1167 (C-F stretch);

¹H NMR (500 MHz, MeOD): δ 4.77 (qd, 3.8 Hz, 8.7 Hz), 3.86 (q, 9.2 Hz) COOCH₂CF₃, 3.45-3.33 (4H, m, NHCH₂CH₂NH), 2.20 – 2.15 (2H, m, CH₂CONH), 1.64 – 1.55 (2H, m, CH₂CH₂CONH), 1.36 – 1.27 (6H, m, CH₂), 0.90 (3H, t, 6.0 Hz, CH₃), ¹³C NMR (125 MHz, MeOD): δ 177.0, 176.9, 176.7 (CH₂CONH), 163.4, 163.1, 162.8 (NHCOCN), 161.4, 161.3 (COO), 146.6, 146.4 (CNOH), 126.3 (q, 278 Hz, CF₃), 124.5 (q, 276 Hz, CF₃), 124.4 (q, 276 Hz, CF₃), 61.9 (q, 37 Hz, CH₂CF₃), 61.7 (q, 37 Hz, CH₂CF₃), 61.2 (q, 35 Hz, CH₂CF₃), 40.3, 40.2, 39.7, 39.6 (NHCH₂CH₂NH), 37.2, 37.1 (CH₂CONH), 32.7, 32.6 (CH₂), 30.0 (CH₂), 26.9, 26.8 (CH₂), 23.6, 23.5 (CH₂), 14.4 (CH₃); ¹⁹F NMR (300 MHz, MeOD): δ -75.1, -75.3, -78.7 (CF₃); HRMS (ESI): calculated for C₁₄H₂₂F₃N₃O₅Na [M+Na]⁺: 392.1404, found: 392.1404.

7.1.3.45 Synthesis of butyl 3-((2-heptanamidoethyl)amino)-2-(hydroxyimino)-3-oxopropanoate (137)

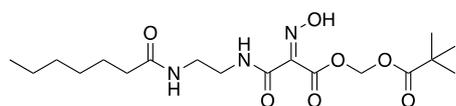


Compound **133** (250 mg, 0.80 mmol) was dissolved in EtOH (5 mL) and concentrated aqueous HCl (12 M, 2.5 mL) added. NaNO₂ (439 mg, 6.36 mmol) in H₂O (1.5 mL) was added dropwise and the solution stirred at room temperature for 1 hour. The solvent was removed *in vacuo*, and the residue diluted with H₂O (25 mL). The aqueous phase was extracted with EtOAc (3 x 20 mL) which was then back extracted with sat. NaCl (aq) (20 mL). The organic phase was dried with MgSO₄ (s), filtered and concentrated. The crude product was purified by silica chromatography, eluting with 1:1 petroleum ether: EtOAc to EtOAc to afford **137** as a white solid (124 mg, 45 %) as a mixture of *cis* and *trans* isomers.

¹H NMR (500 MHz, MeOD): δ 4.28 (2H, dt, 6.5 Hz, 13.1 Hz, COOCH₂), 3.46 – 3.34 (4H, m, NHCH₂CH₂NH), 2.19 (2H, dd, 7.4 Hz, 13.0 Hz, CH₂CONH), 1.73 – 1.66 (2H, m, COOCH₂CH₂), 1.65 – 1.57 (2H, dt, 6.7 Hz, 13.4 Hz, CH₂CH₂CONH), 1.50 – 1.41 (2H, m, COOCH₂CH₂CH₂), 1.38 – 1.29 (6H, m, CH₂), 0.97 (3H, t, 7.4 Hz,

COOCH₂CH₂CH₂CH₃), 0.92 (3H, t, 6.1 Hz, CH₃); ¹³C NMR (125 MHz, MeOD): δ 176.9, 176.7 (CH₂CONH), 163.9, 162.8 (NHCOCN), 163.3, 162.9 (COO), 147.5, 146.7 (CNOH), 66.7, 66.5 (COOCH₂), 40.2, 39.7 (NHCH₂CH₂NH), 37.2, 37.1 (CH₂CONH), 32.7 (CH₂), 31.6 (COOCH₂CH₂), 30.0 (CH₂), 26.9, 26.8 (CH₂CH₂CONH), 23.6 (CH₂), 20.1, 20.0 (COOCH₂CH₂CH₂), 14.4, 14.4 (CH₃), 14.0, 13.9 (COOCH₂CH₂CH₂CH₃); **HRMS(ESI)**: calculated for C₁₆H₂₉N₃O₅Na [M+Na]⁺: 366.1999, found: 366.2002.

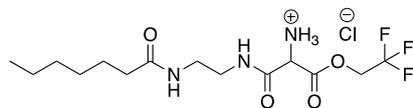
7.1.3.46 Synthesis of (pivaloyloxy)methyl-3-((2-heptanamidoethyl)amino)-2-(hydroxyimino)-3-oxopropanoate (139)



Compound **134** (384 mg, 1.03 mmol) was dissolved in EtOH (6 mL) and concentrated aqueous hydrochloric acid (12 M, 3 mL) added. NaNO₂ (569 mg, 8.25 mmol) in H₂O (4 mL) was added dropwise and stirred at room temperature for 1 hour. The solvent was removed *in vacuo*, and the residue diluted with H₂O (15 mL), and the aqueous layer extracted with DCM (3 x 15 mL). The organic phase was dried with MgSO₄ (s), filtered and concentrated *in vacuo* to afford **139** as a yellow oil (390 mg, 92 %) as mixture a *cis* and *trans* isomers. **IR** (thin film) ν_{max}: 3335 (N-H stretch), 2930 (C-H stretch), 1757 (C=O stretch), 1639 (C=O stretch), 1553 (C=N stretch);

¹H NMR (500 MHz, CDCl₃): δ 8.65 (1H, s, CONH), 7.29, 6.26 (1H, s, CONH), 5.94, 5.92 (2H, s, OCH₂O), 3.54-3.37 (4H, m, NHCH₂CH₂NH), 2.21 – 2.16 (2H, m, CH₂CONH), 1.64 – 1.55 (2H, m, CH₂CH₂CONH), 1.32 – 1.25 (6H, m, CH₂), 1.22, 1.18 (9H, s, COC(CH₃)₃), 0.89 – 0.86 (3H, m, CH₃CH₂); ¹³C NMR (125 MHz, CDCl₃): δ 177.2, 177.0 (OCOC(CH₃)₃), 175.6, 174.6 (CH₂CONH), 163.3, 162.7 (CNCOO), 160.9, 160.2 (NHCOCN), 144.5, 138.0 (COCNCO), 80.9, 79.6 (OCH₂O), 40.9, 40.2 (C(CH₃)₃), 39.9, 39.5, 39.5, 39.0 (NHCH₂CH₂NH), 36.8, 36.8 (CH₂CO), 31.6, 31.6, 29.0, 29.0, 22.6 (CH₂), 25.7 (CH₂CH₂CO), 27.6, 27.0, 27.0 (C(CH₃)₃), 14.2 (CH₃); **HRMS(ESI)**: calculated for C₁₈H₃₁N₃O₇Na [M+Na]⁺: 424.2054, found: 424.2054.

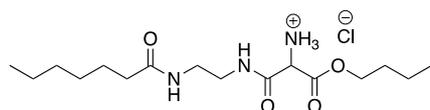
7.1.3.47 Synthesis of 1-((2-heptanamidoethyl)amino)-1,3-dioxo-3-(2,2,2-trifluoroethoxy)propan-2-aminium chloride (**136**)



Anhydrous MeOH (2.5 mL) was added to **135** (16 mg, 0.043 mmol) together with Pd/C (4 mg, 0.035 mmol), then 0.5 M HCl in MeOH (258 μ L) was added dropwise. H_{2(g)} was bubbled through the solution, which was then stirred under a H_{2(g)} atmosphere for 5.5 hours. The reaction mixture was filtered through Celite and concentrated *in vacuo* to afford **136** as a light brown solid (17 mg, 96 %).

¹H NMR (500 MHz, MeOD): δ 4.93 – 4.78 (2H, m, OCH₂CF₃), 3.91 (1H, s, COCHNH₂CO), 3.41-3.34 (4H, m, NHCH₂CH₂NH), 2.21 (2H, t, 7.6 Hz, CH₂CO), 1.67 – 1.55 (2H, m, CH₂CH₂CO), 1.39 – 1.29 (6H, m, CH₂), 0.92 (3H, t, 6.6 Hz, CH₃); ¹³C NMR (125 MHz, MeOD): δ 176.9 (CH₂CO), 164.9 (COO), 163.6, 162.9 (both NHCOCH), 125.2 (q, 277 Hz, CF₃), 63.0 (q, 37 Hz, CH₂CF₃), 54.5 (COCHNH₂CO), 40.9, 39.4 (NHCH₂CH₂NH), 37.1 (CH₂CO), 32.7 (CH₂), 30.0 (CH₂), 26.8 (CH₂CH₂CO), 23.6 (CH₂), 14.3 (CH₃); ¹⁹F NMR (400 MHz, CDCl₃): -71.2; HRMS(ESI): calculated for C₁₄H₂₄F₃O₄Na [M+Na]⁺: 378.1611, found: 378.1603; IR (thin film) ν_{\max} : 3381, 3285, 1769, 1647, 1166.

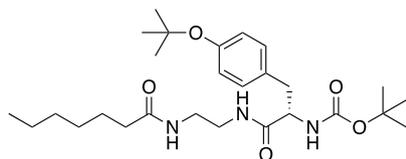
7.1.3.48 Synthesis of 1-butoxy-3-((2-heptanamidoethyl)amino)-1,3-dioxopropan-2-aminium chloride (**138**)



Anhydrous MeOH (8 mL) was added to **137** (105 mg, 0.31 mmol) and Pd/C (24 mg, 0.23 mmol), then 0.5M HCl in MeOH (1.83 mL) was added dropwise. H₂ was bubbled through the solution, then stirred under a H₂ atmosphere for 4.5 hours. The reaction mixture was filtered through Celite and concentrated *in vacuo* to afford **138** as a pale pink solid (112 mg, 94 %). IR (thin film) ν_{\max} : 3295 (N-H stretch), 2929 (C-H stretch), 1747 (C=O stretch), 1633 (C=O stretch);

¹H NMR (500 MHz, MeOD): δ 4.76 (1H, s, COCHNH₂), 4.35 – 4.25 (2H, m, COOCH₂), 3.44-3.30 (4H, m, NHCH₂CH₂NH), 2.20 (2H, t, 7.6 Hz, CH₂CONH), 1.73 – 1.66 (2H, m, COOCH₂CH₂), 1.63 – 1.56 (2H, m, CH₂CH₂CONH), 1.45 – 1.38 (2H, m, COOCH₂CH₂CH₂), 1.35 – 1.28 (6H, m, CH₂), 0.95 (3H, t, 7.4 Hz, COOCH₂CH₂CH₂CH₃), 0.90 (3H, t, 6.5 Hz, CH₃); **¹³C NMR** (125 MHz, MeOD): δ 176.8 (CH₂CONH), 165.8 (COOCH₂), 163.9 (NHCOCHNH₂), 68.3 (COOCH₂), 56.6 (COCHNH₂CO), 40.8 (NHCH₂CH₂NH), 39.5 (NHCH₂CH₂NH), 37.1 (CH₂CONH), 32.7 (CH₂), 31.5 (COOCH₂CH₂), 30.0 (CH₂), 26.8 (CH₂CH₂CONH), 23.6 (CH₂), 19.9 (COOCH₂CH₂CH₂), 14.4 (CH₃), 13.9 (COOCH₂CH₂CH₂CH₃); **HRMS(ESI)**: calculated for C₁₆H₃₁N₃O₄Na [M+Na]⁺: 352.2207, found: 352.2203.

7.1.3.49 Synthesis of *tert*-butyl (*S*)-(3-(4-(*tert*-butoxy)phenyl)-1-((2-heptanamidoethyl)amino)-1-oxopropan-2-yl)carbamate (167)

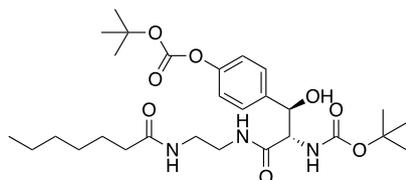


2-Heptanamidoethan-1-aminium chloride (504 mg, 2.41 mmol) and Boc-L-Tyr(^tBu)-OH (1.06 g, 3.14 mmol) under an argon atmosphere were dissolved in anhydrous THF (20 mL), then cooled to 0 °C and DIPEA (1.26 mL, 7.23 mmol) was added. After 15 minutes, HATU (1.19 g, 3.14 mmol) was added and the reaction allowed to return to room temperature. After 16 hours, solvent was removed *in vacuo* and the residue redissolved in EtOAc (30 ml). The organic phase was washed with 1M HCl (40 ml), sat. NaHCO₃ (aq) (40 ml) and sat. NaCl (aq) (40 ml). The organics were dried with MgSO₄ (s) and filtered, then concentrated *in vacuo*. The crude product was purified by column chromatography with a stepwise gradient from 3:1 petroleum ether: EtOAc to EtOAc, to afford *tert*-butyl (*S*)-(3-(4-(*tert*-butoxy)phenyl)-1-((2-heptanamidoethyl)amino)-1-oxopropan-2-yl)carbamate as a white solid (832 mg, 70%). **R_f**: 0.37 in EtOAc; **IR** (thin film) ν_{max}: 2929 (C-H stretch), 1649 (C=O stretch), 1162 (C-O stretch), 845 (C-H bend);

¹H NMR (500 MHz, MeOD): δ 7.15 (2H, t, 7.0 Hz, CCHCH), 6.91 (2H, t, 8.4 Hz, CCHCH), 4.23 – 4.18 (1H, m, COCHNH), 3.28 – 3.21 (4H, m, NHCH₂CH₂NH), 3.04

(1H, dd, 5.9 Hz, 13.7 Hz, CHCH₂C), 2.79 (1H, dd, 9.1 Hz, 13.6 Hz, CHCH₂C). 2.17 (2H, t, 7.6 Hz, CH₂CO), 1.59 (2H, quin, 7.0 Hz, CH₂CH₂CO), 1.39 (9H, s, COC(CH₃)₃), 1.36 – 1.30 (15H, m, COOC(CH₃)₃, CH₂CH₂CH₂), 0.91 (3H, t, 6.6 Hz, CH₃); ¹³C NMR (125 MHz, MeOD): δ 176.6 (CH₂CONH), 174.8 (NHCOCH), 157.7 (NHCOO), 155.3 (COC(CH₃)₃), 133.8 (CHCH₂C), 130.9 (CCHCH), 125.2 (CCHCH), 80.6 (CHCOC(CH₃)₃), 79.5 (COOC(CH₃)₃), 57.87 (COCHNH), 40.1, 39.8 (NHCH₂CH₂NH), 38.7 (CHCH₂C), 37.2 (CH₂CO), 32.7 (CH₂), 30.1 (COOC(CH₃)₃), 29.2 (CH₂), 28.7 (CHCOC(CH₃)₃), 26.9 (CH₂CH₂CO), 23.6 (CH₂), 14.4 (CH₃CH₂); **HRMS(ESI)**: calculated for C₂₇H₄₅N₃O₅Na [M+Na]⁺: 514.3251, found: 514.3250; [α]_D²⁸: -2.9 (0.068, MeOH).

7.1.3.50 Synthesis of *tert*-butyl ((1*R*,2*S*)-1-(4-((*tert*-butoxycarbonyl)oxy)phenyl)-3-((2-heptanamidoethyl)amino)-1-hydroxy-3-oxopropan-2-yl)carbamate (**168**)

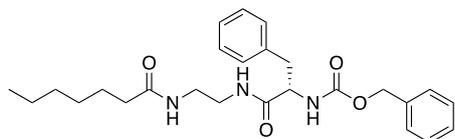


2-Heptanamidoethan-1-aminium chloride (14 mg, 0.066 mmol) and (2*S*,3*R*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((*tert*-butoxycarbonyl)oxy)phenyl)-3-hydroxypropanoic acid (29 mg, 0.073 mmol) were dissolved in anhydrous DMF (1 mL), cooled to 0 °C, and DIPEA (35 μL, 0.20 mmol) added. After 10 minutes, HATU (28 mg, 0.073 mmol) was added, and the reaction stirred at room temperature for 45 minutes. The solvent was removed *in vacuo* by azeotropeing with heptane. The crude product was purified by column chromatography using a stepwise gradient from 1:1 petroleum ether: EtOAc to EtOAc to afford **168** as a white solid (31 mg, 86%). **R_f**: 0.26 (EtOAc).

¹H NMR (500 MHz, CDCl₃): 7.40 (2H, d, 8.4 Hz, OCCHCH), 7.15 (2H, d, 8.6 Hz, OCCH), 6.85 (CONH), 6.04 (CONH), 5.38 (1H, d, 5.4 Hz, CHOH), 5.31 (1H, d, 2.3 Hz, CCHOHCH), 4.31 (1H, d, 5.8 Hz, COCHNH), 3.48 (2H, q, 7.0 Hz, CH₂CH₂NHCOCH), 3.40 – 3.34 (1H, m, CH₂CH₂NHCOCH), 3.28 – 3.22 (1H, m, CH₂CH₂NHCOCH), 2.15 (2H, t, 7.5 Hz, CH₂CO), 1.63 – 1.56 (2H, m, CH₂CH₂CO), 1.55 (18H, s, C(CH₃)₃), 1.36 – 1.31 (6H, m, CH₂), 0.87 (3H, t, 6.8 Hz, CH₃CH₂); ¹³C

NMR (125 MHz, CDCl₃): 174.6 (CH₂CONH), 171.8 (NHCOCH), 156.1 (NHCOO), 150.6 (OCOO), 137.3 (CHOHCCH), 127.0 (CHOHCCH), 121.2 (OCOCCH), 83.6 (OCOOC(CH₃)₃), 80.7 (NHCOOC(CH₃)₃), 72.2 (COCHCHOH), 59.9 (COCHNH), 40.0 (NHCH₂CH₂NH), 39.6 (NHCH₂CH₂NH), 36.7 (CH₂CO), 31.5 (CH₂), 29.0 (CH₂) 27.7 (C(CH₃)₃), 25.5 (CH₂CH₂CO), 22.5 (CH₂), 14.0 (CH₃CH₂); **HRMS (ESI)**: calculated for C₂₈H₄₅N₃O₈Na [M+Na]⁺: 574.3099, found: 574.3091.

7.1.3.51 Synthesis of benzyl (S)-(1-((2-heptanamidoethyl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (169)

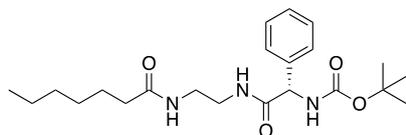


N-(2-aminoethyl)heptanamide (141 mg, 0.82 mmol), *Z*-L-Phe-OH (294 mg, 0.98 mmol), and HATU (373 mg, 0.98 mmol) were dissolved in anhydrous DCM (5 mL) and stirred at room temperature for 10 minutes. DIPEA (0.43 ml, 2.46 mmol) was added, and the reaction mixture stirred at room temperature for 48 hours. The solvent was removed *in vacuo*, and the residue redissolved in EtOAc. The organic phase was washed with sat. NaHCO₃ (aq) (3 x 10 mL) and sat. NaCl (aq) (10 mL), then dried with MgSO₄ (s), filtered, and concentrated. The crude product was purified using column chromatography using an isocratic elution with EtOAc, to afford benzyl (S)-(1-((2-heptanamidoethyl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate as a white solid (95 mg, 25%). **R_f**: 0.39 (EtOAc).

¹H NMR (500 MHz, CDCl₃): δ 7.32 – 7.06 (10H, m, ArH), 6.33 (1H, s, CONH), 5.87 (1H, s, CONH), 5.34 – 5.31 (1H, m, COCHNH), 5.11 – 5.06 (2H, m, COOCH₂Ar), 4.34 (1H, dd, 6.7 Hz, 6.6 Hz, COCHNH), 3.31 – 3.23 (4H, m, NHCH₂CH₂NH), 3.10 – 3.02 (2H, m, CHCH₂Ar), 2.10 (2H, t, 7.7 Hz, CH₂CO), 1.61 – 1.53 (2H, m, CH₂CH₂CO), 1.30 – 1.24 (6H, m, CH₂), 0.87 (3H, t, 6.5 Hz, CH₃CH₂); **¹³C NMR** (125 MHz, CDCl₃): δ 174.3 (CH₂CONH), 171.8 (COCH), 156.1 (NHCOO), 136.5 (CHCCH₂), 136.2 (CHCCH₂), 129.4 (CH), 128.9 (CH), 128.7 (CH), 128.4 (CH), 128.2 (CH), 127.3 (CH), 67.3 (COOCH₂Ar), 56.6 (COCHNH), 40.2 (NHCH₂CH₂NH), 39.7 (NHCH₂CH₂NH), 38.7 (CHCH₂Ar), 31.7 (CH₂), 29.1 (CH₂),

25.8 ($\text{CH}_2\text{CH}_2\text{CO}$), 22.7 (CH_2), 14.2 (CH_3CH_2); **HRMS (ESI)**: calculated for $\text{C}_{26}\text{H}_{36}\text{N}_3\text{O}_4\text{Na}$ $[\text{M}+\text{Na}]^+$: 476.2607, found: 476.2609.

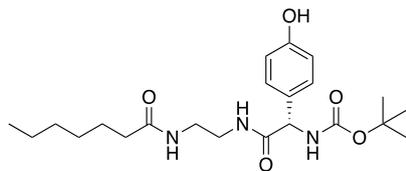
7.1.3.52 Synthesis of *tert*-butyl (*S*)-(2-((2-heptanamidoethyl)amino)-2-oxo-1-phenylethyl)carbamate (170)



2-Heptanamidoethan-1-aminium chloride (250 mg, 1.20 mmol) and Boc-L-Phe-OH (332 mg, 1.32 mmol) under an argon atmosphere were dissolved in anhydrous DMF (3 mL), cooled to 0 °C and DIPEA (627 μ L, 3.60 mmol) added. After 15 minutes, HATU (502 mg, 1.32 mmol) was added and the reaction mixture stirred at room temperature for 4 hours. The solvent was removed *in vacuo* and the residue redissolved in EtOAc. The organic phase was washed with 5% LiCl (aq) (5 x 10 mL), 1M HCl (10 mL), sat. NaHCO₃ (aq) (10 mL) and sat. NaCl (aq) (2 x 20 mL), then dried with MgSO₄ (s), filtered and concentrated. The crude product was purified by column chromatography with a stepwise gradient from 1:1 EtOAc: petroleum ether to EtOAc to afford *tert*-butyl (*S*)-(2-((2-heptanamidoethyl)amino)-2-oxo-1-phenylethyl)carbamate as a white solid (399 mg, 82%).

¹H NMR (500 MHz, MeOD): δ 7.43 – 7.30 (5H, m, ArH), 5.10 (1H, s, COCHNH), 3.33 – 3.25 (4H, m, NHCH₂CH₂NH), 2.14 (2H, t, 7.4 Hz, CH₂CO), 1.61 – 1.54 (2H, m, CH₂CH₂CO), 1.45 (9H, s, C(CH₃)₃), 1.37 – 1.29 (6H, m, CH₂), 0.93 (3H, t, 6.9 Hz, CH₃CH₂); ¹³C NMR (125 MHz, MeOD): δ 176.7 (CH₂CONH), 173.6 (NHCOCH), 157.5 (NHCOO), 139.2 (COCHCCH), 129.8 (CH), 129.3 (CH), 128.5 (CH), 80.9 (C(CH₃)₃), 60.4 (COCHNH), 40.1 (NHCH₂CH₂NH), 39.8 (NHCH₂CH₂NH), 37.2 (CH₂CO), 32.7 (CH₂), 30.0 (CH₂), 28.7 (C(CH₃)₃), 26.8 (CH₂CH₂CO), 23.6 (CH₂), 14.4 (CH₃CH₂); HRMS (ESI): calculated for C₂₂H₃₅N₃O₄Na [M+Na]⁺: 428.2520, found: 428.2522.

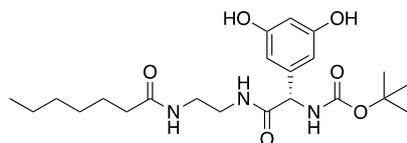
7.1.3.53 Synthesis of *tert*-butyl (S)-2-((2-heptanamidoethyl)amino)-1-(4-hydroxyphenyl)-2-oxoethylcarbamate (171)



2-Heptanamidoethan-1-aminium chloride (400 mg, 1.92 mmol) and Boc-L-Hpg-OH (564 mg, 2.11 mmol) under an argon atmosphere were dissolved in DMF (5 mL), cooled to 0 °C, and DIPEA (1.00 ml, 5.75 mmol) were added. After 15 minutes, HATU (802 mg, 2.11 mmol) was added and stirred for 4 hours. DMF was removed *in vacuo* and the residue redissolved in EtOAc. The organic phase was washed with 5% LiCl_(aq) (5 x 10 mL), 1M HCl (2 x 10 mL), sat. NaHCO_{3(aq)} (2 x 10 mL) and sat. NaCl_(aq) (2 x 10 mL), then dried with MgSO_{4(s)}, filtered and concentrated. The crude product was purified by column chromatography with a stepwise gradient from 1:1 petroleum ether: EtOAc to EtOAc, affording *tert*-butyl (S)-2-((2-heptanamidoethyl)amino)-1-(4-hydroxyphenyl)-2-oxoethylcarbamate as a white solid (275 mg, 34%). **R_f** = 0.32 (EtOAc); **IR** (thin film) ν_{max} : 2929 (C-H stretch), 1641 (C=O stretch), 1161 (C-O stretch);

¹H NMR (500 MHz, MeOD): δ 7.20 (2H, d, 8.5 Hz, COCHCCH), 6.75 (2H, d, 8.5 Hz, COCHCCH), 4.96 (1H, s, COCHNH), 3.35 – 3.22 (4H, m, NHCH₂CH₂NH), 2.12 (2H, t, 7.4 Hz, CH₂CO), 1.58 – 1.52 (2H, m, CH₂CH₂CO), 1.44 (9H, s, C(CH₃)₃), 1.34 – 1.28 (6H, m, CH₂), 0.91 (3H, t, 6.8 Hz, CH₃CH₂); **¹³C NMR** (125 MHz, MeOD): δ 176.7 (CH₂CONH), 174.2 (NHCOCH), 158.7 (CHCOH), 157.4 (NHCOO), 129.7 (COCHC), 116.5 (CHCOH), 80.8 (C(CH₃)₃), 59.9 (COCHNH), 40.1 (NHCH₂CH₂NH), 39.9 (NHCH₂CH₂NH), 37.2 (CH₂CO), 32.7 (CH₂), 30.0 (CH₂), 28.7 (C(CH₃)₃), 26.8 (CH₂CH₂CO), 23.6 (CH₂), 14.4 (CH₃CH₂); **HRMS (ESI)**: calculated for C₂₂H₃₅N₃O₅Na [M+Na]⁺: 444.2469, found: 444.2472; [α]_D²⁸: -57.4 (0.070, MeOH).

7.1.3.54 Synthesis of *tert*-butyl (*S*)-(1-(3,5-dihydroxyphenyl)-2-((2-heptanamidoethyl)amino)-2-oxoethyl)carbamate (173)

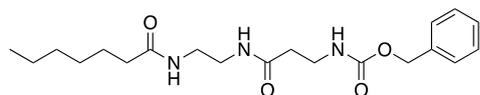


2-Heptanamidoethan-1-aminium chloride (83 mg, 0.39 mmol) and Boc-L-Dhpg-OH (71 mg, 0.25 mmol) under an argon atmosphere were dissolved in anhydrous THF (5 mL) and cooled to 0 °C. NaHCO₃ (s) (84 mg, 1.0 mmol) was added, and 5 minutes later DEPBT (225 mg, 0.75 mmol) was added. The reaction was stirred at room temperature overnight and diluted with H₂O (15 mL) and EtOAc (15 mL). The aqueous layer was isolated and extracted with EtOAc (3 x 10 mL). The combined organic phases were washed with 1M HCl (15 mL), sat. NaHCO₃ (aq) (3 x 10 mL) and sat. NaCl (aq) (15 mL), then dried with MgSO₄ (s), filtered and concentrated. The crude product was purified using column chromatography using an isocratic elution with EtOAc, to afford *tert*-butyl (*S*)-(1-(3,5-dihydroxyphenyl)-2-((2-heptanamidoethyl)amino)-2-oxoethyl)carbamate (29 mg, 27%). **R_f**: 0.38 (EtOAc);

¹H NMR (500 MHz, MeOD): δ 6.32 (2H, d, 1.8 Hz, CHCOHCHC), 6.21 (1H, s, CHCOHCHC), 4.91 (1H, s, COCHNH), 3.34 – 3.26 (4H, m, NHCH₂CH₂NH), 2.15 (2H, t, 7.6 Hz, CH₂CO), 1.60 – 1.53 (2H, m, CH₂CH₂CO), 1.46 (9H, s, C(CH₃)₃), 1.35 – 1.29 (6H, m, CH₂), 0.91 (3H, t, 6.8 Hz, CH₃CH₂); **¹³C NMR** (125 MHz, MeOD): δ 176.8 (CH₂CONH), 173.7 (COCHNH), 159.9 (COH), 157.5 (NHCOO) 141.0 (COCHCCH), 106.9 (CHCCHCOH), 103.4 (COHCH), 80.9 (C(CH₃)₃), 60.2 (COCHNH), 40.1 (NHCH₂CH₂NH), 39.8 (NHCH₂CH₂NH), 37.2 (CH₂CO), 32.7 (CH₂), 30.0 (CH₂), 28.7 (C(CH₃)₃), 26.8 (CH₂CH₂CO), 23.6 (CH₂), 14.4 (CH₃CH₂);

LRMS (ESI): calculated for C₂₂H₃₅N₃O₆Na [M+Na]⁺: 460.2, found: 460.2.

7.1.3.55 Synthesis of benzyl (3-((2-heptanamidoethyl)amino)-3-oxopropyl)carbamate (172)

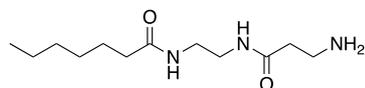


2-Heptanamidoethan-1-aminium chloride (300 mg, 1.44 mmol) and Z- β -Ala-OH (353 mg, 1.58 mmol) under an argon atmosphere were dissolved in DMF (2 mL), cooled to 0 °C, and DIPEA (752 μ L, 4.32 mmol) added. After 15 minutes, HATU (601 mg, 1.58 mmol) was added and stirred at room temperature for 16 hours. The solvent was removed *in vacuo* and the residue redissolved in EtOAc. The organic phase was washed with 5% LiCl_(aq) (5 x 10 mL), 1M HCl (2 x 20 mL), sat. NaHCO_{3(aq)} (2 x 20 mL) and sat. NaCl_(aq) (20 mL). The organic phase was isolated and dried with MgSO_{4(s)}, filtered and concentrated to afford pure benzyl (3-((2-heptanamidoethyl)amino)-3-oxopropyl)carbamate (495 mg, 91%).

¹H NMR (300 MHz, MeOD): δ 7.35 (5H, s, ArH), 5.08 (2H, s, CH₂Ar), 3.40 (2H, t, 6.0 Hz, COCH₂CH₂NH₂), 3.30 – 3.22 (4H, m, NHCH₂CH₂NH), 2.39 (2H, t, 6.4 Hz, COCH₂CH₂NH₂), 2.18 (2H, t, 7.20 Hz, CH₂CO), 1.61 – 1.54 (2H, m, CH₂CH₂CO), 1.35 – 1.27 (6H, m, CH₂), 0.91 (3H, t, 5.9 Hz, CH₃CH₂); LRMS (ESI): calculated for C₂₀H₃₁O₄Na [M+Na]⁺: 400.2, found: 400.3.

The data are in accordance with those reported in the literature.¹⁵³

7.1.3.56 Synthesis of N-(2-(3-aminopropanamido)ethyl)heptanamide (180)

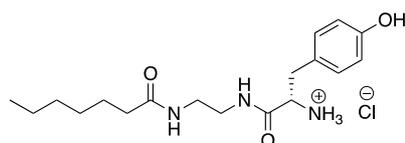


Benzyl (3-((2-heptanamidoethyl)amino)-3-oxopropyl)carbamate (495 mg, 1.31 mmol) and Pd/C (10%) (30 mg, 0.28 mmol) under an argon atmosphere were suspended in anhydrous MeOH (20 mL) and degassed. Hydrogen gas was bubbled through the solution for 5 hours. The reaction mixture was filtered through Celite and concentrated to afford N-(2-(3-aminopropanamido)ethyl)heptanamide (302 mg, 95%) as an off-white solid.

$^1\text{H NMR}$ (300 MHz, MeOD): δ 3.35 – 3.28 (4H, m, $\text{NHCH}_2\text{CH}_2\text{NH}$), 2.94 (2H, t, 6.3 Hz, $\text{COCH}_2\text{CH}_2\text{NH}_2$), 2.38 (2H, t, 6.2 Hz, $\text{COCH}_2\text{CH}_2\text{NH}_2$), 2.19 (2H, t, 7.4 Hz, CH_2CO), 1.66 – 1.57 (2H, m, $\text{CH}_2\text{CH}_2\text{CO}$), 1.38 – 1.30 (6H, m, CH_2), 0.92 (3H, t, 6.0 Hz, CH_3CH_2); **LRMS (ESI)**: calculated for $\text{C}_{12}\text{H}_{26}\text{N}_3\text{O}_2$ $[\text{M}+\text{H}]^+$: 244.2, found: 244.4.

The data are in accordance with those reported in the literature.¹⁵³

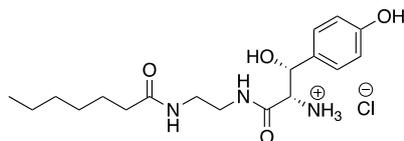
7.1.3.57 Synthesis of (S)-1-((2-heptanamidoethyl)amino)-3-(4-hydroxyphenyl)-1-oxopropan-2-aminium chloride (174)



tert-Butyl (*S*)-(3-(4-(*tert*-butoxy)phenyl)-1-((2-heptanamidoethyl)amino)-1-oxopropan-2-yl)carbamate (638 mg, 1.30 mmol) was dissolved in 5:3 Et₂O: MeOH (8 mL), and 2M HCl in Et₂O (3.24 mL, 6.49 mmol) added slowly. The reaction mixture was stirred at room temperature for 5 hours and the solvent removed *in vacuo*. The residue was washed with Et₂O (3 x 5 mL) to afford (*S*)-1-((2-heptanamidoethyl)amino)-3-(4-hydroxyphenyl)-1-oxopropan-2-aminium chloride (475 mg, 98%) as a foamy brown solid. **R_f**: 0.86 (1:1 DCM: MeOH); **IR** (thin film) ν_{max} : 3244 (N-H stretch), 2926 (C-H stretch), 1613 (C=O stretch), 826 (C-H bend);

$^1\text{H NMR}$ (500 MHz, MeOD): δ 7.09 (2H, d, 8.4 Hz, CH_2CCH), 6.78 (2H, d, 8.4 Hz, CH_2CCHCH), 3.92 (1H, t, 7.3 Hz, COCHNH), 3.30 – 3.15 (4H, m, $\text{NHCH}_2\text{CH}_2\text{NH}$), 3.08 (1H, dd, 6.8 Hz, 14.1 Hz, CHCH_2C), 2.94 (1H, dd, 7.9 Hz, 14.1 Hz, CHCH_2C), 2.18 (2H, t, 7.6 Hz, CH_2CO), 1.63 – 1.53 (2H, m, $\text{CH}_2\text{CH}_2\text{CO}$), 1.37 – 1.25 (6H, m, CH_2), 0.89 (3H, t, 6.6 Hz, CH_3CH_2); $^{13}\text{C NMR}$ (125 MHz, MeOD): δ 176.8 (CH_2CONH), 169.9 (COCH), 158.3 (COH), 131.6 (CH_2CCH), 126.1 (CH_2CCH), 116.8 (CH_2CCHCH), 56.1 (COCHNH), 40.3 ($\text{NHCH}_2\text{CH}_2\text{NH}$), 39.6 ($\text{NHCH}_2\text{CH}_2\text{NH}$), 37.9 (CHCH_2C), 37.2 (CH_2CO), 32.7 (CH_2), 30.0 (CH_2), 26.9 ($\text{CH}_2\text{CH}_2\text{CO}$), 23.6 (CH_2), 14.4 (CH_3CH_2); **HRMS (ESI)**: calculated for $\text{C}_{18}\text{H}_{30}\text{N}_3\text{O}_3$ $[\text{M}+\text{H}]^+$: 336.2282, found: 336.2279; $[\alpha]_{\text{D}}^{28}$: +32.0 (0.077, MeOH).

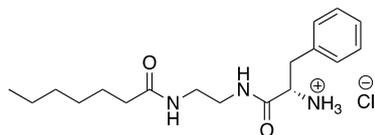
7.1.3.58 Synthesis of (2S,3R)-1-((2-heptanamidoethyl)amino)-3-hydroxy-3-(4-hydroxyphenyl)-1-oxopropan-2-aminium chloride (175)



tert-Butyl ((1R,2S)-1-(4-((*tert*-butoxycarbonyl)oxy)phenyl)-3-((2-heptanamidoethyl)amino)-1-hydroxy-3-oxopropan-2-yl)carbamate (31 mg, 0.056 mmol) was dissolved in Et₂O (2 mL) and 2M HCl in Et₂O (0.28 mL, 0.56 mmol) was added. The reaction was stirred at room temperature for 3 hours, and the solvent was removed *in vacuo* to afford (2S,3R)-1-((2-heptanamidoethyl)amino)-3-hydroxy-3-(4-hydroxyphenyl)-1-oxopropan-2-aminium chloride as a white solid (16 mg, 86%). **R_f**: 0.72 (1:1 MeOH: DCM); **IR** (thin film) ν_{max} : 3251 (N-H stretch), 2954 (C-H stretch), 2852 (O-H stretch), 1613 (C=O stretch), 726 (C-H bend);

¹H NMR (500 MHz, MeOD): δ 7.09 (2H, d, 8.4 Hz, CH₂CCH), 6.78 (2H, d, 8.4 Hz, CH₂CCHCH), 3.92 (1H, t, 7.3 Hz, COCHNH₂), 3.29-3.15 (4H, m, NHCH₂CH₂NH), 3.08 (1H, dd, 6.8 Hz, 14.1 Hz, CH₂CCH), 2.94 (1H, dd, 7.9 Hz, 14.1 Hz, CH₂CCH), 2.18 (2H, t, 7.6 Hz, CH₂CONH), 1.59 – 1.52 (2H, m, CH₂CH₂CO), 1.34 – 1.25 (6H, m, CH₂), 0.89 (3H, t, 6.6 Hz, CH₃); **¹³C NMR** (125 MHz, MeOD): δ 176.8 (CH₂CONH), 169.9 (COCHNH₂), 158.3 (CHCOH), 131.6 (CHCHCOH), 126.1 (CH₂CCH), 116.8 (CHCHCOH), 56.1 (COCHNH₂), 40.3 (NHCH₂CH₂NH), 39.6 (NHCH₂CH₂NH), 37.9 (CHCH₂C), 37.2 (CH₂CO), 32.7 (CH₂), 30.0 (CH₂), 26.9 (CH₂CH₂CO), 25.6 (CH₂), 14.4 (CH₃); **HRMS (ESI)**: calculated for C₁₈H₂₉N₃O₄Na [M+Na]⁺: 374.2050, found: 374.2048; [α]_D²⁸: +16.7 (0.042, MeOH).

7.1.3.59 Synthesis of (S)-1-((2-heptanamidoethyl)amino)-1-oxo-3-phenylpropan-2-aminium chloride (176)

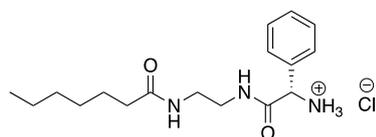


tert-Butyl (*S*)-1-((2-heptanamidoethyl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (588 mg, 1.40 mmol) was dissolved in Et₂O (10 mL), and 2M HCl in

Et₂O (7 mL, 14.0 mmol) was added slowly. The reaction mixture was stirred at room temperature for 4.5 hours and the solvent removed *in vacuo*. The residue was triturated with Et₂O (3 x 5 mL) to afford (*S*)-1-((2-heptanamidoethyl)amino)-1-oxo-3-phenylpropan-2-aminium chloride as a white solid (416 mg, 93%). **R_f** = 0.97 (1:1 DCM: MeOH); **IR** (thin film) ν_{max} : 3222 (N-H stretch), 2927 (C-H stretch), 1635 (C=O stretch), 748 (C-H bend);

¹H NMR (500 MHz, MeOD): δ 7.32 – 7.19 (5H, m, ArH), 3.50 (1H, t, 7.0 Hz, COCHNH₂), 3.25 – 3.15 (4H, m, NHCH₂CH₂NH), 2.99 (1H, dd, 13.5 Hz, 6.5 Hz, CHCH₂), 2.80 (1H, dd, 13.5 Hz, 7.5 Hz, CHCH₂), 2.15 (2H, t, 7.5 Hz, CH₂CO), 1.61 – 1.55 (2H, m, CH₂CH₂CO), 1.34 – 1.28 (6H, m, CH₂), 0.92 – 0.88 (3H, m, CH₃CH₂); **¹³C NMR** (125 MHz, MeOD): δ 177.1 (COCH), 176.6 (CH₂CONH), 138.9 (CH₂C), 130.4 (CCHCHCH), 129.6 (CCHCH), 127.8 (CCHCHCH), 57.9 (COCHNH), 42.5 (CHCH₂), 39.9 (NHCH₂CH₂NH), 39.8 (NHCH₂CH₂NH), 37.2 (CH₂CO), 32.7 (CH₂), 30.0 (CH₂), 26.9 (CH₂CH₂CO), 23.6 (CH₂), 14.4 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₈H₃₀N₃O₂ [M+H]⁺: 319.2020, found: 319.2019; [α]_D²⁶: +8.14 (c = 0.018, MeOH); [α]_D²⁸: +33.4 (0.061, MeOH).

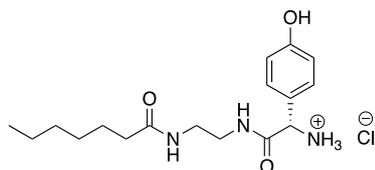
7.1.3.60 Synthesis of (*S*)-2-((2-heptanamidoethyl)amino)-2-oxo-1-phenylethan-1-aminium chloride (177)



tert-Butyl (*S*)-2-((2-heptanamidoethyl)amino)-2-oxo-1-phenylethylcarbamate (391 mg, 0.96 mmol) was dissolved in 2:1 MeOH:Et₂O (6 mL), and 2M HCl in Et₂O (2.4 mL, 4.8 mmol) added. After stirring for 5 hours, the solvent was removed *in vacuo*. The residue was triturated with Et₂O (3 x 3 mL) to afford (*S*)-2-((2-heptanamidoethyl)amino)-2-oxo-1-phenylethan-1-aminium chloride (300 mg, 91%) as an off-white solid. **R_f** = 0.86 (1:1 DCM: MeOH); **IR** (thin film) ν_{max} : 3294 (N-H stretch), 2910 (C-H stretch), 1682 (C=O stretch), 1641 (C=O stretch), 733 (C-H bend); **¹H NMR** (500 MHz, MeOD): δ 7.55 – 7.47 (5H, m, ArH), 4.93 (1H, s, COCHNH), 3.32 – 3.21 (4H, m, NHCH₂CH₂NH), 2.11 (2H, m, CH₂CO), 1.58 – 1.52 (2H, m,

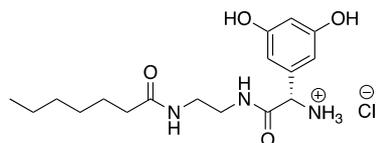
CH₂CH₂CO), 1.36 – 1.29 (6H, m, **CH₂**), 0.93 (3H, t, 6.9 Hz, **CH₃CH₂**); **¹³C NMR** (125 MHz, MeOD): δ 176.7 (**CH₂CONH**), 169.2 (**COCHNH**), 134.5 (**COCHCCH**), 131.1 (**CH**), 130.6 (**CH**), 129.3 (**CH**), 58.0 (**COCHNH**), 40.5 (**NHCH₂CH₂NH**), 39.5 (**NHCH₂CH₂NH**), 37.1 (**CH₂CO**), 32.7 (**CH₂**), 30.0 (**CH₂**), 26.8 (**CH₂CH₂CO**), 23.6 (**CH₂**), 14.4 (**CH₃CH₂**); **HRMS (ESI)**: calculated for C₁₇H₂₈N₃O₂: 306.2176, found: 306.2177; [α]_D²⁶: +64.2 (0.083, MeOH).

7.1.3.61 Synthesis of (*S*)-2-((2-heptanamidoethyl)amino)-1-(4-hydroxyphenyl)-2-oxoethan-1-aminium chloride (178)



tert-Butyl (*S*)-2-((2-heptanamidoethyl)amino)-1-(4-hydroxyphenyl)-2-oxoethyl)carbamate (235 mg, 0.56 mmol) was dissolved in Et₂O (10 mL) and 2M HCl in Et₂O (1.4 mL, 2.8 mmol) and stirred for 3 hours. The solvent was removed *in vacuo* and the residue was washed with Et₂O (2 x 3 mL), to afford (*S*)-2-((2-heptanamidoethyl)amino)-1-(4-hydroxyphenyl)-2-oxoethan-1-aminium chloride (198 mg, 99%) as a pale yellow solid. **R_f** = 0.59 (1:1 MeOH: DCM); **IR** (thin film) ν_{max} : 3270 (N-H stretch), 2924 (C-H stretch), 1639 (C=O stretch), 835 (C-H bend); **¹H NMR** (500 MHz, MeOD): δ 7.30 (2H, d, 8.5 Hz, **CHCCH**), 6.84 (2H, d, 8.5 Hz, **CHCCHCH**), 4.81 (1H, s, **COCHNH₂**), 3.32 – 3.18 (4H, m, **NHCH₂CH₂NH**), 2.10 (2H, t, 7.6 Hz, **CH₂CO**), 1.56 – 1.51 (2H, m, **CH₂CH₂CO**), 1.34 – 1.29 (6H, m, **CH₂**), 0.89 (3H, t, 6.8 Hz, **CH₃**); **¹³C NMR** (125 MHz, MeOD): δ 176.8 (**CH₂CO**), 169.7 (**COCHNH₂**), 160.3 (**COH**), 130.8 (**CCHCH**), 124.9 (**CCHCH**), 117.1 (**CCHCH**), 57.6 (**COCHNH₂**), 40.4, 39.6 (**NHCH₂CH₂NH**), 37.0 (**CH₂CO**), 32.7 (**CH₂**), 30.0 (**CH₂**), 26.8 (**CH₂CH₂CO**), 23.6 (**CH₂**), 14.4 (**CH₃**); **HRMS (ESI)**: calculated for C₁₇H₂₇N₃O₃Na [M+Na]⁺: 344.1945, found: 344.1942; [α]_D²⁶: -21.3 (0.045, MeOH).

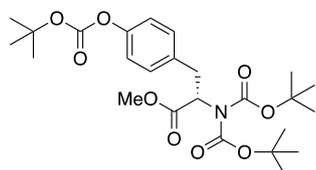
7.1.3.62 Synthesis of (S)-1-(3,5-dihydroxyphenyl)-2-((2-heptanamidoethyl)amino)-2-oxoethan-1-aminium (179)



tert-Butyl (*S*)-1-(3,5-dihydroxyphenyl)-2-((2-heptanamidoethyl)amino)-2-oxoethyl)carbamate (29 mg, 0.066 mmol) was dissolved in EtOH (4 ml), and 2M HCl in Et₂O (2 mL, 4.0 mmol) and the reaction stirred overnight at room temperature. The solvent was removed *in vacuo* to afford (*S*)-1-(3,5-dihydroxyphenyl)-2-((2-heptanamidoethyl)amino)-2-oxoethan-1-aminium chloride (18 mg, 82%) as a pale brown solid. **R_f** = 0.51 (1:1 MeOH: DCM); **IR** (thin film) ν_{max} : 2919 (C-H stretch), 1649 (C=O stretch), 1602 (C=O stretch), 1033 (C-O stretch);

¹H NMR (500 MHz, MeOD): δ 6.42 (2H, d, 2.0 Hz, CHCCHCOH), 6.36 (1H, t, 2.0 Hz, CCHCOHCH), 4.70 (1H, s, COCHNH₂), 3.37 – 3.21 (4H, m, NHCH₂CH₂NH), 2.18 – 2.14 (2H, m, CH₂CO), 1.59 – 1.54 (2H, m, CH₂CH₂CO), 1.36 – 1.28 (6H, m, CH₂), 0.92 (3H, t, 6.8 Hz, CH₃CH₂); **¹³C NMR** (125 MHz, MeOD): δ 176.9 (CH₂CONH), 169.2 (COCHNH), 160.7 (COH), 136.2 (COCHC), 107.4 (CHCCHCOH), 104.9 (COHCHCOH), 57.9 (COCHNH₂), 40.5 (NHCH₂CH₂NH), 39.5 (NHCH₂CH₂NH), 37.1 (CH₂CO), 32.7 (CH₂), 30.0 (CH₂), 26.8 (CH₂CH₂CO), 23.6 (CH₂), 14.4 (CH₃CH₂); **HRMS (ESI)**: calculated for C₁₇H₂₈N₃O₄ [M+H]⁺: 338.2074, found: 338.2075; [α]_D²⁸: +48.4 (0.052, MeOH).

7.1.3.63 Synthesis of methyl (S)-2-(bis(*tert*-butoxycarbonyl)amino)-3-(4-((*tert*-butoxycarbonyl)oxy)phenyl)propanoate (183)



L-Tyr-OMe (1.51 mg, 6.52 mmol) was dissolved in 2:1 THF: H₂O (15 mL), and Na₂CO₃ (2.07 g, 19.5 mmol) was added portion wise. The mixture was cooled to 0 °C

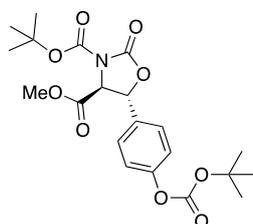
and Boc₂O (2.85 g, 13.04 mmol) was added. The reaction mixture was stirred at room temperature for 2 hours, then diluted with EtOAc (40 mL) and sat. NaCl (aq) (30 mL). The organic layer was concentrated and used directly in the next step following the literature procedure.²⁹⁴

The organic concentrate containing *N*-Boc-L-Tyr-OMe (6.52 mmol) was dissolved in MeCN (15 mL), and 4-dimethylaminopyridine (DMAP, 797 mg, 6.52 mmol) was added. Boc₂O (4.27 g, 19.56 mmol) was added and stirred at room temperature for 16 hours. The solvent was removed *in vacuo* and the residue redissolved in EtOAc (50 mL). The organic layer was washed with sat. NH₄Cl (aq) (50 mL), H₂O (50 mL), and sat. NaCl (aq) (50 mL), dried with MgSO₄ (s), filtered, and concentrated. The crude residue was purified by column chromatography with an isocratic elution using 4:1 petroleum ether: EtOAc to afford (*S*)-2-(bis(*tert*-butoxycarbonyl)amino)-3-(4-((*tert*-butoxycarbonyl)oxy)phenyl)propanoate (2.36 g, 73% over 2 steps).

¹H NMR (400 MHz, CDCl₃): δ 7.18 (2H, d, 7.6 Hz, OCCH), 7.07 (2H, d, 7.5 Hz, OCCHCH), 5.14 (1H, dd, 5.0 Hz, 9.7 Hz, COCHNH), 3.74 (3H, s, COOCH₃), 3.42 (1H, dd, 4.6 Hz, 14.2 Hz, CHCH₂C), 3.23 – 3.16 (1H, m, CHCH₂C), 1.54 (9H, s, C(CH₃)₃), 1.39 (18H, s, C(CH₃)₃); **LRMS (ESI)**: calculated for C₂₅H₃₇NO₉Na [M+Na]⁺: 518.2, found: 518.3.

The data are in accordance with those reported in the literature.²⁹⁴

7.1.3.64 Synthesis of 3-(*tert*-butyl 4-methyl (4*S*,5*R*)-5-(4-((*tert*-butoxycarbonyl)oxy)phenyl)-2-oxooxazolidine-3,4-dicarboxylate (185)



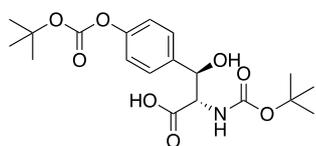
Methyl (*S*)-2-(bis(*tert*-butoxycarbonyl)amino)-3-(4-((*tert*-butoxycarbonyl)oxy)phenyl)propanoate (500 mg, 1.00 mmol), AIBN (3 mg, 0.018 mmol), and *N*-bromosuccinimide (359 mg, 2.00 mmol) were mixed with α,α,α-trifluorotoluene (10 mL), and heated to 120 °C for 30 minutes using microwave

radiation. The reaction mixture was concentrated, redissolved in acetone (10 mL) and AgNO₃ (s) (510 mg, 3.00 mmol) added. The mixture was stirred at room temperature under light-excluding conditions for 24 hours. The mixture was filtered and concentrated. The crude product was purified using column chromatography with a stepwise gradient from 17:3 petroleum ether: EtOAc to 81:19 petroleum ether: EtOAc to afford 3-(*tert*-butyl) 4-methyl (4*S*,5*R*)-5-(4-((*tert*-butoxycarbonyl)oxy)phenyl)-2-oxooxazolidine-3,4-dicarboxylate (92 mg, 21%) as a yellow oil.

¹H NMR (400 MHz, CDCl₃): δ 7.40 (2H, d, 8.3 Hz, OCCH), 7.25 (2H, d, 5.7 Hz, OCCHCH), 5.39 (1H, d, 4.1 Hz, NCHCOOCH₃), 4.62 (1H, d, 4.2 Hz, OCHCH), 3.89 (3H, s, COOCH₃), 1.57 (9H, s, C(CH₃)₃), 1.50 (9H, s, C(CH₃)₃); LRMS (ESI): calculated for C₂₁H₂₇NO₉Na [M+Na]⁺: 460.2, found: 460.2.

The data are in accordance with those reported in the literature.²⁹⁴

7.1.3.65 Synthesis of (2*S*,3*R*)-2-((*tert*-butoxycarbonyl)amino)-3-hydroxy-3-(4-hydroxyphenyl)propanoic acid (186)

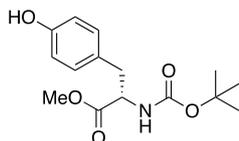


3-(*tert*-butyl) 4-methyl (4*S*,5*R*)-5-(4-((*tert*-butoxycarbonyl)oxy)phenyl)-2-oxooxazolidine-3,4-dicarboxylate (96 mg, 0.22 mmol) was dissolved in 3:1 MeOH:H₂O (8 mL) and Cs₂CO₃ (65 mg, 0.22 mmol) added with stirring at room temperature for 24 hours. The solvent was removed *in vacuo* and the residue redissolved in EtOAc (25 mL). The organic phase was washed with sat. NH₄Cl (aq) (2 x 10 mL), H₂O (2 x 10 mL) and sat. NaCl (aq) (2 x 10 mL). The organic phase was dried with MgSO₄ (s), filtered, and concentrated. The crude was purified using column chromatography with an isocratic elution using EtOAc to give *N*-Boc-β-OH-Tyr-OH (21 mg, 33%). R_f = 0.14 (EtOAc);

¹H NMR (400 MHz, CDCl₃): δ 7.39 (2H, d, 8.3 Hz, OCCH), 7.15 (2H, d, 8.0 Hz, OCCHCH), 5.47 – 5.33 (1H, m, CHOHCH), 4.62 – 4.50 (1H, m, NHCHCOOH), 1.54 (9H, s, C(CH₃)₃), 1.34 (9H, s, C(CH₃)₃); ¹³C NMR (100 MHz, CDCl₃): δ 175.1

(COOH), 155.6 (NHCOO), 150.8 (OCOO), 137.1 (OCCHCHC), 127.3 (OCCHCH), 121.2 (OCCH), 83.6 (COOC(CH₃)₃), 80.1 (COOC(CH₃)₃), 73.1 (CHOH), 59.7 (NHCHCOOH), 28.1 (C(CH₃)₃); **LRMS (ESI)**: calculated for C₁₉H₂₆NO₈ [M-H]⁻: 396.2, found: 396.2.

7.1.3.66 Synthesis of methyl (*tert*-butoxycarbonyl)-L-tyrosinate (182)

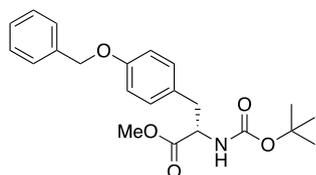


L-Tyr-OMe HCl (4.03 g, 17.37 mmol) was dissolved in MeOH (20 mL), and Et₃N (7.26 mL, 52.11 mmol) and Boc₂O (4.17 g, 19.11 mmol) added at 0 °C. The reaction was returned to room temperature and stirred for 2 hours. The mixture was concentrated and the crude product purified by column chromatography using a isocratic elution of 1:1 petroleum ether: EtOAc to afford *N*-Boc-L-Tyr-OMe (4.87 g, 95%).

¹H NMR (300 MHz, CDCl₃): δ 6.99 (2H, d, 8.1 Hz, HOCCHCH), 6.75 (2H, d, 8.4 Hz, HOCCHCH), 4.58 – 4.48 (1H, m, COCHNH), 3.71 (3H, s, COOCH₃), 3.50 – 3.48 (1H, m, CHCH₂Ar), 3.05 – 2.95 (1H, m, CHCH₂Ar), 1.42 (9H, s, C(CH₃)₃); **LRMS (ESI)**: calculated for C₁₅H₂₂NO₅ [M+H]⁺: 296.3, found: 296.4.

The data are in accordance with those reported in the literature.³²⁷

7.1.3.67 Synthesis of methyl (*S*)-3-(4-(benzyloxy)phenyl)-2-((*tert*-butoxycarbonyl)amino)propanoate (187)



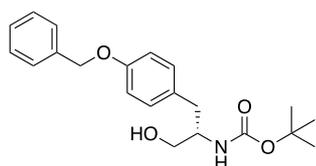
N-Boc-L-Tyr-OMe (4.89 g, 16.54 mmol) and K₂CO₃ (2.74 g, 19.85 mmol) under an argon atmosphere were suspended in anhydrous THF (30 mL). Benzyl bromide (2.36 mL, 19.85 mmol) was added, and the reaction heated to 60 °C for 14 hours. The mixture was concentrated, redissolved in Et₂O (50 mL) and filtered to remove

potassium salts. The crude filtrate was concentrated *in vacuo* and purified by column chromatography with a stepwise solvent gradient (from 4:1 petroleum ether: EtOAc to 2:1 petroleum ether: EtOAc) to afford methyl (*S*)-3-(4-(benzyloxy)phenyl)-2-((*tert*-butoxycarbonyl)amino)propanoate (5.61g, 88%).

$^1\text{H NMR}$ (300 MHz, CDCl_3): δ 7.45 – 7.31 (5H, m, ArH), 7.07 (2H, d, 8.3 Hz, OCCHCH), 6.92 (2H, d, 8.4 Hz, OCCH), 5.04 (2H, s, OCH_2Ar), 4.60 – 4.55 (1H, m, CHCH $_2$ Ar), 3.73 (3H, s, COOCH_3), 3.05 – 3.01 (2H, m, CHCH $_2$ Ar) 1.42 (9H, s, $\text{C}(\text{CH}_3)_3$); **LRMS (ESI)**: calculated for $\text{C}_{22}\text{H}_{28}\text{NO}_5\text{Na}$ $[\text{M}+\text{Na}]^+$: 408.2.2, found: 408.3.

The data are in accordance with those reported in the literature.³²⁸

7.1.3.68 Synthesis of *tert*-butyl (*S*)-(1-(4-(benzyloxy)phenyl)-3-hydroxypropan-2-yl)carbamate (188)



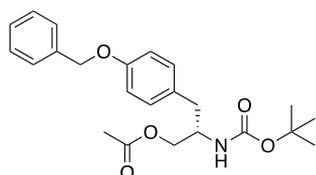
N-Boc-*L*-Tyr(Bzl)-OMe (2.14 g, 5.56 mmol) was dissolved in anhydrous THF (15 mL); LiCl (s) (943 mg, 22.25 mmol) and NaBH $_4$ (842 mg, 22.25 mmol) were added in one portion. The reaction mixture was heated to reflux for 72 hours, then quenched by the addition of MeOH (5 mL) over ice. The mixture was concentrated *in vacuo*, suspended in H $_2$ O (10 mL) and adjusted to pH 6 with 1M HCl. The aqueous phase was extracted with EtOAc (5 x 25 mL), dried with MgSO $_4$ (s), filtered and concentrated. The crude product was purified by column chromatography using a stepwise gradient from 7:3 petroleum ether: EtOAc to 1:1 petroleum ether: EtOAc to afford *tert*-butyl (*S*)-(1-(4-(benzyloxy)phenyl)-3-hydroxypropan-2-yl)carbamate (1.58 g, 80%) as a white solid.

$^1\text{H NMR}$ (400 MHz, CDCl_3): δ 7.45 – 7.29 (5H, m, ArH), 7.12 (2H, d, 7.5 Hz, OCCHCH), 6.92 (2H, d, 7.4 Hz, OCCH), 5.04 (2H, s, OCH_2Ar), 4.72 (1H, d, 7.3 Hz, COCHNH), 3.82 (1H, s, CH $_2$ OH), 3.69 – 3.63 (1H, m, CHCH $_2$ Ar), 3.58 – 3.51 (1H,

m, CHCH₂Ar), 2.77 (2H, d, 6.6 Hz, CH₂OH), 1.42 (9H, s, C(CH₃)₃); **LRMS (ESI)**: calculated for C₂₁H₂₇NO₄Na [M+Na]⁺: 380.2, found: 380.3.

The data are in accordance with those reported in the literature.³²⁹

7.1.3.69 Synthesis of (*S*)-3-(4-(benzyloxy)phenyl)-2-((*tert*-butoxycarbonyl)amino)propyl acetate (189)

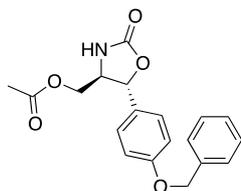


tert-Butyl (*S*)-(1-(4-(benzyloxy)phenyl)-3-hydroxypropan-2-yl)carbamate (1.57 g, 4.45 mmol) under an argon atmosphere was dissolved in anhydrous DCM (20 mL), and Et₃N (1.24 mL, 8.90 mmol) and DMAP (5 mg, 0.04 mmol) added. Acetyl chloride (0.35 mL, 4.90 mmol) was added in dropwise at room temperature, and the solution stirred for 16 hours. The reaction mixture was then diluted with DCM (20 mL), and the organic phase washed with 1M HCl (2 x 20 mL), sat. NaHCO₃ (aq) (20 mL), and sat. NaCl (aq) (2 x 20 mL). The organic layer was dried with MgSO₄ (s), filtered and concentrated. The crude product was purified by column chromatography using an isocratic elution with 4:1 petroleum ether: EtOAc to afford (*S*)-3-(4-(benzyloxy)phenyl)-2-((*tert*-butoxycarbonyl)amino)propyl acetate (1.961 g, 95%) as a white solid.

¹H NMR (400 MHz, CDCl₃): δ 7.45 – 7.29 (5H, m, ArH), 7.09 (2H, d, 8.1 Hz, OCCH₂CH), 6.91 (2H, d, 8.2 Hz, OCCH), 5.02 (2H, m, OCH₂Ar), 4.67 – 4.59 (1H, m, COCHNH), 4.02 (3H, s, CH₃CO), 2.85 – 2.68 (2H, m, CH₂OCO), 1.42 (9H, s, C(CH₃)₃); **LRMS (ESI)**: calculated for C₂₃H₂₉NO₅Na [M+Na]⁺: 422.2, found: 422.3.

The data are in accordance with those reported in the literature.³³⁰

7.1.3.70 Synthesis of ((4*R*,5*R*)-5-(4-(benzyloxy)phenyl)-2-oxooxazolidin-4-yl)methyl acetate (190)

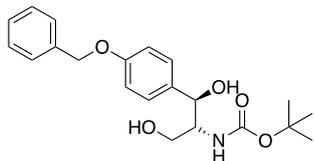


(*S*)-3-(4-(Benzyloxy)phenyl)-2-((*tert*-butoxycarbonyl)amino)propyl acetate (507 mg, 1.27 mmol) under an argon atmosphere was dissolved in anhydrous MeCN (20 mL). $K_2S_2O_8$ (s) (687 mg, 2.54 mmol) in H_2O (17 mL) was added dropwise, followed by dropwise addition of $CuSO_4 \cdot 5H_2O$ (s) (63 mg, 0.25 mmol) in H_2O (3 mL). The reaction mixture was heated to 70 °C for 2.5 hours. The reaction mixture was cooled and diluted with H_2O (30 mL), and the aqueous layer extracted with EtOAc (5 x 25 mL). The organic phase was dried with $MgSO_4$, filtered and concentrated. The crude product was purified by column chromatography using a stepwise gradient from 1:1 petroleum ether: EtOAc to 2:3 petroleum ether: EtOAc to afford ((4*R*,5*R*)-5-(4-(benzyloxy)phenyl)-2-oxooxazolidin-4-yl)methyl acetate (184 mg, 42%) as a pale yellow solid.

1H NMR (400 MHz, $CDCl_3$): δ 7.46 – 7.30 (7H, m, ArH, OCCHCH), 7.01 (2H, d, 7.8 Hz, OCCHCH), 5.21 (1H, d, 5.8 Hz, NHCHCH₂), 5.08 (2H, s, OCH₂Ar), 4.33 (1H, dd, 2.9 Hz, 11.4 Hz, CH₂OCO), 4.10 (1H, dd, 5.9 Hz, 11.6 Hz, CH₂OCO), 4.00 – 3.92 (1H, m, NHCH), 2.10 (3H, s, OCOCH₃); LRMS (ESI): calculated for $C_{19}H_{19}NO_5Na$ $[M+Na]^+$: 364.1, found: 364.2.

The data are in accordance with those reported in the literature.³³⁰

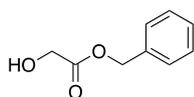
7.1.3.71 Synthesis of *tert*-butyl ((1*R*,2*R*)-1-(4-(benzyloxy)phenyl)-1,3-dihydroxypropan-2-yl)carbamate (192)



((4*R*,5*R*)-5-(4-(Benzyloxy)phenyl)-2-oxooxazolidin-4-yl)methyl acetate (146 mg, 0.43 mmol) and Ba(OH)₂ (123 mg, 0.65 mmol) were dissolved in 4:1 EtOH:H₂O (10 mL) and heated to 80 °C for 16 hours. The mixture was concentrated *in vacuo*, and the residue redissolved in 1:1 THF: H₂O (10 mL). NaHCO₃ (s) (72 mg, 0.86 mmol) and Boc₂O (103 mg, 0.47 mmol) were added, and stirred at room temperature for 16 hours. The mixture was concentrated *in vacuo*, diluted in H₂O (20 mL) and 1M HCl (20 mL), and the aqueous phase extracted with CHCl₃ (3 x 10 mL). The combined organic phase was dried with MgSO₄ (s), filtered and concentrated. The crude product was purified by column chromatography using an isocratic elution of 2:3 petroleum ether: EtOAc to afford *tert*-butyl ((1*R*,2*R*)-1-(4-(benzyloxy)phenyl)-1,3-dihydroxypropan-2-yl)carbamate (42 mg, 26%). **R_f** = 0.72 (EtOAc); **IR** (thin film) ν_{\max} : 3320 (O-H stretch), 2928 (C-H stretch), 1237 (C-O stretch), 1011 (C-O stretch);

¹H NMR (500 MHz, MeOD): δ 7.44 (2H, 7.4 Hz, ArH), 7.37 (2H, t, 7.5 Hz, ArH), 7.33 – 7.28 (3H, m, ArH, OCCH), 6.97 (2H, d, 8.5 Hz, OCCHCH), 5.09 (2H, s, OCH₂), 4.85 (1H, d, 3.9 Hz, HOCH), 3.73 (1H, dd, 5.6 Hz, 10.2 Hz, CHNH), 3.65 (1H, dd, 6.4 Hz, 10.6 Hz, CH₂OH), 3.46 (1H, dd, 5.7 Hz, 10.7 Hz, CH₂OH), 1.38 (9H, s, C(CH₃)₃); **¹³C NMR** (125 MHz, MeOD): δ 159.5 (NHCOO), 158.3 (OCCH), 138.9 (CCH), 136.2 (CCH), 129.5 (CH), 128.8 (CH), 128.6 (CH), 128.5 (CH), 115.6 (CH), 80.2 (C(CH₃)₃), 72.7 (HOCH), 71.0 (OCH₂), 62.9 (CH₂OH), 59.2 (CHNH), 28.7 (C(CH₃)₃); **HRMS (ESI)**: calculated for C₂₁H₂₇NO₅Na [M+Na]⁺: 396.1781, found: 396.1779; [α]_D²⁸: -21.5 (0.086, MeOH).

7.1.3.72 Synthesis of benzyl 2-hydroxyacetate (**199**)

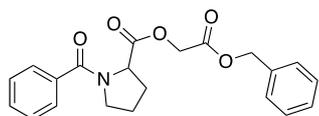


To a solution of glycolic acid (200 mg, 2.63 mmol) and benzyl bromide (250 μ L, 2.10 mmol) in acetonitrile (4 mL) at 0 °C under an inert atmosphere, was added DBU (310 μ L, 2.10 mmol). The solution was allowed to return to room temperature overnight. The solvent was then removed *in vacuo* and the residue dissolved in EtOAc. The organic phase was washed with 1M HCl (10 mL), sat. NaCl (2 x 10 mL), dried with MgSO₄, filtered and concentrated to afford **199** as a colourless oil (265 mg, 76%).

¹H NMR (300 MHz, CDCl₃): δ 7.41 – 7.32 (5H, s, ArH), 5.24 (2H, s, CH₂), 4.20 (2H, d, 5.4 Hz, CH₂OH), 2.34 (1H, t, 5.4 Hz, CH₂OH); LRMS (ESI): calculated for C₉H₁₀O₃Na: 189.1, found: 188.7.

The data are in accordance with those reported in the literature.³³¹

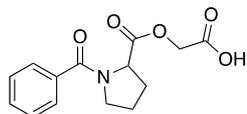
7.1.3.73 Synthesis of 2-(benzyloxy)-2-oxoethyl benzoylprolinate (**200**)



Benzyl 2-hydroxyacetate (18 mg, 0.109 mmol), 1-benzoyl-pyrrolidine-2-carboxylic acid (20 mg, 0.091 mmol) and 4-dimethylaminopyridine (DMAP, 2.2 mg, 0.018 mmol) under an argon atmosphere were dissolved in anhydrous DCM (1 mL). 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC, 21 μ L, 0.119 mmol) was added, and the mixture stirred at room temperature overnight. The mixture was then diluted with DCM and washed with sat. NaHCO₃ (5 mL), 0.1 M HCl (5 mL) and brine (5 mL). The organic layer was dried on MgSO_{4(s)}, filtered and concentrated. The residue was purified using silica gel chromatography (eluting from 3:1 Petroleum ether:EtOAc to 3:2 Petroleum ether:EtOAc) to afford 2-(benzyloxy)-2-oxoethyl benzoylprolinate as a white solid (16 mg, 49 %). R_f = 0.28 (2: 3 EtOAc: petroleum ether).

¹H NMR (500 MHz, MeOD): δ 7.57 – 7.38 (5H, m, ArH), 4.74 (1H, d, 15.8 Hz, COCH₂O), 4.69 (1H, dd, 8.5 Hz, 5.2 Hz, NCHCO), 4.58 (1H, d, 15.8 Hz, COCH₂O), 3.65- 3.53 (2H, m, NCH₂), 2.43- 2.10 (2H, m, NCHCH₂), 2.09- 1.79 (2H, m, NCH₂CH₂); **¹³C NMR** (125 MHz, MeOD): δ 173.2 (COOH), 173.2 (CHCOO), 137.2 (CH₂CCHCH), 131.6 (Ar), 129.6 (Ar), 128.1 (Ar), 62.9 (CH₂COOH), 60.6 (NCH), 51.3 (NCH₂), 30.2 (NCHCH₂), 26.1 (NCH₂CH₂); **HRMS(ESI)**: calculated for C₂₁H₂₁NO₅Na [M+Na]⁺: 390.1312, found: 390.1312.

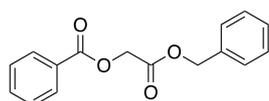
7.1.3.74 Synthesis of 2-((benzoylpropyl)oxy)acetic acid (201)



2-(Benzoyloxy)-2-oxoethyl benzoyl-L -prolinate (16 mg, 0.058 mmol) and 10% Pd/C (5 mg, 0.046 mmol) under an argon atmosphere were suspended in dry MeOH (2 mL), and the atmosphere replaced with hydrogen. The mixture was stirred at room temperature overnight, then filtered through Celite and concentrated to afford 2-((benzoyl-L-prolyl)oxy)acetic acid as a colourless oil (12 mg, 98 %).

¹H NMR (500 MHz, MeOD): δ 7.57- 7.38 (5H, m, ArH), 4.74 (1H, d, 15.8 Hz, CH₂COOH), 4.58 (1H, d, 15.8 Hz, CH₂COOH), 4.69 (1H, dd, 5.2 Hz, 8.5 Hz, NCHCOO), 3.65– 3.53 (2H, m, CH₂N), 3.43 – 2.19 (2H, m, CH₂CHN), 2.09 – 1.79 (2H, m, CH₂CH₂N); **¹³C NMR** (125 MHz, MeOD) δ 173.2 (CH₂COOH), 173.2 (NCHCOO), 173.1 (ArCON), 137.2 (CHCHCCO), 131.6 (CHCCO), 129.6 (CHCHCHCCO), 128.1 (CHCHCCO), 62.9 (CH₂COOH), 60.6 (NCHCOO), 46.9 (CH₂N), 29.7 (CH₂CHN), 25.5 (CH₂CH₂N); **HRMS(ESI)**: calculated for C₁₄H₁₅NO₅Na [M+Na]⁺: 300.0842, found: 300.0843.

7.1.3.75 Synthesis of 2-(benzyloxy)-2-oxoethyl benzoate (203)



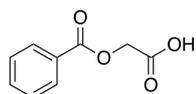
Benzyl-2-hydroxyacetate (1.795 g, 10.80 mmol) was dissolved in DCM (30 mL) and cooled to 0 °C. Et₃N (1.32 mL, 11.88 mmol) was added, and then benzoyl chloride

(1.32 mL, 11.34 mmol) was added dropwise. The reaction was allowed to return to room temperature over 5 hours. The mixture was then concentrated, and the residue dissolved in DCM (30 mL); the organic layer was washed with sat. NaHCO₃ (aq) (30 mL), 1M HCl (30 mL) and sat. NaCl (aq) (30 mL). The organic phase was isolated, dried with MgSO₄ (s), filtered and concentrated. The crude product was purified by column chromatography, eluting with toluene to afford 2-(benzyloxy)-2-oxoethyl benzoate as white solid (2.45 g, 84%).

¹H NMR (300 MHz, CDCl₃): δ 8.13 – 8.08 (2H, m, ArH), 7.50 – 7.42 (1H, m, ArH), 7.53 – 7.46 (2H, m, ArH), 7.36 (5H, s, ArH), 5.24 (2H, s, CH₂OCO), 4.90 (2H, s, COCH₂O); LRMS (ESI): calculated for C₁₆H₁₄O₄Na [M+Na]⁺: 293.1, found: 292.8.

The data are in accordance with those reported in the literature.³³²

7.1.3.76 Synthesis of 2-(benzyloxy)acetic acid (204)

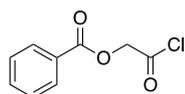


2-(Benzyloxy)-2-oxoethyl benzoate (162 mg, 0.60 mmol) and Pd/C (10% wt) (51 mg, 0.48 mmol) was suspended in anhydrous MeOH (5 mL) under an argon atmosphere, and the solution degassed. H₂ (g) was bubbled through, and the reaction was stirred overnight under an H₂ (g) atmosphere. The reaction mixture was filtered through Celite and the filtrate concentrated to afford **204** as a white solid (100 mg, 100%).

¹H NMR (300 MHz, MeOD): δ 8.16 – 8.00 (2H, m, ArH), 7.64 – 7.55 (1H, m, ArH), 7.54 – 7.39 (2H, m, ArH), 4.87 (2H, s, COCH₂O); LRMS (ESI): calculated for C₉H₇O₄ [M-H]⁻: 179.0, found: 179.0.

The data are in accordance with those reported in the literature.³³²

7.1.3.77 Synthesis of 2-chloro-2-oxoethyl benzoate (205)



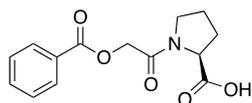
2-(Benzyloxy)acetic acid (108 mg, 0.60 mmol) was dissolved in toluene (3 mL) and thionyl chloride (218 μL, 3.00 mmol) added under an inert atmosphere. The reaction

was heated to reflux overnight, and thionyl chloride removed *in vacuo* to afford **205** as a brown oil (112 mg, 94%). The compound was judged pure by ^1H NMR analysis.

^1H NMR (300 MHz, CDCl_3): δ 8.11 – 8.06 (2H, m, ArH), 7.67 – 7.60 (1H, m, ArH), 7.52 – 7.45 (2H, m, ArH), 5.15 (OCH_2COCl).

The data are in accordance with those reported in the literature.³³³

7.1.3.78 Synthesis of (2-(benzoyloxy)acetyl)-L-proline (196)

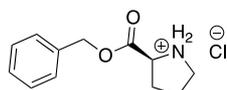


L-Proline (151 mg, 1.32 mmol) was suspended in DCM (10 mL) under an inert atmosphere, Et_3N (203 μL , 1.46 mmol) was added and the mixture was cooled to 0 $^\circ\text{C}$ and stirred. 2-chloro-2-oxoethyl benzoate (276 mg, 1.39 mmol) was dissolved in DCM (6 mL) and added dropwise, and the mixture stirred at room temperature for 3 hours. The reaction mixture was washed with 1M HCl (15 mL) and sat. $\text{NaCl}_{(\text{aq})}$ (15 mL), dried with $\text{MgSO}_4_{(\text{s})}$, filtered and concentrated *in vacuo*. The crude product was purified by column chromatography, eluting with pure EtOAc to afford scleric acid, **196**, as a white solid (269 mg, 73%). $R_f = 0.23$ in 1:9 DCM: MeOH. Before use in biological experiments, **196** was further purified by HPLC. Solvent A was H_2O with 0.1% formic acid, and solvent B was MeOH with 0.1% formic acid. The gradient was from 5% B to 95% B over 50 minutes, then to 5% B over 10 minutes. Scleric acid had a retention time of 30.8 minutes.

Major isomer ^1H NMR (700 MHz, MeOD): δ 8.10 – 8.05 (2H, m, COCCH), 7.64 – 7.59 (1H, m, COCCHCHCH), 7.51 – 7.46 (2H, m, COCCHCH), 5.10 (1H, d, 15.1 Hz, COCH_2O), 4.95 (1H, d, 15.1 Hz, COCH_2O), 4.48 (1H, dd, 9.0 Hz, 3.4 Hz, NCH), 3.70 – 3.63 (2H, m, NCH_2), 2.31 – 2.24 (1H, m, NCHCH_2), 2.10 – 2.01 (3H, m, NCHCH_2 , $\text{NCHCH}_2\text{CH}_2$); ^{13}C NMR (125 MHz, MeOD): 175.4 (CO_2H), 168.1 (NCO), 167.5 ($\text{OCOC}(\text{CH})_2$), 134.5 (COCCHCHCH), 130.8 (COCCH), 130.8 (OCOCCH), 129.6 (COCCHCH), 63.1 (COCH_2O), 60.5 (NCH), 47.1 (NCH_2), 30.0 (NCHCH_2), 25.7 (NCH_2CH_2);

Minor isomer $^1\text{H NMR}$ (700 MHz, MeOD): δ 8.10 – 8.05 (2H, m, COCCH), 7.64 – 7.59 (1H, m, COCCHCHCH), 7.51 – 7.46 (2H, m, COCCHCH), 5.04 (1H, dd, 14.7 Hz, COCH₂O), 4.80 (1H, dd, 14.7 Hz, COCH₂O), 4.69 (1H, dd, 2.3 Hz, 8.5 Hz, NCH), 3.60 (2H, m, NCH₂), 2.37 (1H, m, NCHCH₂), 1.93 (3H, m, NCHCH₂, NCH₂CH₂); $^{13}\text{C NMR}$ (125 MHz, MeOD): 174.8 (CO₂H), 168.6 (NCO), 167.5 (OCOC(CH)₂), 134.5 (COCCHCHCH), 130.8 (COCCH), 130.8 (OCOCCH), 129.6 (COCCHCH), 63.3 (COCH₂O), 60.1 (NCH), 48.0 (NCH₂), 32.3 (NCHCH₂), 23.1 (NCH₂CH₂); **HRMS (ESI)**: calculated for C₁₄H₁₅O₅NNa [M+Na]⁺: 300.0842, found: 300.0842. $[\alpha]_{\text{D}}^{20} = -65$ (c 0.195, MeOH).

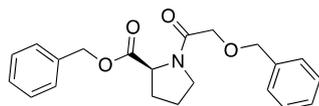
7.1.3.79 Synthesis of (S)-2-((benzyloxy)carbonyl)pyrrolidin-1-ium chloride (207)



Benzyl alcohol (13.5 mL, 130.4 mmol) was cooled to 0 °C under an inert atmosphere, and SOCl₂ (1.27 mL, 17.4 mmol) added dropwise. L-Proline (1.00 g, 8.69 mmol) was added, and the solution stirred at 0 °C for 2 hours and then for 48 hours at room temperature. Diethyl ether (100 mL) was added to precipitate the product, with the flask cooled to -20 °C for 24 hours. The precipitate was collected by filtration to afford a white solid (1.653 g, 79%). $^1\text{H NMR}$ (300 MHz, MeOD): δ 7.42 – 7.37 (5H, m, ArH), 5.29 (2H, dd, 12.1 Hz, 4.0 Hz, ArCH₂O), 4.49 (1H, t, 7.7 Hz, OOCCHNH), 3.41 – 3.33 (2H, m, NCH₂), 2.47 – 2.40 (1H, m, NHCHCH₂), 2.11 – 2.03 (3H, m, NHCHCH₂, NHCH₂); **LRMS (ESI)**: calculated for C₁₂H₁₆O₂N [M+H]⁺: 206.1, found: 205.6.

The data are in accordance with those reported in the literature.³³⁴

7.1.3.80 Synthesis of benzyl (2-(benzyloxy)acetyl)-L-prolinate (208)

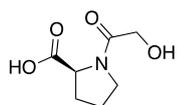


Benzyl L-Proline hydrochloride (500 mg, 2.07 mmol) was suspended in DCM (10 mL) under an inert atmosphere, and Et₃N (606 μL, 4.35 mmol) added. Benzyloxyacetyl chloride (402 μL, 2.18 mmol) was added dropwise, and the reaction returned to room temperature and stirred for 16 hours. The reaction mixture was concentrated, redissolved in EtOAc (25 mL) and washed with 1M HCl (20 mL), sat. NaCl_(aq) (20 mL) and sat. NaHCO_{3(aq)} (20 mL). The crude product was isolated after concentrating and drying the organic layer with MgSO_{4(s)}; purification of the crude residue by column chromatography, eluting with 3:1 petroleum ether: EtOAc to 1:1 petroleum ether: EtOAc, afforded the desired product as a colourless oil (681 mg, 93%) with a 3:1 mixture of isomers. **R_f**: 0.30 in 1:1 petroleum ether: EtOAc; **IR** (thin film) ν_{max} : 1739 (C=O stretch), 1648 (C=O stretch), 1167 (C-O stretch), 736 (C-H bend);

Major isomer ¹H NMR (500 MHz, CDCl₃): δ 7.37 – 7.27 (10H, m, ArH), 5.18 (2H, dd, 12.3 Hz, 8.9 Hz, CCH₂OCO), 4.62 – 4.58 (2H, m, CH₂OCH₂Ar), 4.61 – 4.58 (1H, m, NHCHCOO), 4.16 (2H, dd, 14.3 Hz, 3.2 Hz, COCH₂O), 3.65 – 3.59 (1H, m, NCH₂), 3.55 – 3.49 (1H, m, NCH₂), 2.24 – 2.14 (1H, m, NCHCH₂), 2.06 – 1.92 (3H, m, NCH₂CH₂, NCHCH₂); ¹³C NMR (125 MHz, CDCl₃): δ 172.1 (CHCOO), 168.5 (NCOCH₂), 137.5 (CH₂OCH₂C(CH)₂), 135.8 (COOCH₂C(CH)₂), 128.8 (CH), 128.7 (CH), 128.6 (CH), 128.6 (CH), 128.5 (CH), 128.4 (CH), 128.3 (CH), 128.2 (CH), 128.1 (CH), 128.0 (CH), 73.2 (CH₂OCH₂C(CH)₂), 69.6 (NCOCH₂), 67.0 (COOCH₂C(CH)₂), 59.2 (NCH), 46.5 (NCH₂), 28.9 (NCHCH₂), 25.1 (NCH₂CH₂);

Minor isomer ¹H NMR (500 MHz, CDCl₃): δ 7.37 – 7.27 (10H, m, ArH), 5.04 (1H, d, 12.1 Hz, CCH₂OCO), 4.92 (1H, d, 12.2 Hz, CCH₂OCO), 4.63 – 4.59 (1H, m, NHCHCOO), 4.45 – 4.39 (2H, m, CH₂OCH₂Ar), 4.04 (2H, s, COCH₂O), 3.71 – 3.5 (2H, m, NCH₂), 2.14 – 2.08 (1H, m, NCHCH₂), 1.90 – 1.81 (2H, m, NCHCH₂NCH₂CH₂); ¹³C NMR (125 MHz, CDCl₃): δ 172.1 (CHCOO), 168.6 (NCOCH₂), 137.2 (CH₂OCH₂C(CH)₂), 135.5 (COOCH₂C(CH)₂), 128.8 (CH), 128.7 (CH), 128.6 (CH), 128.6 (CH), 128.5 (CH), 128.4 (CH), 128.3 (CH), 128.2 (CH), 128.1 (CH), 128.0 (CH), 73.4 (CH₂OCH₂C(CH)₂), 70.7 (NCOCH₂), 67.1 (COOCH₂C(CH)₂), 59.2 (NCH), 47.1 (NCH₂), 31.7 (NCHCH₂), 22.0 (NCH₂CH₂); **HRMS (ESI)**: calculated for C₂₁H₂₃O₄NNa [M+Na]⁺: 376.1519, found: 376.1522. [α]_D²⁸: -41.2 (0.073, MeOH).

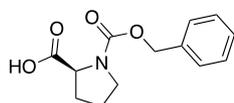
7.1.3.81 Synthesis of (2-hydroxyacetyl)-L-proline (202)



Anhydrous MeOH (5 mL) was added to benzyl (2-(benzyloxy)acetyl)-L-prolinate (51 mg, 0.141 mmol), Pd/C (10%) (12 mg, 0.113 mmol) and ammonium formate (46 mg, 0.707 mmol) under an inert atmosphere, and the reaction heated to reflux for 4 hours. The mixture was cooled and filtered through Celite. The filtrate was concentrated to afford a white powder (23 mg, 92%) with a 1:1 mixture of isomers.

$^1\text{H NMR}$ (400 MHz, MeOD): δ 4.36 (1H, dd, 3.2 Hz, 8.7 Hz, NCH), 4.19 (2H, dd, 15.6 Hz, 6.5 Hz, COCH₂OH), 4.16 (1H, dd, 3.4 Hz, 7.7 Hz, NCH), 4.09 (2H, dd, 15.4 Hz, 16.2 Hz, COCH₂OH), 3.67 – 3.49 (2H, m, NCH₂), 3.57 – 3.38 (2H, m, NCH₂), 2.30 – 2.11 (2H, m, NCHCH₂), 2.19 – 1.83 (2H, m, NCH₂CH₂); $^{13}\text{C NMR}$ (100 MHz, MeOD): δ 179.3 (COOH), 178.8 (COOH), 172.9 (NCOCH₂), 172.1 (NCOCH₂), 62.8 (NCH), 62.2 (NCH), 61.8 (COCH₂OH), 61.5 (COCH₂OH), 47.9 (NCH₂), 46.6 (NCH₂), 33.0 (NCHCH₂), 30.7 (NCHCH₂), 25.6 (NCH₂CH₂), 23.4 (NCH₂CH₂); **HRMS (ESI)**: calculated for C₇H₁₁O₄NNa: 196.0580, found: 196.0583.

7.1.3.82 Synthesis of ((benzyloxy)carbonyl)-L-proline (209)



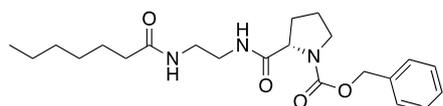
Benzyl chloroformate (1.02 ml, 7.17 mmol) was added dropwise to a solution of L-proline (750 mg, 6.51 mmol) in 1M NaOH (8 mL) at 0 °C. The mixture was stirred at room temperature for 16 hours. The reaction was acidified to pH 3 with 1M HCl, and the aqueous layer extracted with EtOAc (3 x 10 mL). The combined organic phase was dried with MgSO₄ (s), filtered and concentrated. The crude product was purified by column chromatography using a stepwise gradient from 1:1 petroleum ether: EtOAc to EtOAc to afford ((benzyloxy)carbonyl)-L-proline (1.232 g, 76%) as a colourless gum.

$^1\text{H NMR}$ (400 MHz, CDCl₃): δ 7.39 – 7.23 (5H, m, ArH), 5.13 (2H, q, 12.1 Hz, CH₂Ar), 4.70 (1H, dd, 5.2 Hz, 8.4 Hz, NHCH), 3.61 – 3.41 (2H, m, NCH₂), 2.31 –

2.05 (2H, m, CH₂), 1.96 – 1.72 (2H, m, CH₂); LRMS (ESI): calculated for C₁₃H₁₆NO₄: 248.1, found: 248.1.

The data are in accordance with those reported in the literature.

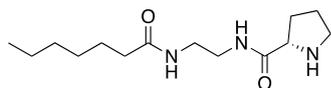
7.1.3.83 Synthesis of benzyl (S)-2-((2-heptanamidoethyl)carbamoyl)pyrrolidine-1-carboxylate (210)



N-(2-aminoethyl)heptanamide (141 mg, 0.82 mmol) and ((benzyloxy)carbonyl)-L-proline (265 mg, 1.06 mmol) under an argon atmosphere were dissolved in anhydrous THF (15 mL), cooled to 0 °C and DIPEA (428 μl, 2.46 mmol) was added. After 10 minutes HATU (404 mg, 1.06 mmol) was added and the reaction stirred at room temperature for 16 hours. The solvent was removed *in vacuo*, and the residue redissolved in EtOAc (30 mL). The organic phase was washed with 1M HCl (15 ml), sat. NaHCO₃ (aq) (15 ml) and sat. NaCl (aq) (15 ml), then dried with MgSO₄ (s), filtered, and concentrated. The crude product was purified by column chromatography using a stepwise gradient from DCM to 19:1 DCM: MeOH to afford benzyl (S)-2-((2-heptanamidoethyl)carbamoyl)pyrrolidine-1-carboxylate (302 mg, 92%) as a colourless gum. *R*_f = 0.10 (EtOAc).

¹H NMR (500 MHz, CDCl₃): δ 7.36 (5H, m, ArH), 6.86, 6.63 (1H, s, CONH), 6.29, 5.95 (1H, s, CONH), 5.23 – 5.17 (2H, d, 12.1 Hz, COOCH₂), 5.13 – 5.03 (2H, d, 12.1 Hz, COOCH₂), 4.26 (1H, d, 4.9 Hz, COCHN), 3.60 – 3.43 (2H, m, OCONCH₂), 3.43 – 3.19 (4H, m, NHCH₂CH₂NH), 2.27 – 1.85 (6H, m, COCHCH₂CH₂, CH₂CONH), 1.63 – 1.53 (2H, m, CH₂CH₂CONH), 1.31 – 1.23 (6H, m, CH₂), 0.86 (3H, t, 6.2 Hz, CH₃); ¹³C NMR (125 MHz, MeOD): δ 174.4 (CH₂CONH), 173.1 (NHCOCH), 156.1 (COO), 136.5 (COOCH₂C), 128.7 (Ar), 128.3 (Ar), 127.9 (Ar), 67.5, 67.3 (COOCH₂), 61.1 (COCHN), 47.6, 47.3 (COCHNCH₂), 40.1, 39.8 (NHCH₂CH₂NH), 36.7 (CH₂CH₂CO), 31.7 (CH₂), 29.2 (COCHCH₂), 29.1 (CH₂), 25.8 (CH₂CH₂CO), 24.7 (COCHCH₂CH₂), 22.6 (CH₂), 14.2 (CH₃); HRMS(ESI): calculated for C₂₂H₃₃N₃O₄Na [M+Na]⁺: 426.2363, found: 426.2361.

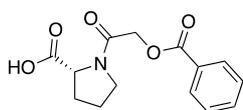
7.1.3.84 Synthesis of (*S*)-*N*-(2-heptanamidoethyl)pyrrolidine-2-carboxamide (211)



Benzyl (*S*)-2-((2-heptanamidoethyl)carbamoyl)pyrrolidine-1-carboxylate (75 mg, 0.19 mmol) and Pd/C (10%) (16 mg, 0.15 mmol) under an argon atmosphere were suspended in anhydrous MeOH (5 ml) and degassed. Hydrogen gas was bubbled through the solution for 4 hours. The reaction mixture was filtered through Celite and concentrated to afford (*S*)-*N*-(2-heptanamidoethyl)pyrrolidine-2-carboxamide (48 mg, 95%) as a colourless solid. $R_f = 0.31$ (1:1 MeOH: DCM); **IR** (thin film) ν_{\max} : 3279 (N-H stretch), 2928 (C-H stretch), 1633 (C=O stretch);

^1H NMR (500 MHz,): δ 3.62 (1H, dd, 5.2 Hz, 8.7 Hz, COCHNH), 3.34 – 3.28 (4H, m, NHCH₂CH₂NH), 2.95 (2H, qt, 6.2 Hz, 10.3 Hz, COCHNHCH₂), 2.18 (2H, t, 7.6 Hz, CH₂CH₂CO), 2.15 – 2.08 (1H, m, COCHCH₂), 1.81 – 1.71 (3H, m, COCHCH₂CH₂), 1.60 (2H, quin, 7.4 Hz, CH₂CH₂CO), 1.38 – 1.28 (6H, m, CH₂), 0.91 (3H, t, 6.8 Hz, CH₃CH₂); **^{13}C NMR** (125 MHz,): δ 177.8 (NHCOCH), 176.7 (CH₂CONH), 61.7 (COCHNH), 48.0 (CH₂CH₂NH), 39.9 (NHCH₂CH₂NH), 37.2 (CH₂CH₂CO), 32.7 (CH₂), 32.0 (COCHCH₂), 30.0 (CH₂), 27.0 (CHCH₂CH₂), 27.0 (CH₂CH₂CO), 23.6 (CH₂), 14.4 (CH₃CH₂); **HRMS(ESI)**: calculated for C₁₄H₂₈N₃O₂ [M+H]⁺: 270.2176, found: 270.2180; [α]_D²⁸: -57.4 (0.070, MeOH).

7.1.3.85 Synthesis of (2-(benzoyloxy)acetyl)-*D*-proline (212)



D-proline (64 mg, 0.56 mmol) was suspended in DCM (2 mL) under an inert atmosphere, Et₃N (85 μ L, 0.62 mmol) was added and the mixture was cooled to 0 °C and stirred. 2-chloro-2-oxoethyl benzoate (116 mg, 0.58 mmol) was dissolved in DCM (2 mL) and added dropwise, and the mixture stirred at room temperature for 16 hours. The reaction mixture was washed with 1M HCl (15 mL) and sat. NaCl (aq) (15 mL), dried with MgSO₄ (s), filtered and concentrated *in vacuo*. The crude product was purified by column chromatography, eluting with EtOAc to afford **212** as a white solid

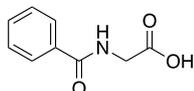
(136 mg, 88%). Before use in biological experiments, **212** was further purified by HPLC. Solvent A was H₂O with 0.1% formic acid, and solvent B was MeCN with 0.1% formic acid. The gradient was from 5% B to 100% B over 45 minutes, then to 5% B over 5 minutes. **212** had a retention time of 21.3 minutes.

IR (thin film) ν_{max} : 1720 (C=O stretch), 1620 (C=O stretch), 1266 (C-O stretch), 708 (C-H bend);

Major isomer **¹H NMR** (700 MHz, MeOD): δ 8.10 – 8.05 (2H, m, COCCH), 7.64 – 7.59 (1H, m, COCCHCHCH), 7.52 – 7.46 (2H, m, COCCHCHCH), 5.10 (1H, d, 15 Hz, COCH₂O), 4.95 (1H, d, 15 Hz, COCH₂O), 4.48 (1H, dd, 9 Hz, 3.4 Hz, NCH), 3.74 – 3.63 (2H, m, NCH₂), 2.34 – 2.23 (1H, m, NCHCH₂), 2.12 – 2.01 (3H, m, NCHCH₂, NCHCH₂CH₂); **¹³C NMR** (125 MHz, MeOD): 175.3 (CO₂H), 168.1 (NCO), 167.5 (OCOC(CH)₂), 134.5 (COCCHCHCH), 130.8 (COCCH), 130.8 (OCOCCH), 129.6 (COCCHCH), 63.1 (COCH₂O), 60.5 (NCH), 47.1 (NCH₂), 30.0 (NCHCH₂), 25.7 (NCH₂CH₂);

Minor isomer **¹H NMR** (700 MHz, MeOD): δ 8.10 – 8.05 (2H, m, COCCH), 7.64 – 7.59 (1H, m, COCCHCHCH), 7.52 – 7.46 (2H, m, COCCHCHCH), 5.02 (1H, d, 14.7 Hz, COCH₂O), 4.77 (1H, d, 14.7 Hz, COCH₂O), 4.70 (1H, dd, 2.3 Hz, 8.5 Hz, NCH), 3.62 – 3.54 (2H, m, NCH₂), 2.41 – 2.34 (1H, m, NCHCH₂), 2.00 – 1.83 (3H, m, NCHCH₂, NCH₂CH₂); **¹³C NMR** (125 MHz, MeOD): 174.8 (CO₂H), 168.6 (NCO), 167.4 (OCOC(CH)₂), 134.5 (COCCHCHCH), 130.8 (COCCH), 130.8 (OCOCCH), 129.6 (COCCHCH), 63.3 (COCH₂O), 60.1 (NCH), 47.9 (NCH₂), 32.3 (NCHCH₂), 23.1 (NCH₂CH₂); $[\alpha]_{\text{D}}^{28}$: +45.0 (0.067, MeOH).

7.1.3.86 Synthesis of benzoylglycine (**219**)



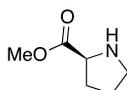
Glycine (1.50 g, 19.98 mmol) was dissolved in 1M NaOH (20 mL), cooled to 0 °C, and benzoyl chloride (2.32 mL, 19.98 mmol) was added dropwise. 1M NaOH was added to maintain pH >10 while stirring at room temperature for 2 hours. 1M HCl (50

mL) was added, and the mixture stirred for 15 minutes. The precipitate was collected by filtration and washed with H₂O and Et₂O, to afford benzoylglycine (3.23 g, 90%).

¹H NMR (300 MHz, MeOD): δ 7.88 (2H, d, 7.7 Hz, ArH), 7.57 (1H, t, 7.2 Hz, ArH), 7.49 (2H, t, 7.4 Hz, ArH), 4.11 (2H, s, NHCH₂COOH); **LRMS (ESI)**: calculated for C₉H₈NO₃ [M-H]⁻: 178.1, found: 178.1.

The data are in accordance with those reported in the literature.³³⁵

7.1.3.87 Synthesis of methyl L-prolinate (220)

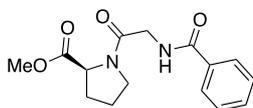


L-Proline (2 g, 17.36 mmol) was dissolved in MeOH (15 mL) and H₂SO₄ (0.5 mL) added. The mixture was heated to reflux for 16 hours. The solvent was removed *in vacuo*, H₂O (10 mL) added, and adjusted to pH 10 with 6M NaOH. The aqueous phase was extracted with EtOAc (5 x 20 mL), adjusted to pH 12, and further extracted with EtOAc (3 x 20 mL). The combined organic phases were dried with MgSO₄(s), filtered and concentrated to afford methyl L-prolinate (1.63 g, 73%).

¹H NMR (300 MHz, CDCl₃): δ 3.81 – 3.77 (1H, m, NHCH), 3.71 (3H, s, COOCH₃), 3.06 – 2.91 (2H, m, NHCH₂), 2.14 – 2.01 (1H, m, NHCHCH₂), 1.92 – 1.61 (3H, m, NHCHCH₂, NHCH₂CH₂); **LRMS (ESI)**: calculated for C₆H₁₂NO₂ [M+H]⁺: 130.1, found: 130.2.

The data are in accordance with those reported in the literature.³³⁶

7.1.3.88 Synthesis of methyl benzoylglycyl-L-prolinate (221)

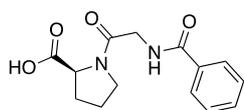


Methyl L-prolinate (256 mg, 1.98 mmol) and benzoylglycine (460 mg, 2.57 mmol) under an argon atmosphere were dissolved in DMF (2.5 mL), cooled to 0 °C and DIPEA (1.03 mL, 5.94 mmol) added. After 15 minutes, HATU (977 mg, 2.57 mmol) was added and the reaction stirred at room temperature for 16 hours. The solvent was removed *in vacuo* and the residue redissolved in EtOAc (20 ml). The organic phase

was washed with 1M HCl (2 x 20 mL), sat. NaHCO₃ (aq) (2 x 20 mL) and sat. NaCl (aq) (3 x 20 mL). The organic layer was dried with MgSO₄ (s), filtered and concentrated to afford methyl benzoylglycyl-L-prolinate (169 mg, 29%) as an orange solid. $R_f = 0.31$ (EtOAc);

¹H NMR (500 MHz, MeOD): δ 7.88 (2H, d, 7.3 Hz, COCCH), 7.56 (1H, t, 7.4 Hz, COCCHCHCH), 7.48 (2H, d, 7.6 Hz, COCCHCHCH), 4.79 (1H, dd, 1.8 Hz, 8.5 Hz, OCOCHN, minor), 4.50 (1H, dd, 4.1 Hz, 8.7 Hz, OCOCHN, major), 4.34 (2H, d, 16.9 Hz, COCH₂NH, major), 4.23 (2H, d, 16.4 Hz, COCH₂NH, minor), 4.16 (2H, d, 16.9 Hz, COCH₂NH, major), 3.94 (2H, d, 16.5 Hz, COCH₂NH, minor), 3.78 (3H, s, COOCH₃, minor), 3.72 (3H, s, COOCH₃, major), 3.76 – 3.66 (2H, m, CONCH₂, major), 3.64 – 3.56 (2H, m, CONCH₂, minor), 2.42 – 2.33 (1H, m, CONCHCH₂, minor), 2.32 – 2.24 (1H, m, CONCHCH₂, major), 2.13 – 2.05 (2H, m, CONCH₂CH₂, major), 2.05 – 1.96 (1H, m, CONCHCH₂, major and CONCH₂CH₂, minor), 1.94 – 1.83 (1H, m, CONCH₂CH₂, minor); **¹³C NMR** (125 MHz, MeOD): δ 174.3 (COOCH₃), 173.8 (COOCH₃), 170.3 (NCOCH₂), 170.1 (NCOCH₂), 169.7 (NHCOC), 135.2 (NHCOC), 132.9 (COCCHCHCH), 129.6 (COCCHCH), 128.4 (COCCH), 60.6 (OCOCHN), 60.3 (OCOCHN), 53.2 (COOCH₃), 52.8 (COOCH₃), 47.9 (CONCH₂), 47.5 (CONCH₂), 43.1 (COCH₂NH), 43.0 (COCH₂NH), 32.2 (COCHCH₂), 30.0 (COCHCH₂), 25.8 (COCHCH₂CH₂), 23.2 (COCHCH₂CH₂); **HRMS (ESI)**: calculated for C₁₅H₁₈N₂O₄Na [M+Na]⁺:313.1159, found: 313.1152.

7.1.3.89 Synthesis of benzoylglycyl-L-prolinate (213)



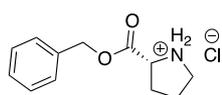
Methyl benzoylglycyl-L-prolinate (158 mg, 0.54 mmol) was dissolved in 3:1 MeOH:H₂O (4 mL) and LiOH (26 mg, 1.09 mmol) added. The reaction was stirred at room temperature for 4 hours, concentrated, and the aqueous phase acidified to pH 1 with 6M HCl. The aqueous phase was extracted with CHCl₃ (5 x 5 mL), dried with MgSO₄ (s), filtered and concentrated to afford benzoylglycyl-L-prolinate (138 mg, 93%) as a white solid. Before use in biological experiments, **213** was further purified by HPLC. Solvent A was H₂O with 0.1% formic acid, and solvent B was MeCN with 0.1%

formic acid. The gradient was from 5% B to 100% B over 45 minutes, then to 5% B over 5 minutes. Compound **213** had a retention time of 17.3 minutes.

IR (thin film) ν_{\max} : 3358 (O-H stretch), 1739 (C=O stretch), 1622 (C=O stretch);

¹H NMR (500 MHz, MeOD): δ 7.87 (2H, d, 7.4 Hz, COCCH), 7.54 (1H, t, 7.4 Hz, COCCHCHCH), 7.46 (2H, t, 7.6 Hz, COCCHCH), 4.72 (1H, dd, 2.2 Hz, 8.5 Hz, HOOCCH, major), 4.48 (1H, dd, 3.2 Hz, 8.8 Hz, HOOCCH), 4.34 (1H, d, 16.9 Hz, COCH₂NH, major), 4.27 (1H, d, 16.5 Hz, COCH₂NH), 4.14 (1H, d, 16.9 Hz, COCH₂NH, major), 3.94 (1H, d, 16.5 Hz, COCH₂NH), 3.74 - 3.64 (2H, m, NCH₂, major), 3.62 - 3.54 (2H, m, NCH₂), 2.42 - 2.32 (1H, m, NCHCH₂), 2.32 - 2.22 (1H, m, NCHCH₂, major), 2.12 - 2.01 (3H, m, NCHCH₂, NCH₂CH₂, major), 1.99 - 1.84 (3H, m, NCHCH₂, NCH₂CH₂); **¹³C NMR** (125 MHz, MeOD): δ 175.6 (COOH, major), 175.0 (COOH), 170.3 (NCOCH₂, major), 170.1 (NCOCH₂), 169.6 (NHCOC), 135.2 (NHCOC), 135.1 (NHCOC), 132.9 (CCHCHCH), 129.6 (CCHCH), 128.4 (CCH), 60.5 (HOOCCH, major), 60.3 (HOOCCH), 48.0 (NCH₂), 47.5 (NCH₂, major), 43.1 (COCH₂NH, major), 43.0 (COCH₂NH), 32.3 (NCHCH₂), 30.2 (NCHCH₂, major), 25.7 (NCH₂CH₂, major), 23.3 (NCH₂CH₂); **HRMS (ESI)**: calculated for C₁₄H₁₆N₂O₄Na [M+Na]⁺: 299.1002, found; 299.1003; [α]_D³⁰: -71.3 (0.067, MeOH).

7.1.3.90 Synthesis of (*R*)-2-((benzyloxy)carbonyl)pyrrolidin-1-ium chloride (**223**)



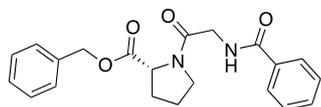
SOCl₂ (0.26 mL, 3.62 mmol) was added dropwise to benzyl alcohol (1.85 mL, 17.4 mmol) at 0 °C. After 15 minutes, D-proline (208 mg, 1.81 mmol) was added and stirred for 2 hours at 0 °C, then 16 hours at room temperature. Et₂O (35 mL) was added and the mixture cooled to -20 °C. The precipitate was collected to afford of (*R*)-2-((benzyloxy)carbonyl)pyrrolidin-1-ium chloride (359 mg, 82%).

¹H NMR (300 MHz, MeOD): δ 7.47 - 7.34 (5H, m, ArH), 5.34 - 5.30 (2H, m, ArCH₂O), 4.51 (1H, t, 7.6 Hz, OOCCHNH), 3.46 - 3.30 (2H, m, NCH₂), 2.52 - 2.36

(1H, m, NHCHCH₂), 2.21 – 1.98 (3H, m, NHCHCH₂, NHCH₂); **LRMS (ESI)**: calculated for C₁₂H₁₆O₂N [M+H]⁺: 206.1, found: 205.3.

The data are in accordance with those reported in the literature.³³⁷

7.1.3.91 Synthesis of benzyl benzoylglycyl-D-prolinate (224)



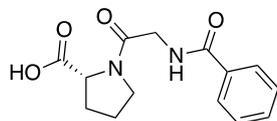
(*R*)-2-((Benzyloxy)carbonyl)pyrrolidin-1-ium chloride (148 mg, 0.61 mmol) and benzoylglycine (121 mg, 0.67 mmol) under an argon atmosphere were dissolved in anhydrous DMF (1.5 mL), cooled to 0 °C, and DIPEA (0.32 mL, 1.83 mmol) was added. After 15 minutes, HATU (255 mg, 0.67 mmol) was added and the reaction stirred at room temperature for 16 hours. The solvent was removed *in vacuo* and the residue redissolved in EtOAc (30 mL). The organic phase was washed with 5% LiCl (aq) (5 x 10 mL), 1M HCl (2 x 15 mL), sat. NaHCO₃ (aq) (2 x 15 mL) and sat. NaCl (aq) (2 x 20 mL). The organic phase was dried with MgSO₄ (aq), filtered and concentrated. The crude product was purified by column chromatography with an isocratic elution of Et₂O to afford benzyl benzoylglycyl-D-prolinate (141 mg, 63%) as a white solid. **R_f** = 0.50 (EtOAc); **IR** (thin film) ν_{max} : 1738 (C=O stretch), 1636 (C=O stretch), 1165 (C-O stretch), 693 (C-H bend);

>9:1 mixture of isomers, major isomer assigned

¹H NMR (500 MHz, CDCl₃): δ 7.83 (2H, t, 6.2 Hz, COCCH), 7.50 (1H, t, 7.4 Hz, CCHCHCH), 7.43 (2H, t, 7.5 Hz, COCCHCH), 7.39 – 7.31 (5H, m, ArH), 7.21 (1H, s, CONH), 5.22 (1H, d, 12.3 Hz, COCH₂Ar), 5.16 (1H, d, 12.3 Hz, COCH₂Ar), 4.62 (1H, dd, 3.4 Hz, 8.9 Hz, NCH), 4.31 (1H, dd, 4.4 Hz, 17.8 Hz, COCH₂NH), 4.21 (1H, dd, 3.6 Hz, 17.8 Hz, COCH₂NH), 3.72 – 3.51 (2H, m, NCH₂CH₂), 2.31 – 2.18 (1H, m, NCHCH₂), 2.14 – 2.01 (3H, m, NCHCH₂, NCH₂CH₂); **¹³C NMR** (125 MHz, CDCl₃): δ 171.7 (COOCH₂), 167.3 (NHCOC), 167.3 (NCOCH₂), 135.6 (CCH), 134.0 (CCH), 131.8 (CHCHCH), 128.8 (CH), 128.7 (CH), 128.5 (CH), 128.3 (CH), 127.2 (CH), 67.2 (OCH₂Ar), 59.3 (NCH), 46.2 (NCH₂), 42.7 (COCH₂NH), 29.2

(NCHCH₂), 24.8 (NCH₂CH₂); **HRMS (ESI)**: calculated for C₂₁H₂₂N₂O₄Na [M+Na]⁺: 389.1472, found: 389.1742.

7.1.3.92 Synthesis of benzoylglycyl-D-prolinate (**214**)



Benzyl benzoylglycyl-D-prolinate (42 mg, 0.12 mmol), Pd/C (10%) (10 mg, 0.09 mmol) and ammonium formate (36 mg, 0.57 mmol) under an argon atmosphere were suspended in anhydrous MeOH (5 mL) and heated to reflux for 16 hours. The reaction mixture was filtered through Celite and concentrated to afford benzoylglycyl-D-prolinate (31 mg, 97%) as a white solid. Before use in biological experiments, **214** was further purified by HPLC. Solvent A was H₂O with 0.1% formic acid, and solvent B was MeCN with 0.1% formic acid. The gradient was from 5% B to 100% B over 45 minutes, then to 5% B over 5 minutes. Compound **214** had a retention time of 17.0 minutes.

IR (thin film) ν_{\max} : 3290 (O-H stretch), 1726 (C=O stretch), 1633 (C=O stretch), 1179 (C-O stretch), 678 (C-H bend);

4:1 mixture of isomers

Major isomer ¹H NMR (500 MHz, MeOD): δ 7.89 (2H, d, 7.3 Hz, COCCH), 7.56 (1H, t, 7.4 Hz, CHCHCH), 7.49 (2H, t, 7.6 Hz, CCHCH), 4.51 – 4.47 (1H, m, NCH), 4.36 (1H, d, 16.9 Hz, COCH₂NH), 4.17 (1H, d, 16.9 Hz, COCH₂NH), 3.76 – 3.66 (2H, m, NCH₂), 2.34 – 2.25 (1H, m, NCHCH₂), 2.14 – 2.03 (NCHCH₂, NCH₂CH₂);

¹³C NMR (125 MHz, MeOD): δ 174.2 (COOH), 168.9 (NCOC), 168.2 (NCOCH₂), 133.8 (COCCH), 131.5 (CH), 128.2 (CH), 127.0 (CH), 59.2 (NCH), 46.1 (NCH₂), 41.7 (COCH₂NH), 28.8 (NCHCH₂), 24.3 (NCH₂CH₂);

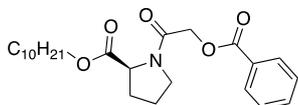
Minor isomer ¹H NMR (500 MHz, MeOD): δ 7.89 (2H, d, 7.3 Hz, COCCH), 7.56 (1H, t, 7.4 Hz, CHCHCH), 7.49 (2H, t, 7.6 Hz, CCHCH), 4.72 – 4.69 (1H, m, NCH), 4.29 (1H, d, 16.5 Hz, COCH₂NH), 3.96 (1H, d, 16.5 Hz, COCH₂NH), 3.64 – 3.57

(2H, NCH₂), 2.44 – 2.34 (1H, m, NCHCH₂), 2.03 – 1.86 (3H, m, NCHCH₂, NCH₂CH₂);

¹³C NMR (125 MHz, MeOD): δ 173.9 (COOH), 168.9 (NHCOC), 168.7 (NCOCH₂), 133.7 (CCH), 131.5 (CH), 128.2 (CH), 127.0 (CH), 59.1 (NCH), 46.6 (NCH₂), 41.7 (COCH₂NH), 31.0 (NCHCH₂), 21.9 (NCH₂CH₂);

HRMS (ESI): calculated for C₁₄H₁₅N₂O₄ [M-H]⁻: 275.1037, found: 275.1039; [α]_D³⁰: +64.6 (0.063, MeOH).

7.1.3.93 Synthesis of decyl (2-(benzoyloxy)acetyl)-L-proline (215)



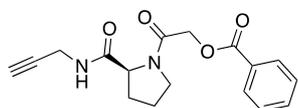
(2-(Benzoyloxy)acetyl)-L-proline (38 mg, 0.137 mmol) and 1-decanol (29 μL, 0.151 mmol) under an argon atmosphere was dissolved in anhydrous DCM (3 mL), and DCC (31 mg, 0.151 mmol) and DMAP (1.7 mg, 0.014 mmol) were added. The reaction mixture was stirred at room temperature overnight. The solvent was removed *in vacuo*, the residue dissolved in EtOAc (10 mL) and the solution filtered through Celite to remove dicyclohexylurea. The organic phase was washed with sat. NaHCO₃ (aq) (2 x 10 mL) and sat. NaCl (aq) (2 x 10 mL), dried with MgSO₄ (s), filtered and concentrated to afford of decyl (2-(benzoyloxy)acetyl)-L-proline (26 mg, 46%) as a colourless oil. Before use in biological experiments, **215** was further purified by HPLC. Solvent A was H₂O with 0.1% formic acid, and solvent B was MeCN with 0.1% formic acid. The gradient was from 50% B to 100% B over 45 minutes, then to 100% B over 5 minutes, then to 50% B over 5 minutes. Compound **215** had a retention time of 40.0 minutes. IR (thin film) ν_{max}: 2924 (C-H stretch), 1736 (C=O stretch), 1652 (C=O stretch), 681 (C-H bend);

¹H NMR (500 MHz, MeOD): δ 8.10 (2H, t, 7.9 Hz, COCCH), 7.65 (1H, t, 7.4 Hz, CHCHCH), 7.51 (2H, t, 7.8 Hz, CCHCH), 5.09 (1H, d, 15.0 Hz, COCH₂NH), 4.99 (1H, d, 15.0 Hz, COCH₂NH), 4.50 (1H, dd, 4.1 Hz, 8.7 Hz, NCH), 4.19 – 4.08 (2H, m, COOCH₂), 3.76 – 3.66 (2H, m, NCH₂), 2.40 – 2.25 (2H, m, NCHCH₂), 2.05 – 1.96

(2H, m, NCH₂CH₂), 1.69 – 1.61 (2H, m, OCH₂CH₂), 1.42 – 1.24 (14H, m, CH₂), 0.92 – 0.89 (3H, m, CH₃);

¹³C NMR (125 MHz, MeOD): δ 173.7 (COOCH₂), 168.1 (CONH), 167.4 (OCOC), 134.5 (CH), 130.8 (CH), 130.8 (CCH), 129.6 (CH), 66.5 (OCH₂CH₂), 63.1 (NHCH₂CO), 60.8 (NCH), 47.1 (NCH₂), 33.1 (CH₂), 32.2 (CH₂), 30.7 (CH₂), 30.7 (CH₂), 30.4 (CH₂), 30.4 (CH₂), 29.9 (NCHCH₂), 29.7 (CH₂), 27.0 (NCH₂CH₂), 25.8 (CH₂), 23.7 (CH₂), 14.4 (CH₂); HRMS (ESI): calculated for C₂₄H₃₅NO₅Na [M+Na]⁺: 440.2407, found: 440.2406; [α]_D³⁰: -88.6 (0.022, MeOH).

7.1.3.94 Synthesis of (S)-2-oxo-2-(2-(prop-2-yn-1-ylcarbamoyl)pyrrolidin-1-yl)ethyl benzoate (216)



(2-(Benzoyloxy)acetyl-L-proline (190 mg, 0.68 mmol) and propargylamine (48 μL, 0.75 mmol) were dissolved in anhydrous THF (2.5 mL) and EDC.HCl (182 mg, 0.95 mmol), HOBT (130 mg, 0.95 mmol) and Et₃N (0.28 mL, 2.0 mmol) were added. The mixture was stirred at room temperature for 48 hours. The solvent was removed in vacuo and the residue redissolved in EtOAc. The organic phase was then washed with 1M HCl (5 mL), H₂O (5 mL), sat. NaHCO₃ (aq) (5 mL) and sat. NaCl (aq) (5 mL), dried with MgSO₄ (s), filtered and concentrated *in vacuo* to afford (S)-2-oxo-2-(2-(prop-2-yn-1-ylcarbamoyl)pyrrolidin-1-yl)ethyl benzoate as a white solid (139 mg, 65 %). Before use in biological experiments, **216** was further purified by HPLC. Solvent A was H₂O with 0.1% formic acid, and solvent B was MeCN with 0.1% formic acid. The gradient was from 5% B to 100% B over 45 minutes, then to 5% B over 5 minutes. Compound **216** had a retention time of 23.8 minutes.

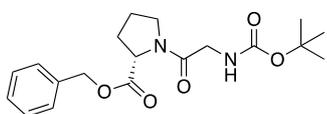
IR (thin film) ν_{max}: 3285 (N-H stretch), 1727 (C=O stretch), 1659 (C=O stretch), 710 (C-H bend);

¹H NMR (500 MHz, MeOD): δ 7.98 (2H, t, 8.6 Hz, COCCH), 7.53 (1H, t, 7.4 Hz, CHCHCH), 7.39 (2H, t, 7.8 Hz, CCHCH), 5.00 (1H, d, 15.0 Hz, COCH₂O), 4.86 (1H, d, 15.0 Hz, COCH₂O), 4.34 (1H, dd, 3.8 Hz, 8.5 Hz, NCH), 3.91 – 3.81 (2H, m,

NHCH₂C), 3.65 – 3.54 (2H, m, NCH₂), 3.21 (1H, s, CH₂CCH), 2.02 – 1.82 (1H, m, NCHCH₂), 1.32 – 1.09 (3H, m, NCHCH₂, NCH₂CH₂);

¹³C NMR (125 MHz, MeOD): δ 174.0 (COOCH₂), 168.4 (CON), 167.6 (CONHCH₂), 134.5 (CH), 130.8 (CH), 130.8 (CCH), 129.6 (CH), 72.3 (CH₂CCH), 63.3 (COCH₂O) 61.7 (CH₂CCH), 49.9 (NCH), 47.4 (NCH₂), 30.5 (CH₂CCH), 29.6 (NCHCH₂), 25.7 (NCH₂CH₂); [α]_D³⁰: -56.5 (0.057, MeOH); HRMS (ESI): calculated for C₁₇H₁₈N₂O₄Na [M+Na]⁺: 337.1159, found: 337.1162.

7.1.3.95 Synthesis of benzyl (*tert*-butoxycarbonyl)glycyl-L-prolinate (226)



(*S*)-2-((Benzyloxy)carbonyl)pyrrolidin-1-ium chloride (311 mg, 1.29 mmol) and Boc-Gly-OH (249 mg, 1.42 mmol) under an argon atmosphere were dissolved in DMF (2 mL), cooled to 0 °C and DIPEA (674 μL, 3.87 mmol) added. After 15 minutes, HATU (540 mg, 1.42 mmol) was added, and the solution stirred at room temperature for 2 hours. The solvent was removed *in vacuo* and the residue redissolved in CHCl₃ (25 mL). The organic phase was washed with 5% LiCl (aq) (5 x 10 mL), 1M HCl (2 x 10 mL), sat. NaHCO₃ (aq) (2 x 10 mL) and sat. NaCl (aq) (2 x 10 mL), then dried with MgSO₄ (s), filtered and concentrated. The crude product was purified by column chromatography with a stepwise gradient from 1:1 petroleum ether: EtOAc to 2:3 petroleum ether: EtOAc to afford benzyl (*tert*-butoxycarbonyl)glycyl-L-prolinate (370 mg, 79%) as a colourless oil. R_f: 0.24 (1:1 petroleum ether: EtOAc).

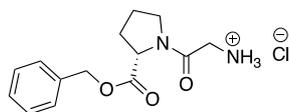
>9:1 mixture of isomers

¹H NMR (500 MHz, CDCl₃): δ 7.38 – 7.32 (5H, m, ArH), 5.42 (1H, s, NHCOO), 5.16 (2H, q, 12.3 Hz, ArCH₂O), 4.59 – 4.56 (1H, m, OOCCHN), 3.96 (2H, ddd, 4.3 Hz, 17.2 Hz, 21.0 Hz, NCOCH₂NH), 3.61 – 3.55 (1H, m, OCNCH₂CH₂), 3.49 – 3.43 (1H, m, OCNCH₂CH₂), 2.22 – 2.15 (1H, m, OOCCHCH₂), 2.09 – 1.97 (3H, m, OOCCHCH₂, OOCCHCH₂CH₂), 1.44 (9H, s, C(CH₃)₃); ¹³C NMR (125 MHz, CDCl₃): δ 171.8 (CON), 167.5 (COO), 156.0 (NCOO), 135.7 (CHCCH), 128.7 (CH),

128.5 (CH), 128.2 (CH), 79.8 (C(CH₃)₃), 67.7 (ArCH₂O), 67.1 (ArCH₂O), 59.1 (OOCCHN), 58.8 (OOCCHN), 46.8 (OCNCH₂), 46.0 (OCNCH₂), 43.2 (NCOCH₂NH), 43.0 (NCOCH₂NH), 31.5 (OOCCHCH₂), 29.1 (OOCCHCH₂), 28.5 (C(CH₃)₃), 24.8 (OCNCH₂CH₂), 22.3 (OCNCH₂CH₂); **LRMS (ESI)**: calculated for C₁₉H₂₆N₂O₅Na [M+Na]⁺: 385.2, found: 385.3.

The data are in accordance with those reported in the literature.³³⁸

7.1.3.96 Synthesis of (S)-2-(2-((benzyloxy)carbonyl)pyrrolidin-1-yl)-2-oxoethan-1-aminium chloride (227)



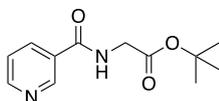
Benzyl (*tert*-butoxycarbonyl)glycyl-L-prolinate (360 mg, 0.99 mmol) was dissolved in 1:1 Et₂O: MeOH (6 mL), and 2M HCl in Et₂O (3.97 mL, 7.95 mmol) was added and stirred at room temperature for 4.5 hours. The solvent was removed *in vacuo* to afford (S)-2-(2-((benzyloxy)carbonyl)pyrrolidin-1-yl)-2-oxoethan-1-aminium chloride (286 mg, 97%).

¹H NMR (300 MHz, D₂O): δ 7.35 (5H, s, ArH), 5.20 – 5.13 (2H, m, CH₂Ar), 4.61 – 4.54 (1H, m, NCH), 4.06 – 3.86 (2H, m, NCH₂), 3.64 – 3.40 (2H, m, COCH₂NH₂), 2.26 – 2.11 (1H, m, NCHCH₂), 2.08 – 1.94 (3H, m, NCHCH₂, NCH₂CH₂);

LRMS (ESI): calculated for C₁₄H₁₉N₂O₃ [M+H]⁺: 263.1, found: 263.2.

The data are in accordance with those reported in the literature.³³⁹

7.1.3.97 Synthesis of *tert*-butyl nicotinoylglycinate (230)



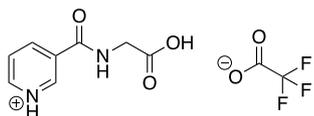
Gly-O^tBu.AcOH (200 mg, 1.05 mmol) and nicotinoyl chloride hydrochloride (203 mg, 1.15 mmol) were dissolved in DCM (10 mL) and cooled to 0 °C. Et₃N (0.51 mL, 3.66 mmol) was added, and the reaction mixture stirred at room temperature for 16 hours. The solvent was removed *in vacuo* and redissolved in EtOAc (20 mL). The organic phase was washed with 1M HCl (2 x 20 mL), sat. NaHCO₃ (aq) (2 x 20 mL),

and sat. NaCl (aq) (20 mL), then dried with MgSO₄ (aq), filtered and concentrated. The crude product was purified by column chromatography with an isocratic elution using EtOAc to afford *tert*-butyl nicotinoylglycinate (199 mg, 80%). **R_f** = 0.33 (EtOAc);

¹H NMR (500 MHz, CDCl₃): δ 9.03 (1H, d, 1.6 Hz, NCHC), 8.74 (1H, dd, 1.2 Hz, 4.7 Hz, NCH), 8.13 (1H, dd, 1.8 Hz, 6.1 Hz, OCCCH), 7.39 (1H, dd, 4.9 Hz, 7.8 Hz, NCHCH), 4.15 (2H, d, 4.9 Hz, NHCH₂CO), 1.50 (9H, s, C(CH₃)₃);

¹³C NMR (125 MHz, CDCl₃): δ 169.1 (COO), 165.5 (NHCO), 152.7 (NCHC), 148.3 (NCH), 135.2 (OCCCH), 129.7 (NCHC), 123.6 (NCHCH), 83.0 (C(CH₃)₃), 42.6 (NHCH₂CO), 28.2 (C(CH₃)₃); **HRMS (ESI)**: calculated for C₁₂H₁₇N₂O₃ [M+H]⁺: 237.1234, found: 237.1228.

7.1.3.98 Synthesis of 3-((carboxymethyl)carbamoyl)pyridin-1-ium trifluoroacetate (231)

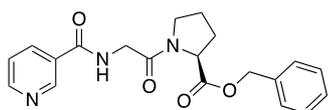


tert-Butyl nicotinoylglycinate (132 mg, 0.56 mmol) was dissolved in DCM (3 mL), and TFA (0.86 mL, 11.2 mmol) and H₂O (0.5 mL) was added. The reaction mixture was stirred overnight at room temperature. The solvent was removed *in vacuo* to afford 3-((carboxymethyl)carbamoyl)pyridin-1-ium trifluoroacetate (162 mg, 98%).

¹H NMR (500 MHz, MeOD): δ 9.17 (1H, d, 1.1 Hz, NCHC), 8.88 (1H, d, 4.8 Hz, NCH), 8.66 (1H, d, 8.1 Hz, OCCCH), 7.90 (1H, dd, 5.4 Hz, 7.9 Hz, NCHCH), 4.17 (2H, s, NHCH₂COOH);

¹³C NMR (125 MHz, MeOD): δ 172.7 (COOH), 166.3 (NHCO), 149.2 (NCHC), 146.3 (NCH), 141.3 (OCCCH), 133.2 (OCCCH), 126.8 (NCHCH), 42.3 (NHCH₂CO); **HRMS (ESI)**: calculated for C₈H₇N₂O₃ [M-H]⁻: 179.0462, found: 179.0462.

7.1.3.99 Synthesis of benzyl nicotinoylglycyl-L-prolinate (228)



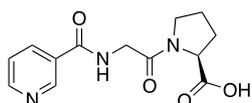
3-((Carboxymethyl)carbonyl)pyridin-1-ium trifluoroacetate (101 mg, 0.34 mmol) and L-Pro-OBn (75 mg, 0.31 mmol) under an argon atmosphere were suspended in anhydrous DMF (1 mL) and cooled to 0 °C, then DIPEA (216 μ L, 1.24 mmol) added. After 15 minutes, HATU (130 mg, 0.34 mmol) was added and the reaction stirred at room temperature for 3 hours. The solvent was removed *in vacuo* and the residue redissolved in EtOAc (20 mL). The organic phase was washed with 5% LiCl_(aq) (5 x 10 mL), sat. NaHCO_{3(aq)} (2 x 10 mL) and sat. NaCl_(aq) (2 x 10 mL) then dried with MgSO_{4(s)}, filtered and concentrated. The crude residue was purified by column chromatography with an isocratic elution using 1:1 EtOAc: acetone to afford benzyl nicotinoylglycyl-L-prolinate (64 mg, 56%). **R_f**: 0.34 (acetone: EtOAc);

¹H NMR (500 MHz, MeOD): δ 9.04 (1H, t, 3.3 Hz, NCHC), 8.71 (1H, dd, 1.3 Hz, 4.9 Hz, NCH), 8.31 – 8.28 (1H, m, OCCCH), 7.56 (1H, dd, 5 Hz, 7.9 Hz, NCHCH), 5.18 (2H, m, OCH₂Ar), 4.57 (1H, dd, 4.0 Hz, 8.7 Hz, NCH), 4.36 (1H, d, 1.9 Hz, COCH₂NH), 4.21 (1H, d, 16.9 Hz, COCH₂NH), 3.77 – 3.68 (2H, m, NCH₂), 2.34 – 2.24 (1H, m, NCHCH₂), 2.12 – 1.99 (3H, m, NCHCH₂, NCH₂CH₂);

¹³C NMR (125 MHz, MeOD): δ 173.5 (COOCH₂), 169.5 (COCH₂), 168.0 (CCONH), 152.8 (NCHC), 149.3 (NCH), 137.3 (CCH), 137.2 (OCCCH), 131.6 (OCCCH), 129.6 (CH), 129.3 (CH), 129.1 (CH), 125.1 (NCHCH), 67.9 (OCH₂Ar), 60.7 (NCH), 47.5 (NCH₂), 43.0 (NHCH₂CO), 30.0 (NCHCH₂), 25.8 (NCH₂CH₂);

HRMS (ESI): calculated for C₂₀H₂₁N₃O₂Na [M+Na]⁺: 390.1424, found: 390.1429.

7.1.3.100 Synthesis of nicotiniylglycyl-L-proline (217)



Benzyl nicotinoylglycyl-L-prolinate (64 mg, 0.174 mmol) and Pd/C (10%) (15 mg, 0.139 mmol) under an argon atmosphere were suspended in anhydrous MeOH (8 mL) and the mixture degassed. Hydrogen gas was bubbled through the solution for 5 hours, then the reaction mixture was filtered through Celite and concentrated to afford nicotiniylglycyl-L-proline (49 mg, 95%) as a white solid. Before use in biological experiments, **217** was further purified by HPLC. Solvent A was H₂O with 0.1% formic acid, and solvent B was MeCN with 0.1% formic acid. The gradient was from 5% B to 95% B over 45 minutes, then to 5% B over 5 minutes. Compound **217** had a retention time of 18.7 minutes.

IR (thin film) ν_{\max} : 3077 (O-H stretch), 1630 (C=O stretch), 1176 (C-N stretch), 719 (C-H bend);

4:1 mixture of isomers

Major isomer **¹H NMR** (500 MHz, MeOD): δ 9.11 – 9.09 (1H, m, NCHC), 8.81 (1H, d, 4.7 Hz, NCH), 8.52 (1H, t, 7.8 Hz, OCCCH), 7.76 (1H, dt, 4.6 Hz, 8.6 Hz, NCHCH), 4.51 (1H, dd, 3.4 Hz, 8.8 Hz, NCH), 4.40 (1H, d, 17.0 Hz, COCH₂NH), 4.22 (1H, d, 17.0 Hz, COCH₂NH), 3.77 – 3.67 (2H, m, NCH₂), 2.35 – 2.26 (1H, m, NCHCH₂), 2.15 – 2.04 (3H, m, NCHCH₂, NCH₂CH₂);

¹³C NMR (125 MHz, MeOD): δ 174.1 (COOH), 167.9 (COCH₂), 165.5 (CCONH), 149.1 (NCHC), 146.0 (NCH), 138.4 (OCCCH), 131.3 (OCCCH), 124.8 (NCHCH), 59.2 (NCH), 46.1 (NCH₂), 41.7 (COCH₂NH), 28.8 (NCHCH₂), 24.3 (NCH₂CH₂);

Minor isomer **¹H NMR** (500 MHz, MeOD): δ 9.11 – 9.09 (1H, m, NCHC), 8.81 (1H, d, 4.7 Hz, NCH), 8.52 (1H, t, 7.8 Hz, OCCCH), 7.76 (1H, dt, 4.6 Hz, 8.6 Hz, NCHCH), 4.73 (1H, dd, 3.4 Hz, 8.8 Hz, NCH), 4.34 (1H, d, 16.6 Hz, COCH₂NH), 4.01 (1H, d, 16.6 Hz, COCH₂NH), 3.66 – 3.57 (2H, m, NCH₂), 2.45 – 2.35 (1H, m, NCHCH₂), 2.03 – 1.90 (3H, m, NCHCH₂, NCH₂CH₂);

¹³C NMR (125 MHz, MeOD): δ 173.5 (COOH), 168.3 (COCH₂), 165.5 (CCONH), 149.2 (NCHC), 146.1 (NCH), 138.3 (OCCCH), 131.2 (OCCCH), 124.7 (NCHCH), 59.2 (NCH), 46.1 (NCH₂), 41.7 (COCH₂NH), 28.8 (NCHCH₂), 24.3 (NCH₂CH₂);

HRMS (ESI): calculated for C₁₃H₁₅N₃O₄Na [M+Na]⁺: 300.0955, found: 300.0952; [α]_D³⁰: -56.4 (0.151, MeOH).

7.2 Feeding experiments of bacterial strains with chemical probes

7.2.1 General microbiology methods

For all feeding experiments on plate, the probe was dissolved in MeOH and introduced to the agar before forming the plate. All feeding experiments were conducted in duplicate. All glassware and media were sterilised by autoclaving (Astell). Liquid cultures were grown with shaking in an Innova 44 incubator/ shaker (New Brunswick Scientific), solid cultures were grown in an Heraeus incubator (Thermo).

7.2.2 Bacterial strains used with chemical probes

Strain	Description
<i>Escherichia coli</i> MG1655 pBACpks ⁺	Strain containing the colibactin gene cluster (Gift from the Piel group, ETH Zürich) (Kan ^R)
<i>Escherichia coli</i> MG1655 pBACpks ⁺ Δ <i>clbA</i>	Strain containing the colibactin gene cluster with the deletion of <i>clbA</i> (Gift from the Piel group) (Kan ^R)
<i>Escherichia coli</i> MG1655 pBACpks ⁺ :: <i>clbL</i>	Strain containing the colibactin gene cluster with an insertional inactivation of <i>clbL</i> (Gift from the Piel group) (Kan ^R Cm ^R)

<i>Escherichia coli</i> MG1655 pBACpks ⁺ :: <i>clbO(KS)</i>	Strain containing the colibactin gene cluster with an insertional inactivation of <i>clbO(KS)</i> (Gift from the Piel group) (Kan ^R Cm ^R)
<i>Escherichia coli</i> MG1655 pBACpks ⁺ :: <i>clbO(ACP)</i>	Strain containing the colibactin gene cluster with an insertional inactivation of <i>clbO(ACP)</i> (Gift from the Piel group) (Kan ^R Cm ^R)
<i>Escherichia coli</i> MG1655 pBACpks ⁺ :: <i>clbQ</i>	Strain containing the colibactin gene cluster with an insertional inactivation of <i>clbQ</i> (Gift from the Piel group) (Kan ^R Cm ^R)
<i>Escherichia coli</i> MG1655 pBACpks ⁺ :: <i>clbP</i>	Strain containing the colibactin gene cluster with an insertional inactivation of <i>clbP</i> (Gift from the Piel group) (Kan ^R Cm ^R)
<i>Amycolatopsis orientalis</i> DSM40040	Wild-type strain producing vancomycin
<i>Streptomyces lasaliensis</i> NRRL3382	Wild-type strain producing echinomycin and lasalocid A
<i>Streptomyces lasaliensis</i> ACP12 S970A	Strain producing echinomycin, with a point mutation in ACP12 to prevent lasalocid production (Tosin group) ¹⁴⁹
<i>Streptomyces albus/scl</i>	Strain containing the scleric acid gene cluster from <i>S. sclerotialus</i> (Corre group) (Kan ^R)
<i>Streptomyces albus/scl</i> Δ <i>M4</i>	Strain containing the scleric acid gene cluster with the deletion of <i>scIM4</i> (Corre group) (Kan ^R)

<i>Streptomyces albus/scl ΔM4 ΔA</i>	Strain containing the scleric acid gene cluster with the deletions of <i>sclM4</i> and <i>sclA</i> (Corre group) (Kan ^R)
<i>Streptomyces albus/scl ΔM4 ΔN</i>	Strain containing the scleric acid gene cluster with the deletions of <i>sclM4</i> and <i>sclN</i> (Corre group) (Kan ^R)
<i>Streptomyces albus/scl ΔM4 ΔQ1-4</i>	Strain containing the scleric acid gene cluster with the deletions of <i>sclM4</i> and <i>sclQ1-4</i> (Corre group) (Kan ^R)

Table 6 - Table of bacterial strains used in this work.

7.2.3 Bacterial culture media

Media	Components
M79 medium	10 g glucose, 10g peptone, 2 g yeast extract, 6 g NaCl, 10 g casamino acids. dH ₂ O up to 1 L. Adjusted to pH 7.0.
MYM medium	4 g maltose, 4 g yeast extract, 10 g malt extract. dH ₂ O up to 1 L. Adjusted to pH 7.2.
MYM agar medium	4 g maltose, 4 g yeast extract, 10 g malt extract, 15 g agar. dH ₂ O up to 1 L. Adjusted to pH 7.2.
LB medium	10 g tryptone, 5 g yeast extract, 10 g NaCl. dH ₂ O up to 1 L.
LB agar medium	10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g agar. dH ₂ O up to 1 L.
TSB medium	30 g tryptic soy broth. dH ₂ O up to 1 L.
TSB agar medium	30 g tryptic soy broth, 15 g agar. dH ₂ O up to 1 L.

SMMS agar medium	2 g casamino acids, 8.68 g TES buffer, 15 g agar. dH ₂ O up to 956 mL. After autoclaving, add 10 mL each of 50 mM NaH ₂ PO ₄ and K ₂ HPO ₄ , 5 mL of 1M MgSO ₄ , 18 mL of 50% w/v glucose, and 1 mL of trace element solution (0.1 gL ⁻¹ of ZnSO ₄ .7H ₂ O, FeSO ₄ .7H ₂ O, MnCl ₂ .4H ₂ O, CaCl ₂ .6H ₂ O and NaCl).
SFM agar medium	20 g soya flour, 20 g mannitol, 20 g agar. dH ₂ O up to 1 L.
2 x YT medium	16 g tryptone, 10 g yeast extract, 5 g NaCl. dH ₂ O up to 1 L. Adjusted to pH 7.0.
SOC medium	20 g tryptone, 5 g yeast extract, 0.584 g NaCl, 0.186 g KCl, 2.4 g MgSO ₄ , 3.6 g glucose. dH ₂ O up to 1 L.
YPD medium	5 g yeast extract, 10 g peptone, 20 g glucose. dH ₂ O up to 1 L.
YPD agar medium	5 g yeast extract, 10 g peptone, 20 g glucose, 15 g agar. dH ₂ O up to 1 L.
R5 medium	103 g sucrose, 0.25 g K ₂ SO ₄ , 10.12 g MgCl ₂ .6H ₂ O, 10 g glucose, 0.1 g casamino acids, 5 g yeast extract, 5.73 g TES buffer. dH ₂ O up to 893 mL. After autoclaving, add 2 mL trace elements solution (0.2 gL ⁻¹ FeCl ₃ .6H ₂ O, 0.01 gL ⁻¹ of Na ₂ B ₄ O ₇ .10H ₂ O, (NH ₄) ₆ Mo ₇ O ₂₄ .5H ₂ O, CuCl ₂ .2H ₂ O, MnCl ₂ .4H ₂ O, 0.04 gL ⁻¹ ZnCl ₂), and 80 mL of 0.33 M CaCl ₂ , 10 mL of 3.96 M KH ₂ PO ₄ , and 15 mL of 173.7 M L-proline.
R5 agar medium	103 g sucrose, 0.25 g K ₂ SO ₄ , 10.12 g MgCl ₂ .6H ₂ O, 10 g glucose, 0.1 g casamino acids, 5 g yeast extract, 5.73 g TES buffer, 15 g agar. dH ₂ O up to 893 mL. After autoclaving, add 2 mL trace elements solution (0.2 gL ⁻¹ FeCl ₃ .6H ₂ O, 0.01 gL ⁻¹ of Na ₂ B ₄ O ₇ .10H ₂ O, (NH ₄) ₆ Mo ₇ O ₂₄ .5H ₂ O, CuCl ₂ .2H ₂ O, MnCl ₂ .4H ₂ O, 0.04 gL ⁻¹ ZnCl ₂), and 80 mL of

0.33 M CaCl₂, 10 mL of 3.96 M KH₂PO₄, and 15 mL of 173.7 M L-proline.

Table 7 - Table of bacterial growth media used in this work. Autoclaved after addition of all components unless specified otherwise.

7.2.4 Feeding experiments set up

Seed cultures of *S. lasaliensis* ACP12(S970A) (10 mL, with springs) were grown in M79 media for 5 days at 30 °C shaking at 250 rpm. MYM agar plates containing the probes **54 – 56**, **60 – 62**, or **75** at 0.2 mM, 1.0 mM and/ or 2.0 mM concentrations were inoculated with 50 µL seed culture and grown for 5 days at 30 °C. For liquid feedings, MYM media (5 mL, with springs) was inoculated with 50 µL seed culture and grown overnight at 30 °C with shaking at 250 rpm. The culture was then supplemented once daily with probe (0.05, 0.25 or 0.5 mM in 25 µL MeOH) for 4 days, and then extracted with 10 mL EtOAc, concentrated, and the residue redissolved in 1 mL HPLC-grade MeOH. The plates were extracted with EtOAc (10 mL), concentrated, and redissolved in 1 mL HPLC-grade MeOH for HPLC-MSⁿ analysis on an Orbitrap Fusion instrument.

Colibactin *E. coli* strains (*E. coli* MG1665 pBAC *pks*⁺, Δ *clbA*, *::clbO(KS)*, *::clbO(ACP)*, *::clbL*, *::clbP*, *::clbQ*) (10 mL cultures) were grown overnight in LB medium containing the appropriate antibiotics as indicated in Table 6 at 37 °C shaking at 180 rpm. 0.5 mL of overnight culture was used to inoculate 50 mL fresh LB media with antibiotic and grown to OD₆₀₀ = 0.4 – 0.6. The 50 mL culture was aliquoted into 5 mL cultures, which were was grown overnight with 1.0 mM or 2.0 mM concentrations of probes **75**, **81**, **136**, **138**, or **141**. Cultures were extracted with EtOAc (10 mL), concentrated, and redissolved in 1 mL HPLC-grade MeOH for HPLC-MSⁿ analysis on an Orbitrap Fusion instrument.

Seed cultures of *A. orientalis* (10 mL, with springs) were grown in R5 media for 5 days at 30 °C with shaking at 250 rpm. For feedings on solid media, R5 agar plates containing the probes (one of **81**, **174 – 180**) at 2.0 mM concentration were inoculated with 50 µL seed culture and grown for 5 days at 30 °C. For chemical supplementation of liquid cultures, R5 media (5 mL, with springs) was inoculated with 50 µL seed

culture and grown overnight at 30 °C with shaking at 250 rpm. The culture was then supplemented once daily with probe (one of **81**, **174** – **180**) (0.5 mM in 25 µl MeOH) for 4 days, and then lyophilized. The lyophilised solid was extracted with 10 mL MeOH, concentrated, and redissolved in 1 mL HPLC-grade MeOH. The plates were extracted with MeOH (10 mL), concentrated, and redissolved in 1 mL HPLC-grade MeOH for HPLC-MSⁿ analysis on a MaXis Impact and on an Orbitrap Fusion instrument.

Seed cultures of *scl S. albus* strains (10 mL, with springs) were grown in tryptic soy broth (TSB) medium with 50 µg mL⁻¹ kanamycin for 3 days at 30 °C shaking at 250 rpm. Supplemented solid minimal medium (SMMS) plates containing 5 mM scleric intermediate, **202**, or 1.0 mM and 2.0 mM chemical probe **67**, **81**, or **211**, and 50 µg mL⁻¹ kanamycin were inoculated with 50 µL seed culture and grown for 5 days at 30 °C. The plates were extracted with EtOAc acidified with to pH 3 with conc. HCl (10 mL), concentrated, and redissolved in 1 mL HPLC-grade MeOH for HPLC-MSⁿ analysis.

7.2.5 HR-LC-MSⁿ analysis of bacterial extracts

Detailed analysis of all the extracts deriving from chemical probe feeding experiments was performed on a Thermo Orbitrap Fusion (Q-OT-qIT) instrument. Reverse phase chromatography was used to separate the mixtures prior to MS analysis. Two columns were used: an Acclaim PepMap µ-precolumn cartridge 300 µm i.d. x 5 mm 5 µm 100 Å and an Acclaim PepMap RSLC 75 mm x 15 cm 2 µm 100 Å (Thermo Scientific). The columns were installed on an Ultimate 3000 RSLCnano system (Dionex). Mobile phase buffer A was composed of 0.1% aqueous formic acid and mobile phase buffer B was composed of 100% acetonitrile containing 0.1% formic acid. Samples (2 µl) were loaded onto the µ-precolumn equilibrated in 2% aqueous acetonitrile containing 0.1% trifluoroacetic acid for 8 mins at 10 µL min⁻¹ after which compounds were eluted onto the analytical column following a gradient described below. Eluting cations were converted to gas-phase ions by electrospray ionization and analysed. Scans of precursors from 150 to 1500 m/z were performed at 60K resolution (at 200 m/z) with a 4 x 10⁵ ion count target. Tandem MS were performed by isolation at 1.6 Th with the

quadrupole, HCD fragmentation with normalized collision energy of 32, and rapid scan MS analysis in the ion trap. The MS² ion count target was set to 2 x 10⁴ and the maximum injection time was 50 ms. A filter targeted inclusion mass list was used to select the precursor ions. The dynamic exclusion duration was set to 45 s with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top speed mode with 5 s cycles, meaning the instrument would continuously perform MS² events until the list of non-excluded precursors diminishes to zero or 5 s, whichever is shorter. Fusion runs were performed with Survey scans of precursors from 150 to 1500 m/z 60K resolution.

The gradients used to separate the mixtures varied according to the probes and system being analysed, as were as shown in Tables 8 – 10. 2 µl of extract was injected, and a flow rate of 0.3 µl/min was used.

Time (mins)	Solvent B (%)
0	3
5	5
25	90
33.9	90
34	3
50	3

Table 8 - Solvent gradient used to separate echinomyacin extracts containing chemical probes.

Time (mins)	Solvent B (%)
0	3
5	5
30	80
40	80
41	3

50	3
----	---

Table 9 - Solvent gradient used to separate colibactin extracts containing aminomalonate- and NRPS chemical probes, and vancomycin, and scleric acid extracts containing chemical probes.

Time (mins)	Solvent B (%)
0	3
5	3
45	80
59	80
60	3
75	3

Table 10 - Solvent gradient used to separate colibactin extracts containing the fluoromalonate chemical probe.

7.2.6 MaXis Impact LC-MS Analysis

Preliminary analysis of the supplementation of chemical probes **81** and **174 – 180** to *Amycolatopsis orientalis* and the outcome of the supplementation of scleric acid intermediate **202** to *S. albus scl ΔMΔN* was established *via* analysis of the organic extracts using a MaXis Impact UHR-TOF (Bruker Daltonics) instrument. Samples (5 μL) were injected onto an Agilent Eclipse C18 column (1.8 μm, 100 mm x 2.1 mm). Mobile phase buffer A was composed of 0.1% aqueous formic acid and mobile phase buffer B was composed of 100% acetonitrile containing 0.1% formic acid. The following solvent gradient was applied: 10% B 0 - 2.7 min; 10 - 100% B 2.7 - 42.7 min; 100% B 42.7 - 52.7 min; 100 - 10% B 52.7 - 55.7 min; 10% B 55.7 - 67.7 min, at a flow rate of 0.5 mL/min. Spectra were recorded in positive ionisation mode, scanning from m/z 0 to 2000 with the resolution set at 45K. Gathered data were internally calibrated once they have been acquired from the instrument using the Bruker Daltonics Data Analysis software. A calibration curve was fitted to the peaks and applied to the spectrum.

7.3 General molecular biology and protein methods

7.3.1 Reagents and equipment

Chemicals were purchased from Sigma-Aldrich unless otherwise specified. Either GeneJET Gel Extraction Kit and GeneJET Plasmid Miniprep Kit from Thermo Fisher Scientific, or Monarch DNA Gel Extraction Kit and Monarch Plasmid Miniprep Kit from New England Biolabs were used. Q5 High-Fidelity DNA polymerase from New England Biolabs, or Phusion High-Fidelity DNA polymerase, purchased from Thermo Fisher Scientific, were used. Restriction enzymes and T4 ligase was purchased from Thermo Fisher Scientific. All enzymes were used with the recommended buffers. All oligonucleotides were purchased from Sigma-Aldrich, and a gBlock was purchased from Integrated DNA Technologies. Champion pET151 Directional TOPO Expression Kit was purchased from Thermo Fisher Scientific. Electrophoresis-grade agarose was purchased from Invitrogen or Geneflow. Agarose gel electrophoresis was conducted using a Mini-Sub Cell GT Cell and PowerPac from BioRad. DNA was visualised using GelRed from Thermo Fisher Scientific and a UVP Ultraviolet transilluminator at 365 nm. DNA size was estimated using GeneRuler 1 kb DNA ladder from Thermo Fisher Scientific. Polyacrylamide gel electrophoresis was conducted using a Mini-PROTEAN Tetra Vertical Electrophoresis Cell from BioRad. Protein size was estimated using PageRuler Plus Prestained Protein Ladder, 10 to 250 kDa from Thermo Fisher Scientific. Bacterial strains and plasmids used are property of the Tosin or Corre groups at the University of Warwick, the Piel group at ETH Zurich. PCR was conducted using an Eppendorf Mastercycler nexus X2. DNA concentration was measured using a NanoDrop Lite spectrophotometer from Thermo Scientific.

7.3.2 Bacterial strains used for molecular biology and protein biochemistry

Strain	Description
<i>Escherichia coli</i> BL21(DE3)	<i>E. coli</i> for the production of protein
<i>Escherichia coli</i> TOP10	<i>E. coli</i> for DNA cloning

<i>Escherichia coli</i> BAP1	<i>E.coli</i> for protein production containing phosphopantetheinyl transferase Sfp
<i>Escherichia coli</i> Stbl4	<i>E. coli</i> for the cloning of unstable DNA
<i>Escherichia coli</i> ET12567/pUZ8002	<i>E.coli</i> used for the biparental conjugation of plasmids with <i>Streptomyces</i> strains (Kan ^R , Cm ^R)
<i>Escherichia coli</i> ET12567/pUZ8002/pCRISPomyces2-C1	<i>E. coli</i> strain used for transfer of pCRISPomyces2-C1 into <i>S. lasaliensis</i> by conjugation (Kan ^R , Cm ^R , Apra ^R)
<i>Escherichia coli</i> ET12567/pUZ8002/pCRISPomyces2-C2	<i>E. coli</i> strain used for transfer of pCRISPomyces2-C2 into <i>S. lasaliensis</i> by conjugation (Kan ^R , Cm ^R , Apra ^R)
<i>Escherichia coli</i> ET12567/pUZ8002/CrisprCodA	<i>E. coli</i> strain used for transfer of CRISPR/Cas9(CodA) into <i>S. lasaliensis</i> by conjugation (Kan ^R , Cm ^R , Apra ^R)

Table 11 - Table of bacterial strains used for molecular biology and protein expression in this work.

7.3.3 Oligonucleotides

Name	Sequence (5'-3')
FabC FF	CACCATGGCCGCCACTCAGGAAGA
FabC RR	TCAGGCCTGGTGCTTGAGGA
Ecm1 FF	CACCATGCTTGACGGTTTCGTGCACTG
Ecm1 RR	AAAAAACGGAGGCAGCTGTGCACAG
Ecm6mod1 FF	CACCATGGGTTGTACCGGCGTCA
Ecm6mod1 RR	GGAGACGTCGAGGCGTACGG
Ecm6mod1_2 FF	TATGGATCCATGGGTTGTACCGGCGTCACATCC

Ecm6mod1_2 RR	TATGAATTCTTCCGGCGCCGGGATCCAGC
Ecm6mod1 seq1	TACCGGCCGGCGCACCGAA
Ecm6mod1 seq2	GTCGTCAACGTCCTCCCCGC
Ecm6mod1 seq3	TCGCCTTCATGCTCGCCGAC
Ecm6mod1 seq4	CTGCACAACCTGTACGGGCC
Ecm6mod1 seq5	GTTCGTCGCACTCGACCGCC
Ecm6C1 FF	ATAGGATCCGTCACATCCGCGCAGGAAG
Ecm6C1 RR	ATACTCGAGCTACGCGGTGACGGTCCCGAC
C1 gRNA P1	ACGCGGACGCCGAACGCGTCGGTG
C1 gRNA P2	AAACCACCGACGCGTTCGGCGTCC
C1 Homology Left FF	TCGGTTGCCGCCGGGCGTTTTTTATTTGAACACAA TGCCCTCCAG
C1 Homology Left RR	AGCAGCGTGAGGACGCCGAAGAAGACGAGCACG AGCAGGT
C1 Homology Right FF	ACCTGCTCGTGCTCGTCTTCTTCGGCGTCCTCACG CTGCT
C1 Homology Right RR	GCGGCCTTTTTACGGTTCCTGGCCTCGGACGTACA GGTCGCTGCC
C2 gRNA P1	ACGCGCCCATCACGTGCTGCTCGA
C2 gRNA P2	AAACTCGAGCAGCACGTGATGGGC
C2 Homology Left FF	TCGGTTGCCGCCGGGCGTTTTTTATACGACATCGGC TTCAACTTC
C2 Homology Left RR	CCGATGAGCGGCACGGACCAGGCGGTGAGGGCCAG TTCGG

C2 Homology Right FF	CCGAACTGGCCCTCACCGCCTGGTCCGTGCCGCTCA TCGG
C2 Homology Right RR	GCGGCCTTTTTACGGTTCCTGGCCTCGGTA GGAGATCCG
gBlock for pCAP03	CCCATGGTATAAATAGTGGCCTGTCTTTGTCCAGAC GAAGTCACTAGCCTGGCTTCGTCAGGGGTTGGTCAC CGACCGTCGTTTAAACGGAACTGGATGGCCGTCCGA ATCCTCGGCAAGGAGCTGGGCGACCGCTTCGAAGGC GGCGTATGTCGAAAGCTACATATA
Ecm7A1 FF	ATAGAATTTCGGCAGGTCCGGCAACTCGCT
Ecm7A2 RR	ATACTCGAGCTAGGTGTCGGGGACGACGTAGG
pCAP1000 Homology Left FF	GCGCCGATGGTTTCTACAAAGATCGGGGCGGCATC AACTGGGT
pCAP1000 Homology Left RR	TAGCCATTTTCTTAGTTGTTGTAGGCGACGGTCCGGT GACCAAC
pCAP1000 Homology Right FF	GAGAAGATGCGGCCAGCAAACTAAACCTGGGCG GGACAGCTC
pCAP1000 Homology Left RR	TTTTCTAAATACAGGTACCTCAAGTGGCTGATCAGA GGGAAGGTGTG
pCAP03 Sequencing	TACTTGGCGGATAATGCC

Table 12 - Table of oligonucleotides (mostly PCR primers) used in this work.

7.3.4 Plasmids

Plasmid (Resistance)	Description
----------------------	-------------

pET28a(+) (Kan ^R)	Expression vector with T7 promoter for N-terminal His ₆ -proteins
pET151-D-TOPO (Amp ^R)	pET151 vector linearised by DNA topoisomerase I
pCRISPomyces2 (Apra ^R)	Vector for the expression of <i>S. pyogenes</i> Cas9 (Corre group)
pCRISPomyces2-C1 (Apra ^R)	pCRISPomyces2 vector containing a DNA to code a guide RNA for Ecm6 condensation domain 1 and homology arms for HDR
pCRISPomyces2-C2 (Apra ^R)	pCRISPomyces2 vector containing a DNA to code a guide RNA for Ecm6 condensation domain 2 and homology arms for HDR
pCAP03 (Kan ^R)	Empty TAR capture vector (Corre group)
pCAP1000 (Kan ^R)	Empty TAR capture vector (Corre group)
pCAP03 Capture vector (Kan ^R)	TAR capture vector containing two 60 bp arms for homologous recombination (generated in this PhD work)
pCAP1000 Capture vector (Kan ^R)	TAR capture vector containing two 1000 bp arms for homologous recombination (generated in this PhD work)
pCAP03- <i>gbn</i> (Kan ^R)	TAR capture vector containing the gaburedin BGC (Corre group)

pET151-FabC (Amp ^R)	pET151 vector containing <i>fabC</i> from the echinomycin gene cluster (generated in this PhD work)
pET151-Ecm1 (Amp ^R)	pET151 vector containing <i>ecm1</i> from the echinomycin gene cluster (generated in this PhD work)
pET151-Ecm6mod1 (Amp ^R)	pET151 vector containing <i>ecm6</i> module 1 from the echinomycin gene cluster (generated in this PhD work)
pET28a(+)-Ecm6C1 (Kan ^R)	pET28a vector containing <i>ecm6</i> module 1 condensation domain from the echinomycin gene cluster (generated in this PhD work)

Table 13 - Table of plasmids used and generated in this work.

7.3.5 Buffers

Solution	Components
Tris-Borate EDTA (TBE) buffer (10 x)	108 g Tris, 55 g boric acid, 40 mL 0.5 M EDTA, filled to 1 L with H ₂ O and adjusted to pH 8.3.
Tris-EDTA (TE) buffer	100 µl 1 M Tris pH 8.0, 20 µL 0.5 M EDTA pH 8, 9.88 ml H ₂ O
Sodium chloride-TE-EDTA (STE) buffer	44 mg NaCl, 500 µL 0.5 M EDTA pH 8, 200 µL 1M Tris pH 7.5, 9.30 mL dH ₂ O

1.5 M Tris.HCl (pH 8.8)	18.1 g Tris, 100 mL H ₂ O, and adjusted to pH 8.8 with HCl.
1 M Tris.HCl (pH 6.8)	12.1 g Tris, 100 mL H ₂ O, and adjusted to pH 6.8 with HCl.
SDS-PAGE loading dye (4 x)	2 mL 1M Tris.HCl pH 6.8, 0.8 g SDS, 4 mL glycerol, 4 mL 2-mercaptoethanol, 10 µL 10% bromophenol blue.
SDS-PAGE running buffer (10 x)	30 g Tris, 144 g glycine, 10 g SDS, filled to 1 L with H ₂ O, and adjusted to pH 8.3.
SDS-PAGE staining solution	10 mL AcOH, 45 mL H ₂ O, 45 mL MeOH, 0.25 g Coomassie brilliant blue R250.
SDS-PAGE destaining solution	10 mL AcOH, 45 mL H ₂ O, 45 mL MeOH.
Protein lysis buffer (20 mM Tris, 150 mM NaCl, 20 % glycerol, pH 8.0)	2.42 g Tris, 8.76 g NaCl, 200 mL glycerol, filled to 1 L with H ₂ O, and adjusted to pH 8.0.
Protein elution buffer (20 mM Tris, 150 mM NaCl, 20 % glycerol, 300 mM imidazole, pH 8.0)	2.42 g Tris, 8.76 g NaCl, 200 mL glycerol, 20.42 g imidazole, filled to 1 L with H ₂ O, and adjusted to pH 8.0.
Protein desalting buffer (20 mM Tris, 100 mM NaCl, 5% glycerol)	2.42 g Tris, 5.84 g NaCl, 50 mL glycerol, filled to 1 L H ₂ O, and adjusted to pH 8.0.

Table 14 - Table of buffers and solutions used in this work. Not autoclaved unless specified.

7.3.6 Antibiotics

Antibiotic	Stock concentration mg mL ⁻¹ (solvent)	Working concentration µg mL ⁻¹
Kanamycin (Kan)	50 (H ₂ O)	50

Chloramphenicol (Cm)	35 (EtOH)	35
Ampicillin (Amp)	50 (H ₂ O)	50
Nalidixic acid (Nal)	25 (0.15 M NaOH _(aq))	25
Apramycin (Apra)	50 (H ₂ O)	50

Table 15 - Table of antibiotics used in this work.

7.2.7 Procedures

7.2.7.1 Isolation of genomic DNA from *S. lasaliensis*

10 mL of M79 medium was inoculated with 100 μ L of *S. lasaliensis* glycerol stock and incubated at 30 °C with shaking for 3 days. *S. lasaliensis* culture (500 μ L) was spun down in a 2 mL Eppendorf tube and the supernatant discarded. The cells were resuspended in 300 μ L STE buffer (75 mM NaCl, 25 mM EDTA pH 8, 20 mM Tris-HCl pH 7.5) and 10 μ L of lysozyme solution (50 mg ml⁻¹) was added. After 30 minutes of incubation at 37 °C, 10 % SDS solution (60 μ L) was added and the cells were mixed. Proteinase K solution (7 μ L, 25mg/ml) was then added and the samples were incubated for a further 1 hour at 55 °C. This was mixed with 5 M NaCl (150 μ L) and chloroform (500 μ L) before being spun down by centrifugation at 13,000 rpm for 10 minutes. The upper aqueous layer was transferred to a new 1.5 mL Eppendorf tube and the DNA was precipitated by addition of iso-propanol (700 μ L) and spun down by centrifugation at 13,000 x g for 10 minutes. The DNA was washed twice with 200 μ L 70 % ethanol, air dried and resuspended in 50 μ L TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

7.2.7.2 Polymerase chain reactions (Q5 polymerase)

For PCR reactions using Q5 polymerase, a master mix consisting of 5 μ L 5x Q5 reaction buffer, 5 μ L 5x Q5 GC enhancer, 1 μ L of 2 mM dNTPs, 1.25 μ L forward primer (10 μ M), 1.25 μ L reverse primer (10 μ M), 1 μ L template DNA (50 ng/ μ L), 0.25 μ L Q5 polymerase, and 9.25 μ L dH₂O. PCR was conducted using the conditions in the Table 7 and annealing temperatures in Table 8, calculated using NEB T^m

calculator. PCR products were analysed by agarose gel electrophoresis in TBE buffer using a gel of 1.0% agarose in TBE buffer with GelRed under UV light for detection.

Initial denaturation	98 °C	120 s
Denaturation	98 °C	10 s
Annealing	Variable	20 s
Extension	72 °C	20 s/kb
Final Extension	72 °C	300 s

Table 16 - Table of PCR cycle conditions used with Q5 polymerase.

DNA amplified	Annealing temperature
FabC	72 °C
Ecm6mod1	71 °C
Ecm6mod1_2	72 °C
Ecm6C1	72 °C
C1 left arm	67 °C
C1 right arm	72 °C
C2 left arm	67 °C
C2 right arm	67 °C
pCAP1000 left arm	70 °C
pCAP1000 right arm	70 °C

Table 17 - Annealing temperatures for genes cloned by PCR.

Cloning of Ecm1 using Phusion polymerase

Ecm1 was amplified using Phusion polymerase with conditions derived from literature.³⁴⁰

Mixture: 1.25 μL forward primer (Ecm1 FF, 10 μM), 1.25 μL reverse primer (Ecm1 RR, 10 μM), 1 μL template DNA (50 $\text{ng}/\mu\text{L}$), 2.5 μL dNTPs (2 mM), 5 μL Phusion GC buffer, 0.25 μL Phusion polymerase, 1.5 μL DMSO, 12.25 μL dH_2O .

Conditions: 98 $^\circ\text{C}$ 30s Denaturation
98 $^\circ\text{C}$ 30s Denaturation
71 $^\circ\text{C}$ 30s Annealing
72 $^\circ\text{C}$ 120s Extension
72 $^\circ\text{C}$ 600s Extension

PCR products were analysed by agarose gel electrophoresis in TBE buffer using a gel of 1.0% agarose in TBE buffer with GelRed under UV light for detection.

7.2.7.3 TOPO cloning

Purified PCR product (15 ng) was mixed with TOPO vector (1 μL), salt solution (1.2 M NaCl , 0.06 M MgCl_2) (1 μL), and made up to 6 μL with dH_2O . The reaction was left at room temperature for 10 minutes, and 3 μL used to transform TOP10 chemically competent cells.

7.2.7.4 Digestion and ligation

PCR products were digested after purification using agarose gel electrophoresis and extracted using a gel extraction kit. Plasmids were digested after isolation using a miniprep kit. Restriction digests of plasmids and PCR products were conducted at 37 $^\circ\text{C}$ for 3 hours using 1 μL of restriction enzyme, and the buffer conditions suggested by the manufacturer. Double digests were conducted using appropriate buffer as suggested by New England Biolabs NEBcloner or Thermo Scientific Double Digest Calculator. The digested fragments were purified by gel electrophoresis and gel extraction.

Ligation was performed using a 3:1 molar ratio of insert to vector. T4 ligase (1 μL) and 10x ligase buffer (2 μL) were mixed with vector (50 ng) and insert and made up to 20 μL with dH_2O . The reaction was incubated for 40 minutes at room temperature, and 5 μL used to transform TOP10 chemically competent cells.

7.2.7.5 Transformation

TOP10 chemically competent cells (50 μ L aliquots) were thawed on ice. When just thawed, DNA was added, and the mixture placed on ice for 30 minutes. The mixture was then heat shocked at 42 °C for 30 s and placed on ice for 2 minutes. 200 μ L of SOC media was added to each tube, and cells recovered for 1 hour at 37 °C with shaking. 50 – 200 μ L of transformation mixture were spread onto plates with appropriate selection and incubated overnight at 37 °C. Plasmids were checked by restriction digestion and analysis of diagnostic DNA fragments; and sequenced using Eurofins Genomics SupremeRun service.

7.2.7.6 Protein expression and purification

For protein expression, FabC-pET151 and Ecm6mod1-pET151 plasmids were used to transform *E. coli* BAP1. Ecm6mod1-pET151, Ecm6C1 and Ecm1-pET151 were transformed into *E. coli* BL21. 5 mL of an overnight culture was used to inoculate 500 mL LB media containing appropriate selection. This culture was grown at 37 °C at 180 rpm until an OD₆₀₀ of 0.4 – 0.8 was reached. Protein expression was then induced adding isopropyl- β -D-thiogalactoside (IPTG, 0.5 mM final concentration) and the cultures grown overnight at 15 °C. *E. coli* cells were pelleted (9,000 rpm, 20 mins, 4 °C) and the supernatant discarded. Cells were resuspended in wash buffer (20 mM Tris, 150 mM NaCl, 20 % glycerol, pH 8.0) by vortexing, and cells lysed using a cell disruptor. The suspension was then pelleted (20,000 rpm, 20 mins, 4 °C) and the supernatant passed through a Ni-nitriloacetic acid (NTA) resin affinity column to bind His₆-tagged recombinant protein. The column bound-protein was then eluted with elution buffer (20 mM Tris, 150 mM NaCl, 20 % glycerol, 300 mM imidazole, pH 8.0). Ecm1 and Ecm6C1 protein samples were run on an 10 % SDS-PAGE gel, whereas FabC was run on a 15 % SDS-PAGE gel to confirm the presence of proteins of the correct size.

Ecm6mod1 was expressed as described above; it was also expressed after inoculating 1 L of LB media with 1 mL of overnight preculture. The 1L culture was incubated for 60 hours at 15 °C, before inducing expression with IPTG (0.5 mM final concentration) for 11 hours at 15 °C before cell lysis and protein purification as mentioned above. Ecm6mod1 samples were analysed on an 8 % SDS-PAGE gel.

The elution mixture was concentrated to 2.5 mL using Vivaspin protein concentrator spin columns, and the solution loaded onto a PD-10 desalting column containing Sephadex G-25. The protein was eluted using elution buffer (20 mM Tris, 100 mM NaCl, 5% glycerol). The desalted protein solution was further concentrated, then frozen with liquid N₂ and stored at -80 °C.

7.2.7.7 SDS-PAGE analysis of expressed proteins

Separating gel	8% (mL)	10% (mL)	15% (mL)
H ₂ O	2.3	2.0	1.2
30% Acrylamide	1.3	1.7	2.5
1.5 M Tris (pH 8)	1.3	1.3	1.3
10% SDS	0.05	0.05	0.05
10% APS	0.05	0.05	0.05
Tetramethylethylenediamine (TEMED)	0.003	0.002	0.002

Stacking gel (3 mL): 2.1 mL H₂O, 0.5 mL 30% acrylamide mix, 0.38 mL 1 M Tris (pH 6.8), 30 µL 10% SDS, 30 µL 10% SDS, 3 µL TEMED.

20 µL of protein solution was run on the SDS-PAGE gel, loaded onto the stacking gel. For elution protein fractions this consisted of 10 µL protein solution and 10 µL protein loading dye. For insoluble, flow-through, and soluble protein fractions, this consisted of 4 µL protein solution, 10 µL protein loading dye, and 6 µL dH₂O. The gel was run at 180 V for 45 mins in 1x SDS-PAGE buffer. The gel was stained using a Coomassie blue solution, before de-staining and viewing with white light.

7.2.7.8 Adenylation domain assays

Hydroxylamine trapping assay¹⁰⁹

50 mM Tris buffer, pH 8

300 mM MgCl₂ in 25 mM Tris, pH 8

30 mM ATP in H₂O

2 M hydroxylamine in 3.5 M NaOH, adjusted to pH 8 with 12M HCl

60 mM carboxylic acid, adjusted to pH 8

10% w/v FeCl₃.6H₂O and 3.3% w/v trichloroacetic acid in 0.7 M HCl (stopping solution)

Ecm1: 50 μM in 25 mM Tris buffer, pH 8

Procedure: To a 1.5 mL Eppendorf tube, 150 μL Tris buffer, 15 μL MgCl₂ solution, 22.5 μL ATP solution, 22.5 μL hydroxylamine solution, and 15 μL carboxylic acid solution were added and the volume adjusted to 300 μL with dH₂O. The reaction was initiated by adding 15 μL of 50 μM enzyme solution, or 15 μL of boiled enzyme solution to the control, and the tube kept at 37 °C for 1 hour. 300 μL of stopping solution was added to each tube. The tubes were centrifuged at 14,000 rpm for 5 minutes to pellet the enzymes. The supernatant was transferred to 1.5 mL plastic cuvette and the UV-Vis absorbance spectra recorded.

Malachite Green assay

Malachite green solution: 44 mg of malachite green oxalate were dissolved in 30 mL dH₂O; 6 mL of conc. H₂SO₄ were added, and returned to room temperature after addition of the acid.

Ammonium molybdate solution: 750 mg ammonium molybdate tetrahydrate was dissolved in 10 mL dH₂O.

Tween 20 solution: 110 μL of Tween 20 was mixed with 890 μL dH₂O.

The colour reagent was formed by mixing malachite green: ammonium molybdate: Tween 20 in a 1: 0.25 : 0.02 ratio.

2 x assay buffer: 100 mM Tris.HCl, 200 mM NaCl, 20 mM MgCl₂, adjusted to pH 8.
1 x Assay buffer: 50 mM Tris.HCl, 100 mM NaCl, 10 mM MgCl₂, adjusted to pH 8.
20 mM carboxylic acid, adjusted to pH 8.
25 mM ATP in 2 x assay buffer
1 U/ml yeast inorganic pyrophosphatase (YIPP) in 1x assay buffer
12.5 μM Ecm1 in 1x assay buffer

Procedure: 15 μL of carboxylic acid solution, 5 μL 2x assay buffer, 10 μL YIPP solution, and 10 μL of Ecm1 solution were mixed in a transparent 96-well plate. The assay was initiated by addition of 5 μL ATP solution, and incubated at 37 °C for 1 hour. 12 μL of colour reagent was added and gently mixed. After 1 minute, 10 μL of 34% w/v trisodium citrate was added and gently mixed. The colour was allowed to develop for 30 minutes before the absorbance at 620 nm was recorded on a Hidex Sense Microplate Reader.

7.2.7.9 Condensation domain assay

2x Assay buffer: 100 mM Tris.HCl, 200 mM NaCl, 20 mM MgCl₂, adjusted to pH 8.
1 x Assay buffer: 50 mM Tris.HCl, 100 mM NaCl, 10 mM MgCl₂, adjusted to pH 8.
20 mM carboxylic acid, adjusted to pH 8.
50 mM ATP in 2x assay buffer
20 mM chemical probe in 1x buffer with 15% DMSO
10 mM TCEP in 1x assay buffer
2.5 μM Ecm6C1 in 1x buffer
10 μM FabC in 1x buffer
5 μM Ecm1 in 1x buffer

Procedure: 15 μL of 1x assay buffer, 5 μL carboxylic acid solution, 5 μL chemical probe solution, 5 μL TCEP solution, 5 μL Ecm6C1 solution, 5 μL FabC solution, and 5 μL Ecm1 solution were gently mixed. The assay was initiated by the addition of 5 μL ATP solution, and incubated at 37 °C for 1 hour. The reactions were then quenched with 50 μL 0.1M KOH, and heated to 60 °C for 5 minutes. The reactions were cooled

on ice and extracted with 0.5 mL EtOAc. The EtOAc extracts were concentrated *in vacuo* and dissolved in 130 μ L HPLC-grade MeOH. 100 μ L was run on an Agilent 1200 HPLC with a Phenomenex Synergi C18 Fusion-RP analytic column, recording UV absorbance at 254 nm, and quantified by comparison against an authentic standard.

7.2.7.10 pCRISPomyces2

The protospacer oligonucleotides were designed according the literature.¹²⁷ The oligonucleotides were diluted to 100 μ M in dH₂O. 5 μ L of both oligonucleotides were mixed, and added to 90 μ L 30 mM HEPES, pH 7.8. The mixture was heated to 95 °C for 5 minutes and cooled to 4 °C at 0.1 °C/s.

The protospacer was inserted into the vector by Golden Gate assembly. 100 ng of vector was mixed with 0.3 μ L protospacer, 2 μ L T4 ligase buffer, 1 μ L T4 ligase, 1 μ L BbsI, and made up to 20 μ L with dH₂O. This program used was as follows:

Temperature (°C)	Time (mins)
37	10
16	10
Go to step 1, 9 times	
50	5
65	20

3 μ L of the reaction were used to transform TOP10 chemically competent cells, and plated onto apramycin selective plates. The plasmid was extracted using a miniprep kit, and the purified plasmid digested by XbaI. The homology arms were cloned by PCR and inserted using Gibson assembly. 75 ng of vector were mixed with 18 ng of each homology arm, 10 μ L of 2x Gibson master mix, and adjusted to 20 μ L with dH₂O. The reaction was incubated at 50 °C for 1 hour, and 2 μ L used to transform TOP10 chemically competent cells and plated onto apramycin selective plates. The plasmid was extracted using a miniprep kit and transformed into *E. coli* ET12567/pUZ8002 by

electroporation. 2 μ L of plasmid was added to 50 μ L electrocompetent cells and transferred to an ice-cold cuvette. The cells were electroporated at 2.5 kV, 25 μ F. 200 μ L SOC medium was immediately added, and cells recovered at 37 °C with shaking for 1 hour. The cells were plated onto chloramphenicol, apramycin, and kanamycin selective plates.

The strain was then used to attempt conjugation into *S. lasaliensis*.

7.2.7.11 Streptomyces spore preparation

S. lasaliensis was grown for 5 days at 30 °C on SFM medium. 3.5 mL of sterile dH₂O was added to the plates, and the spores collected with a sterile spreader. The suspension was filtered through sterile cotton wool into a falcon tube. The suspension was centrifuged at 3,000 rpm for 5 minutes and the supernatant removed. The spores were resuspended in 500 μ L 30% glycerol at used directly or stored at -80 °C.

7.2.7.12 Conjugation

1. An *E. coli* colony was used to inoculate 10 mL LB containing apramycin, chloramphenicol, and kanamycin, and incubated overnight at 37 °C.
2. 200 μ L of overnight culture was used to inoculate 10 mL LB + Kan + Cm + Apra, and grown at 37 °C to OD₆₀₀ = 0.4.
3. The cells were washed twice with LB and resuspended in residual LB.
4. 10 μ L of *S. lasaliensis* spores were added to 500 μ L 2xYT medium, heat shocked at 50 °C for 10 minutes and allowed to cool.
5. 500 μ L of *E. coli* suspension and 500 μ L heat-shocked spores and spun down for 10 seconds. The majority of the supernatant was removed, and the residual liquid used to resuspend.
6. 50 μ L and 100 μ L were plated onto SFM with 10 mM MgCl₂ and incubated for 18 hours at 30 °C.
7. The plates were overlaid with 1 mL H₂O containing 0.5 mg nalidixic acid and 1.25 mg apramycin, and incubated at 30 °C.
8. Plates were checked after 72 hours for colonies.

7.2.7.13 TAR cloning

TAR cloning was carried out according to the literature procedure.¹¹⁸ *S. lasaliensis* genomic DNA was digested using NsiI and MseI. 50 µL genomic DNA (2,000 ng/ µL) was mixed with 2 µL NsiI, 2 µL MseI, and 6 µL CutSmart buffer. After overnight incubation at 37 °C, 6 µL of NaOAc (3M, pH 5.2) was added, and 165 µL 95% EtOH. The mixture was incubated at room temperature for 30 minutes, then centrifuged for 30 minutes at 15,000 rpm, 10 °C. The supernatant was discarded, and the DNA pellet washed with 200 µL 70% EtOH. The mixture was centrifuged again for 15 minutes at 15,000 rpm, 10 °C, the supernatant discarded, and the pellet dissolved in TE buffer (30 µl).

After digestion of genomic DNA and PmeI digestion of pCAP03 and pCAP1000 capture vectors and transformation of *S. cerevisiae* VL6-48N, colonies growing on selective plates containing 5-fluoroorotic acid were re-plated on selective plates. DNA was extracted according to the literature protocol and screened by PCR.¹¹⁸

8. Bibliography

- 1 *Nat. Chem. Biol.*, 2007, **3**, 351–351.
- 2 J. Choi, J. Chen, S. L. Schreiber and J. Clardy, *Science*, 1996, **273**, 239–242.
- 3 A. Fleming, *Br. J. Exp. Pathol.*, 1929, 226–236.
- 4 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2016, **79**, 629–661.
- 5 I. Paterson and E. A. Anderson, *Science (80-.)*, 2005, **310**, 451–454.
- 6 E. Banin, D. Hughes and O. P. Kuipers, *FEMS Microbiol. Rev.*, 2017, **41**, 450–452.
- 7 B. T. Hovemann, R. Ryseck, U. Walldorf and K. F. Sto, *Gene*, 1998, **221**, 1–9.
- 8 B. Mach, E. Reich and E. L. Tatum, *Proc. Natl. Acad. Sci.*, 1963, **50**, 175–181.
- 9 U. Spaeren, L. Froholm and S. Laland, *Biochem. J.*, 1967, **102**, 586–592.
- 10 W. Gevers, H. Kleinkauf and F. Lipmann, *Proc. Natl. Acad. Sci. U. S. A.*, 1968, **60**, 269–76.
- 11 W. Gevers, H. Kleinkauf and F. Lipmann, *Proc. Natl. Acad. Sci. U. S. A.*, 1969, **63**, 1335–1342.
- 12 Frøshov, T. L. Zimmer and S. G. Laland, *FEBS Lett.*, 1970, **7**, 68–71.
- 13 C. C. Gilhuus-Moe, T. Kristensen, J. E. Bredesen, T.-L. Zimmer and S. G. Laland, *FEBS Lett.*, 1970, **7**, 287–290.
- 14 H. Kleinkauf, W. Gevers, R. Roskoski and F. Lipmann, *Biochem. Biophys. Res. Commun.*, 1970, **41**, 1218–1222.
- 15 S. G. Lee, R. Roskoski, K. Bauer and F. Lipmann, *Biochemistry*, 1973, **12**, 398–405.
- 16 H. D. Mootz and M. A. Marahiel, *J. Bacteriol.*, 1997, **179**, 6843–6850.

- 17 G. Mittenhuber, R. Weckermann and M. A. Marahiel, *J. Bacteriol.*, 1989, **171**, 4881–4887.
- 18 M. A. Fischbach and C. T. Walsh, *Chem. Rev.*, 2006, **106**, 3468–3496.
- 19 H. Jenke-Kodama and E. Dittmann, *Nat. Prod. Rep.*, 2009, **26**, 874–883.
- 20 S. Caboche, M. Pupin, V. Leclère, A. Fontaine, P. Jacques and G. Kucherov, *Nucleic Acids Res.*, 2008, **36**, 326–331.
- 21 F. H. C. Crick, *J. Mol. Biol.*, 1976, **38**, 367–379.
- 22 R. Dieckmann, Y.-O. Lee, H. van Liempt, H. von Döhren and H. Kleinkauf, *FEBS Lett.*, 1995, **357**, 212–216.
- 23 A. Lawen and R. Traber, *J. Biol. Chem.*, 1993, **268**, 20452–20465.
- 24 M. a Marahiel, T. Stachelhaus and H. D. Mootz, *Chem. Rev.*, 1997, **97**, 2651–2674.
- 25 E. Conti, T. Stachelhaus, M. A. Marahiel and P. Brick, *EMBO J.*, 1997, **16**, 4174–4183.
- 26 J. J. May, N. Kessler, M. A. Marahiel and M. T. Stubbs, *Proc. Natl. Acad. Sci.*, 2002, **99**, 12120–12125.
- 27 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 1999, **6**, 493–505.
- 28 G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol.*, 2000, **7**, 211–224.
- 29 M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, *Nucleic Acids Res.*, 2011, **39**, 362–367.
- 30 K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano and T. Weber, *Nucleic Acids Res.*, 2013, **41**, W204–W212.
- 31 H. Yonus, P. Neumann, S. Zimmermann, J. J. May, M. A. Marahiel and M. T. Stubbs, *J. Biol. Chem.*, 2008, **283**, 32484–32491.
- 32 C. A. Mitchell, C. Shi, C. C. Aldrich and A. M. Gulick, *Biochemistry*, 2012, **51**, 3252–3263.

- 33 S. Meyer, J. C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süßmuth and E. Dittmann, *Cell Chem. Biol.*, 2016, **23**, 462–471.
- 34 A. Degen, F. Mayerthaler, H. D. Mootz and B. Di Ventura, *Sci. Rep.*, 2019, **9**, 1–12.
- 35 T. Stachelhaus, A. Hüser and M. A. Marahiel, *Chem. Biol.*, 1996, **3**, 913–921.
- 36 R. H. Lambalot and C. T. Walsh, *J. Biol. Chem.*, 1995, **270**, 24658–24661.
- 37 T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel and T. A. Holak, *Structure*, 2000, **8**, 407–418.
- 38 V. V Rogov, B. Scha, A. Koglin, M. R. Mofid, F. Lo, M. Blum, T. Mittag, M. A. Marahiel and F. Bernhard, *Science (80-.)*, 2006, 273–276.
- 39 K. J. Weissman and R. Müller, *ChemBioChem*, 2008, **9**, 826–848.
- 40 V. De Crecy-Lagard, P. Marliere and W. Saurin, *Comptes Rendus l'Académie des Sci. - Ser. III*, 1995, **318**, 927–936.
- 41 T. Stachelhaus, H. D. Mootz, V. Bergendahl and M. a Marahiel, *J. Biol. Chem.*, 1998, **273**, 22773–22781.
- 42 V. Bergendahl, U. Linne and M. A. Marahiel, *Eur. J. Biochem.*, 2002, **269**, 620–629.
- 43 C. G. Marshall, N. J. Hillson and C. T. Walsh, *Biochemistry*, 2002, **41**, 244–250.
- 44 T. A. Keating, C. G. Marshall, C. T. Walsh and A. E. Keating, *Nat. Struct. Biol.*, 2002, **9**, 522.
- 45 K. Bloudoff, D. Rodionov and T. M. Schmeing, *J. Mol. Biol.*, 2013, **425**, 3137–3150.
- 46 K. Bloudoff, D. A. Alonzo and T. M. Schmeing, *Cell Chem. Biol.*, 2016, **23**, 331–339.
- 47 P. J. Belshaw, C. T. Walsh and T. Stachelhaus, *Science (80-.)*, 1999, **284**, 486–489.

- 48 D. E. Ehmman, J. W. Trauger, T. Stachelhaus and C. T. Walsh, *Chem. Biol.*, 2000, **7**, 765–772.
- 49 U. Linne and M. A. Marahiel, *Biochemistry*, 2000, **39**, 10439–10447.
- 50 C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, *BMC Evol. Biol.*, 2007, **7**, 78.
- 51 M. Schoppet, M. Peschke, A. Kirchberg, V. Wiebach, R. D. Süßmuth, E. Stegmann and M. J. Cryle, *Chem. Sci.*, 2018, 118–133.
- 52 A. M. Gehring, I. Mori, R. D. Perry and C. T. Walsh, *Biochemistry*, 1998, **37**, 11637–11650.
- 53 T. Duerfahrt, K. Eppelmann, R. Müller and M. A. Marahiel, *Chem. Biol.*, 2004, **11**, 261–271.
- 54 L. F. Povirk, M. Hogan and N. Dattagupta, *Biochemistry*, 1979, **18**, 96–101.
- 55 T. L. Schneider, B. Shen and C. T. Walsh, *Biochemistry*, 2003, **42**, 9722–9730.
- 56 T. A. Keating, D. A. Miller and C. T. Walsh, *Biochemistry*, 2000, **39**, 4729–4739.
- 57 K. Bloudoff, C. D. Fage, M. A. Marahiel and T. M. Schmeing, *Proc. Natl. Acad. Sci.*, 2017, **114**, 95–100.
- 58 R. Bessalle, A. Kapitkovsky, A. Gorea, I. Shalit and M. Fridkin, *FEBS Lett.*, 1990, **274**, 151–155.
- 59 T. Stachelhaus and M. A. Marahiel, *J. Biol. Chem.*, 1995, 270, 6163–6169.
- 60 L. Luo, R. M. Kohli, M. Onishi, U. Linne, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2002, **41**, 9184–96.
- 61 T. Stachelhaus and C. T. Walsh, *Biochemistry*, 2000, **39**, 5775–5787.
- 62 S. A. Samel, P. Czodrowski and L.-O. Essen, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2014, **70**, 1442–1452.
- 63 C. J. Balibar, F. H. Vaillancourt and C. T. Walsh, *Chem. Biol.*, 2005, **12**,

1189–1200.

- 64 K. Haslinger, M. Peschke, C. Brieke, E. Maximowitsch and M. J. Cryle, *Nature*, 2015, **521**, 105–109.
- 65 G. H. Hur, C. R. Vickery and M. D. Burkart, *Nat. Prod. Rep.*, 2012, **29**, 1074–1098.
- 66 T. Velkov, J. Horne, M. J. Scanlon, B. Capuano, E. Yuriev and A. Lawen, *Chem. Biol.*, 2011, **18**, 464–475.
- 67 C. Mahlert, F. Kopp, J. Thirlway, J. Micklefield and M. A. Marahiel, *J. Am. Chem. Soc.*, 2007, **129**, 12011–12018.
- 68 K. J. Labby, S. G. Watsula and S. Garneau-Tsodikova, *Nat. Prod. Rep.*, 2015, **32**, 641–653.
- 69 R. M. Kohli and C. T. Walsh, *Chem. Commun.*, 2003, **3**, 297–307.
- 70 S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh and M. T. Stubbs, *Structure*, 2002, **10**, 301–310.
- 71 T. A. Keating, D. E. Ehmman, R. M. Kohli, C. G. Marshall, J. W. Trauger and C. T. Walsh, *ChemBioChem*, 2001, **2**, 99–107.
- 72 L. Robbel, K. M. Hoyer and M. A. Marahiel, *FEBS J.*, 2009, **276**, 1641–1653.
- 73 K. M. Hoyer, C. Mahlert and M. A. Marahiel, *Chem. Biol.*, 2007, **14**, 13–22.
- 74 H. D. Mootz, J. W. Trauger, R. M. Kohli, C. T. Walsh and M. A. Marahiel, *Nature*, 2002, **407**, 215–218.
- 75 J. W. Trauger, R. M. Kohli and C. T. Walsh, *Biochemistry*, 2001, **40**, 7092–7098.
- 76 D. Schwarzer, H. D. Mootz, U. Linne and M. A. Marahiel, *Proc. Natl. Acad. Sci.*, 2002, **99**, 14083–14088.
- 77 E. Yeh, R. M. Kohli, S. D. Bruner and C. T. Walsh, *ChemBioChem*, 2004, **5**, 1290–1293.
- 78 K. Haslinger, E. Maximowitsch, C. Brieke, A. Koch and M. J. Cryle,

- ChemBioChem*, 2014, **15**, 2719–2728.
- 79 C. Brieke, M. Peschke, K. Haslinger and M. J. Cryle, *Angew. Chemie - Int. Ed.*, 2015, **54**, 15715–15719.
- 80 M. Peschke, C. Brieke, M. Heimes and M. J. Cryle, *ACS Chem. Biol.*, 2018, **13**, 110–120.
- 81 O. Yushchuk, B. Ostash, T. H. Pham, A. Luzhetskyy, V. Fedorenko, A. W. Truman and L. Horbal, *ACS Chem. Biol.*, 2016, **11**, 2254–2264.
- 82 H. D. Mootz, D. Schwarzer and M. A. Marahiel, *ChemBioChem*, 2002, **3**, 490.
- 83 P. Groves, M. S. Searle, J. P. Mackay and D. H. Williams, *Structure*, 1994, **2**, 747–754.
- 84 J. Recktenwald, R. Shawky, O. Puk, F. Pfennig, U. Keller, W. Wohlleben and S. Pelzer, *Microbiology*, 2002, **148**, 1105–1118.
- 85 C. C. Forneris and M. R. Seyedsayamdost, *Angew. Chemie - Int. Ed.*, 2018, **57**, 8048–8052.
- 86 H. Van Liempt and H. Kleinkauf, *Biochemistry*, 1989, 3680–3684.
- 87 V. Aharonowitz, J. Bergmeyer, J. M. Cantoral, G. Cohen, A. L. Demain, U. Fink, H. Kleinkauf, A. Maccabe, H. Palissa, E. Pfeifer, T. Schwecke, H. Van Liempt, H. Von Dohren, S. Wolfe and J. Zhang, *Nat. Biotechnol.*, 1993, **11**, 807–810.
- 88 R. Corbaz, L. Ettliger, E. Gaumann, W. Keller-Schierlein, F. Kradolfer, L. Neipp, V. Prelog, P. Reusser and H. Zahner, *Helv. Chim. Acta*, 1957, **40**, 199–204.
- 89 D. C. Ward, R. E and I. H. Golberg, *Science (80-.)*, 1965, **149**, 1259–1263.
- 90 K. Watanabe, K. Hotta, A. P. Praseuth, K. Koketsu, A. Migita, C. N. Boddy, C. C. C. Wang, H. Oguri and H. Oikawa, *Nat. Chem. Biol.*, 2006, **2**, 423–428.
- 91 K. Koketsu, H. Oguri, K. Watanabe and H. Oikawa, *Chem. Biol.*, 2008, **15**, 818–828.

- 92 G. F. Gause and M. G. Brazhnikova, *Nature*, 1944, 154, 703.
- 93 R. Schwyzer and P. Sieber, *Helv. Chim. Acta*, 1946, **715**, 624–639.
- 94 D. A. Miller, L. Luo, N. Hillson, T. A. Keating and C. T. Walsh, *Chem. Biol.*, 2002, **9**, 333–344.
- 95 A. O. Brachmann, C. Garcie, V. Wu, P. Martin, R. Ueoka, E. Oswald and J. Piel, *Chem. Commun.*, 2015, **51**, 13138–13141.
- 96 J. E. Baldwin, R. M. Adlington, J. W. Bird, R. A. Field, N. M. O’Callaghan and C. J. Schofield, *Tetrahedron*, 1992, **48**, 1099–1108.
- 97 S. W. Fuchs, A. Proschak, T. W. Jaskolla, M. Karas and H. B. Bode, *Org. Biomol. Chem.*, 2011, **9**, 3130–3132.
- 98 H. B. Bode, D. Reimer, S. W. Fuchs, F. Kirchner, C. Dauth, C. Kegler, W. Lorenzen, A. O. Brachmann and P. Grün, *Chem. - A Eur. J.*, 2012, **18**, 2342–2348.
- 99 A. Klitgaard, J. B. Nielsen, R. J. N. Frandsen, M. R. Andersen and K. F. Nielsen, *Anal. Chem.*, 2015, **87**, 6520–6526.
- 100 L. Zha, Y. Jiang, M. T. Henke, M. R. Wilson, J. X. Wang, N. L. Kelleher and E. P. Balskus, *Nat. Chem. Biol.*, 2017, **13**, 1063–1065.
- 101 M. Xue, C. S. Kim, A. R. Healy, K. M. Wernke, Z. Wang, M. C. Frischling, E. E. Shine, W. Wang, S. B. Herzon and J. M. Crawford, *bioRxiv*, 2019, 574053.
- 102 F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. E. O’Connor and C. T. Walsh, *Nature*, 2005, **436**, 1191–1194.
- 103 S. A. Sieber, C. T. Walsh and M. A. Marahiel, *J. Am. Chem. Soc.*, 2003, **125**, 10862–10866.
- 104 P. C. Schmartz, K. Wölfel, K. Zerbe, E. Gad, E. S. El-Tamany, H. K. Ibrahim, K. Abou-Hadeed and J. A. Robinson, *Angew. Chemie - Int. Ed.*, 2012, **51**, 11468–11472.
- 105 S. G. Lee and F. Lipmann, *Methods Enzymol.*, 1975, **43**, 585–602.

- 106 T. P. Geladopoulos, T. G. Sotiroudis and A. E. Evangelopoulos, *Anal. Biochem.*, 1991, **192**, 112–116.
- 107 T. J. Mcquade, A. D. Shallop, A. Sheoran, J. E. Delproposto and O. V Tsodikov, *Anal. Biochem.*, 2009, **386**, 244–250.
- 108 D. J. Wilson and C. C. Aldrich, *Anal. Biochem.*, 2010, **404**, 56–63.
- 109 N. Kadi and G. L. Challis, *Chapter 17 Siderophore Biosynthesis. A Substrate Specificity Assay for Nonribosomal Peptide Synthetase-Independent Siderophore Synthetases Involving Trapping of Acyl-Adenylate Intermediates with Hydroxylamine*, Elsevier Inc., 1st edn., 2009, vol. 458.
- 110 B. P. Duckworth, D. J. Wilson and C. C. Aldrich, in *Methods in Molecular Biology*, ed. B. S. Evans, Springer New York, New York, NY, 2016, vol. 1401, pp. 53–61.
- 111 E. D. Roche and C. T. Walsh, *Biochemistry*, 2003, **42**, 1334–1344.
- 112 W. L. Kelly and C. A. Townsend, *J. Biol. Chem.*, 2004, **279**, 38220–38227.
- 113 D. Bischoff, S. Pelzer, B. Bister, G. J. Nicholson, S. Stockert, M. Schirle, W. Wohlleben, G. Jung and R. D. Süßmuth, *Angew. Chemie - Int. Ed.*, 2001, **40**, 4688–4691.
- 114 H. Zhang, B. A. Boghigian, J. Armando and B. A. Pfeifer, *Nat. Prod. Rep.*, 2011, **28**, 125–151.
- 115 F. Malpartida and D. A. Hopwood, *Nature*, 1984, **309**, 462–464.
- 116 L. Huo, S. Rachid, M. Stadler, S. C. Wenzel and R. Müller, *Chem. Biol.*, 2012, **19**, 1278–1287.
- 117 P. D'Agostino and T. A. M. Gulder, *ACS Synth. Biol.*, 2018, acssynbio.8b00151.
- 118 K. Yamanaka, K. A. Reynolds, R. D. Kersten, K. S. Ryan, D. J. Gonzalez, V. Nizet, P. C. Dorrestein and B. S. Moore, *Proceeding Natl. Acad. Sci. United States Am.*, 2014, **111**, 1957–1962.
- 119 X. Tang, J. Li and B. S. Moore, *ChemBioChem*, 2017, **18**, 1072–1076.

- 120 T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*, 2000.
- 121 M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier, *Science (80-.)*, 2012, **337**, 816–822.
- 122 G. Gasiunas, R. Barrangou, P. Horvath and V. Siksnys, *Proc. Natl. Acad. Sci.*, 2012, **109**, E2579–E2586.
- 123 R. Kanaar, J. H. J. Hoeijmakers and D. C. Van Gent, *Trends Cell Biol.*, 1998, **8**, 483–489.
- 124 F. Alberti and C. Corre, *Nat. Prod. Rep.*, , DOI:10.1039/c8np00081f.
- 125 C. Engler, R. Kandzia and S. Marillonnet, *PLoS One*, , DOI:10.1371/journal.pone.0003647.
- 126 D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith, *Nat. Methods*, 2009, **6**, 343–5.
- 127 R. E. Cobb, Y. Wang and H. Zhao, *ACS Synth. Biol.*, 2015, **4**, 723–8.
- 128 J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse, *Science (80-.)*, 1989, **246**, 64–71.
- 129 N. L. Kelleher, H. Y. Lin, G. A. Valaskovic, D. J. Aaserud, E. K. Fridriksson and F. W. McLafferty, *J. Am. Chem. Soc.*, 1999, **121**, 806–812.
- 130 M. T. Mazur, C. T. Walsh and N. L. Kelleher, *Biochemistry*, 2003, **42**, 13393–13400.
- 131 A. G. Marshall, C. L. Hendrickson and G. S. Jackson, *Mass Spectrom. Rev.*, 1998, **17**, 1–35.
- 132 S. Garneau, P. C. Dorrestein, N. L. Kelleher and C. T. Walsh, *Biochemistry*, 2005, **44**, 2770–2780.
- 133 P. C. Dorrestein, S. B. Bumpus, C. T. Calderone, S. Garneau-Tsodikova, Z. D. Aron, P. D. Straight, R. Kolter, C. T. Walsh and N. L. Kelleher, *Biochemistry*, 2006, **45**, 12756–12766.

- 134 D. Meluzzi, W. H. Zheng, M. Hensler, V. Nizet and P. C. Dorrestein, *Bioorganic Med. Chem. Lett.*, 2008, **18**, 3107–3111.
- 135 M. J. Meehan, X. Xie, X. Zhao, W. Xu, Y. Tang and P. C. Dorrestein, *Biochemistry*, 2011, **50**, 287–299.
- 136 S. B. Bumpus, B. S. Evans, P. M. Thomas, I. Ntai and N. L. Kelleher, *Nat. Biotechnol.*, 2009, **27**, 951–956.
- 137 J. L. Meier, S. Niessen, H. S. Hoover, T. L. Foley, B. F. Cravatt and M. D. Burkart, *ACS Chem. Biol.*, 2009, **4**, 948–957.
- 138 C. Qiao, D. J. Wilson, E. M. Bennett and C. C. Aldrich, *J. Am. Chem. Soc.*, 2007, **129**, 6350–6351.
- 139 A. S. Worthington, H. Rivera, J. W. Torpey, M. D. Alexander and M. D. Burkart, *ACS Chem. Biol.*, 2006, **1**, 687–691.
- 140 R. Finking, A. Neumüller, J. Solsbacher, D. Konz, G. Kretzschmar, M. Schweitzer, T. Krumm and M. A. Marahiel, *ChemBioChem*, 2003, **4**, 903–906.
- 141 F. Ishikawa and H. Kakeya, *Bioorganic Med. Chem. Lett.*, 2014, **24**, 865–869.
- 142 S. Konno, F. Ishikawa, T. Suzuki, N. Dohmae, M. D. Burkart and H. Kakeya, *Chem. Commun.*, 2015, **51**, 2262–2265.
- 143 Y. Liu, T. Zheng and S. D. Bruner, *Chem. Biol.*, 2011, **18**, 1482–1488.
- 144 K. Haslinger, C. Brieke, S. Uhlmann, L. Sieverling, R. D. Süßmuth and M. J. Cryle, *Angew. Chemie - Int. Ed.*, 2014, **53**, 8518–8522.
- 145 G. Castro-Falcón, D. Hahn, D. Reimer and C. C. Hughes, *ACS Chem. Biol.*, 2016, **11**, 2328–2336.
- 146 G. Castro-Falcón, N. Millán-Aguiñaga, C. Roullier, P. R. Jensen and C. C. Hughes, *ACS Chem. Biol.*, 2018, **13**, 3097–3106.
- 147 N. A. Moss, G. Seiler, T. F. Leão, G. Castro-Falcón, L. Gerwick, C. C. Hughes and W. H. Gerwick, *Angew. Chemie Int. Ed.*, 2019, 1–6.

- 148 M. Tosin, Y. Demydchuk, J. S. Parascandolo, C. B. Per, F. J. Leeper and P. F. Leadlay, *Chem. Commun. (Camb)*., 2011, **47**, 3460–3462.
- 149 M. Tosin, L. Smith and P. F. Leadlay, *Angew. Chemie - Int. Ed.*, 2011, **50**, 11930–11933.
- 150 J. S. Parascandolo, J. Havemann, H. K. Potter, F. Huang, E. Riva, J. Connolly, I. Wilkening, L. Song, P. F. Leadlay and M. Tosin, *Angew. Chemie Int. Ed.*, 2016, **55**, 3463–3467.
- 151 E. Riva, I. Wilkening, S. Gazzola, W. M. A. Li, L. Smith, P. F. Leadlay and M. Tosin, *Angew. Chemie - Int. Ed.*, 2014, **53**, 11944–11949.
- 152 I. Wilkening, S. Gazzola, E. Riva, J. S. Parascandolo, L. Song and M. Tosin, *Chem. Commun.*, 2016, **52**, 1–4.
- 153 Y. T. C. C. Ho, D. J. Leng, F. Ghiringhelli, I. Wilkening, D. P. Bushell, O. Kostner, E. Riva, J. Havemann, D. Passarella and M. Tosin, *Chem. Commun.*, 2017, **53**, 7088–7091.
- 154 J. Fernández, L. Marín, R. Álvarez-Alonso, S. Redondo, J. Carvajal, G. Villamizar, C. J. Villar and F. Lombó, *Mar. Drugs*, 2014, **12**, 2668–2699.
- 155 G. J. Quigley, G. Ughetto, G. A. Van Der Marel, J. H. Van Boom, A. H. J. Wang and A. Rich, *Science (80-)*., 1986, **232**, 1255–1258.
- 156 L. G. May, M. A. Madine and M. J. Waring, *Nucleic Acids Res.*, 2004, **32**, 65–72.
- 157 M. M. Van Dyke and P. B. Dervan, *Science (80-)*., 1984, **225**, 1122–1127.
- 158 G. Ughetto, A. H. Wang, G. J. Quigley, G. A. Van Der Marel, J. H. Van and A. Rich, *Nucleic Acids Res.*, 1985, **13**, 2305–2323.
- 159 G. L. Semenza, *Nat. Rev. Cancer*, 2003, **3**, 721–732.
- 160 D. Kong, E. J. Park, A. G. Stephen, M. Calvani, J. H. Cardellina, A. Monks, R. J. Fisher, R. H. Shoemaker and G. Melillo, *Cancer Res.*, 2005, **65**, 9047–9055.
- 161 Y.-S. Park, W.-S. Shin, C.-S. Kim, C. M. Ahn, X.-F. Qi and S.-K. Kim, *Mol.*

- Cell. Toxicol.*, 2018, **14**, 9–18.
- 162 Y. S. Park, W. S. Shin and S. K. Kim, *J. Antimicrob. Chemother.*, 2008, **61**, 163–168.
- 163 Y. S. Park, Y. H. Kim, S. K. Kim and S.-J. Choi, *Bioorg. Med. Chem. Lett.*, 1998, **8**, 731–734.
- 164 J. B. Kim, G. S. Lee, Y. H. Y. B. Y. H. Kim, S. K. Kim and Y. H. Y. B. Y. H. Kim, *Int. J. Antimicrob. Agents*, 2004, **24**, 613–615.
- 165 U. Castillo, J. K. Harper, G. A. Strobel, J. Sears, K. Alesi, E. Ford, J. Lin, M. Hunter, M. Maranta, H. Ge, D. Yaver, J. B. Jensen, H. Porter, R. Robison, D. Millar, W. M. Hess, M. Condrón and D. Teplow, *FEMS Microbiol. Lett.*, 2003, **224**, 183–190.
- 166 H. Jayasuriya, D. L. Zink, J. D. Polishook, G. F. Bills, A. W. Dombrowski, O. Genilloud, F. F. Pelaez, L. Herranz, D. Quamina, R. B. Lingham, R. Danzeisen, P. L. Graham, J. E. Tomassini and S. B. Singh, *Chem. Biodivers.*, 2005, **2**, 112–122.
- 167 P. D. Minor and N. J. Dimmock, *Virology*, 1977, **78**, 393–406.
- 168 D. G. Reid, D. M. Doddrell, D. H. Williams and K. R. Fox, *Biochim. Biophys. Acta*, 1984, **798**, 111–114.
- 169 M. Sato, T. Nakazawa, Y. Tsunematsu, K. Hotta and K. Watanabe, *Curr. Opin. Chem. Biol.*, 2013, **17**, 537–545.
- 170 H. Chen, B. K. Hubbard, S. E. O'Connor and C. T. Walsh, *Chem. Biol.*, 2002, **9**, 103–112.
- 171 O. Kurnasov, L. Jablonski, B. Polanuyer, P. Dorrestein, T. Begley and A. Osterman, *FEMS Microbiol. Lett.*, 2003, **227**, 219–227.
- 172 C. Zhang, L. Kong, Q. Liu, X. Lei, T. Zhu, J. Yin, B. Lin, Z. Deng and D. You, *PLoS One*, , DOI:10.1371/journal.pone.0056772.
- 173 K. Koketsu, H. Oguri, K. Watanabe and H. Oikawa, *Org. Lett.*, 2006, **8**, 4719–4722.

- 174 Y. Hirose, K. Watanabe, A. Minami, T. Nakamura, H. Oguri and H. Oikawa, *J. Antibiot. (Tokyo)*, 2011, **64**, 117–122.
- 175 G. Schmooch, F. Pfennig, J. Jewiarz, W. Schlumbohm, W. Laubinger, F. Schauwecker and U. Keller, *J. Biol. Chem.*, 2005, **280**, 4339–4349.
- 176 A. Cornish, M. J. Waring and R. D. Nolan, *J. Antibiot. (Tokyo)*, 1983, **36**, 1664–1670.
- 177 G. A. Patani and E. J. LaVoie, *Chem. Rev.*, 1996, **96**, 3147–3176.
- 178 A. Reiner, D. Wildemann, G. Fischer and T. Kiefhaber, *J. Am. Chem. Soc.*, 2008, **130**, 8079–8084.
- 179 H. Verma, B. Khatri, S. Chakraborti and J. Chatterjee, *Chem. Sci.*, , DOI:10.1039/C7SC04671E.
- 180 N. Mahanta, D. M. Szantai-Kis, E. J. Petersson and D. A. Mitchell, *ACS Chem. Biol.*, 2019, **14**, 142–163.
- 181 A. Okano, R. C. James, J. G. Pierce, J. Xie and D. L. Boger, *J. Am. Chem. Soc.*, 2012, **134**, 8790–8793.
- 182 T. Ozturk, E. Ertas and O. Mert, *Chem. Rev.*, 2007, **107**, 5210–5278.
- 183 C. J. Van Leeuwen, J. L. Maas-Diepeveen, G. Niebeek, W. H. A. Vergouw, P. S. Griffioen and M. W. Luijken, *Aquat. Toxicol.*, 1985, **7**, 145–164.
- 184 A. Pini, L. Lozzi, A. Bernini, J. Brunetti, C. Falciani, S. Scali, S. Bindi, T. Di Maggio, G. M. Rossolini, N. Niccolai and L. Bracci, *Amino Acids*, 2012, **43**, 467–473.
- 185 Y. Liu, C. Zhao, D. E. Bergbreiter and D. Romo, *J. Org. Chem.*, 1998, **63**, 3471–3473.
- 186 K. M. J. De Mattos-Shiple, C. Greco, D. M. Heard, G. Hough, N. P. Mulholland, J. L. Vincent, J. Micklefield, T. J. Simpson, C. L. Willis, R. J. Cox and A. M. Bailey, *Chem. Sci.*, 2018, **9**, 4109–4117.
- 187 F. Schauwecker, F. Pfennig, N. Grammel and U. Keller, *Chem. Biol.*, 2000, **7**, 287–297.

- 188 E. Valeur and M. Bradley, *Chem. Soc. Rev.*, 2009, **38**, 606–31.
- 189 P. Palladino and D. A. Stetsenko, *Org. Lett.*, 2012, **14**, 6346–6349.
- 190 J. Thirlway, R. Lewis, L. Nunns, M. Al Nakeeb, M. Styles, A. W. Struck, C. P. Smith and J. Micklefield, *Angew. Chemie - Int. Ed.*, 2012, **51**, 7181–7184.
- 191 B. K. Leskiw, E. J. Lawlor, J. M. Fernandez-Abalos and K. F. Chater, *Proc. Natl. Acad. Sci.*, 1991, **88**, 2461–2465.
- 192 K. Watanabe, K. Hotta, A. P. Praseuth, K. Koketsu, A. Migita, C. N. Boddy, C. C. C. Wang, H. Oguri and H. Oikawa, *Nat. Chem. Biol.*, 2006, **2**, 423–428.
- 193 H. Zeng, S. Wen, W. Xu, Z. He, G. Zhai, Y. Liu, Z. Deng and Y. Sun, *Appl. Microbiol. Biotechnol.*, 2015, **99**, 10575–10585.
- 194 F. Alberti, D. J. Leng, I. Wilkening, L. Song, M. Tosin and C. Corre, *Chem. Sci.*, 2019, **10**, 453–463.
- 195 J. D. Sidda, L. Song, V. Poon, M. Al-Bassam, O. Lazos, M. J. Buttner, G. L. Challis and C. Corre, *Chem. Sci.*, 2014, **5**, 86–89.
- 196 C. Cheng and S. Shuman, *Mol. Cell. Biol.*, 2000, **20**, 8059–8068.
- 197 S. Shuman, *J. Biol. Chem.*, 1994, **269**, 32678–32684.
- 198 B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science (80-.)*, 2001, **291**, 1790–1792.
- 199 B. O. Bachmann and J. Ravel, *Methods Enzymol.*, 2009, **458**, 181–217.
- 200 L. A. Kelley, S. Mezulis, C. M. Yates and M. J. E. Sternberg, *Nat. Protoc.*, 2016, **10**, 845–858.
- 201 A. A. Baykov, A. Evtushenko and S. M. Avaeva, *Anal. Biochem.*, 1988, **171**, 266–270.
- 202 J.-P. Nougayrède, S. Homburg, F. Taieb, M. Boury, E. Bruszkiewicz, G. Gottschalk, C. Buchreiser, J. Jacker, U. Dobrindt and E. Oswald, *Science (80-.)*, 2006, **313**, 848–851.
- 203 G. Cuevas-Ramos, C. R. Petit, I. Marcq, M. Boury, E. Oswald and J.-P.

- Nougayrede, *Proc. Natl. Acad. Sci.*, 2010, **107**, 11537–11542.
- 204 D. Reimer, K. M. Pos, M. Thines, P. Grun and H. B. Bode, *Nat. Chem. Biol.*, 2011, **7**, 888–890.
- 205 D. Dubois, O. Baron, A. Cougnoux, J. Delmas, N. Pradel, M. Boury, B. Bouchon, M. A. Bringer, J. P. Nougayrède, E. Oswald and R. Bonnet, *J. Biol. Chem.*, 2011, **286**, 35562–35570.
- 206 C. A. Brotherton and E. P. Balskus, *J. Am. Chem. Soc.*, 2013, **135**, 3359–3362.
- 207 X. Bian, J. Fu, A. Plaza, J. Herrmann, D. Pistorius, A. F. Stewart, Y. Zhang and R. Müller, *ChemBioChem*, 2013, **14**, 1194–1197.
- 208 M. I. Vizcaino, P. Engel, E. Trautman and J. M. Crawford, *J. Am. Chem. Soc.*, 2014, **136**, 9244–9247.
- 209 C. A. Brotherton, M. Wilson, G. Byrd and E. P. Balskus, *Org. Lett.*, 2015, **17**, 1545–1548.
- 210 X. Bian, A. Plaza, Y. Zhang and R. Müller, *Chem. Sci.*, 2015, **6**, 3154–3160.
- 211 L. A. Wessjohann, W. Brandt and T. Thiemann, *Chem. Rev.*, 2003, **103**, 1625–1647.
- 212 A. R. Healy and S. B. Herzon, *J. Am. Chem. Soc.*, 2017, **139**, 14817–14824.
- 213 P. Martin, I. Marcq, G. Magistro, M. Penary, C. Garcie, D. Payros, M. Boury, M. Olier, J. P. Nougayrède, M. Audebert, C. Chalut, S. Schubert and E. Oswald, *PLoS Pathog.*, , DOI:10.1371/journal.ppat.1003437.
- 214 Y. A. Chan, M. T. Boyne, A. M. Podevels, A. K. Klimowicz, J. Handelsman, N. L. Kelleher and M. G. Thomas, *Proc. Natl. Acad. Sci.*, 2006, **103**, 14349–14354.
- 215 L. Zha, M. R. Wilson, C. A. Brotherton and E. P. Balskus, *ACS Chem. Biol.*, 2016, **11**, 1287–1295.
- 216 E. P. Trautman, A. R. Healy, E. E. Shine, S. B. Herzon and J. M. Crawford, *J. Am. Chem. Soc.*, 2017, **139**, 4195–4201.

- 217 Z. R. Li, J. Li, J. P. Gu, J. Y. H. Lai, B. M. Duggan, W. P. Zhang, Z. L. Li, Y. X. Li, R. B. Tong, Y. Xu, D. H. Lin, B. S. Moore and P. Y. Qian, *Nat. Chem. Biol.*, 2016, **12**, 773–775.
- 218 N. S. Guntaka, A. R. Healy, J. M. Crawford, S. B. Herzon and S. D. Bruner, *ACS Chem. Biol.*, 2017, **12**, 2598–2608.
- 219 N. Kondo, A. Takahashi, K. Ono and T. Ohnishi, *J. Nucleic Acids*, , DOI:10.4061/2010/543531.
- 220 S. Balbo, S. S. Hecht, P. Upadhyaya and P. W. Villalta, *Anal. Chem.*, 2014, **86**, 1744–1752.
- 221 M. R. Wilson, Y. Jiang, P. W. Villalta, A. Stornetta, P. D. Boudreau, A. Carrá, C. A. Brennan, E. Chun, L. Ngo, L. D. Samson, B. P. Engelward, W. S. Garrett, S. Balbo and E. P. Balskus, *Science (80-.)*, , DOI:10.1126/science.aar7785.
- 222 M. Xue, E. Shine, W. Wang, J. M. Crawford and S. B. Herzon, *Biochemistry*, 2018, **57**, 6391–6394.
- 223 D. W. Brooks, L. D.-L. Lu and S. Masamune, *Angew. Chemie Int. Ed. English*, 1979, **18**, 72–74.
- 224 V. Kubyshkin and N. Budisa, *Beilstein J. Org. Chem.*, 2017, **13**, 2442–2457.
- 225 F. Buron, A. Turck, N. Ple, L. Bischoff and F. Marsais, *Tetrahedron Lett.*, 2007, **48**, 4327–4330.
- 226 W. G. Laver, A. Neuberger and J. J. Scott, *J. Chem. Soc.*, 1959, 1483–1491.
- 227 J. Havemann, M. E. Yurkovich, R. Jenkins, S. Harringer, W. Tao, S. Wen, Y. Sun, P. F. Leadlay and M. Tosin, *Chem. Commun.*, 2017, **53**, 1912–1915.
- 228 N. Bossuet-Greif, D. Dubois, C. Petit, S. Tronnet, P. Martin, R. Bonnet, E. Oswald and J. P. Nougayrède, *Mol. Microbiol.*, 2016, **99**, 897–908.
- 229 P. Tripathi, E. E. Shine, A. R. Healy, C. S. Kim, S. B. Herzon, S. D. Bruner and J. M. Crawford, *J. Am. Chem. Soc.*, 2017, **139**, 17719–17722.
- 230 Y. Jiang, A. Stornetta, P. W. Villalta, M. R. Wilson, P. D. Boudreau, L. Zha,

- S. Balbo and E. P. Balskus, .
- 231 A. N. Chatterjee and H. R. Perkins, *Biochem. Biophys. Res. Commun.*, 1966, **24**, 489–494.
- 232 M. Nieto and H. R. Perkins, *Biochem. J.*, 1971, **123**, 789–803.
- 233 F. J. Marshall, *J. Med. Chem.*, 1965, **8**, 18–22.
- 234 R. M. Smith, A. W. Johnson and R. D. Guthrie, *J. Chem. Soc. Chem. Commun.*, 1972, 361–362.
- 235 A. W. Johnson, R. M. Smith and R. D. Guthrie, *J. Chem. Soc. Perkin Trans. 1*, 1972, 2153–2159.
- 236 G. A. Smith, K. A. Smith and D. H. Williams, *J. Chem. Soc. Perkin Trans. 1*, 1975, 2108–2115.
- 237 D. H. Williams and J. R. Kalman, *J. Am. Chem. Soc.*, 1977, **99**, 2768–2774.
- 238 G. M. Sheldrick, P. G. Jones, O. Kennard, D. H. Williams and G. A. Smith, *Nature*, 1978, **271**, 223–225.
- 239 R. W. Fairbrother and B. L. Williams, *Lancet*, 1956, **268**, 1177–1179.
- 240 H. I. C. P. A. Committee, *Infect. Control Hosp. Epidemiol.*, 1995, **16**, 105–113.
- 241 H. J. Rogers, H. R. Perkins and J. B. Ward, *Microbial Cell Walls and Membranes*, 1980.
- 242 P. E. Reynolds, *Eur. J. Clin. Microbiol. Infect. Dis.*, 1989, **8**, 943–950.
- 243 R. Nagarajan, *Antimicrob. Agents Chemother.*, 1991, **35**, 605–609.
- 244 R. Leclercq, E. Derlot, J. Duval and P. Courvalin, *N. Engl. J. Med.*, 1988, **319**, 157–161.
- 245 S. J. McKessar, A. M. Berry, J. M. Bell, J. D. Turnidge and J. C. Paton, *Antimicrob. Agents Chemother.*, 2000, **44**, 3224–3228.
- 246 S. Evers and P. Courvalin, *J. Bacteriol.*, 1996, **178**, 1302–1309.
- 247 H. S. Gold and R. C. Moellering, *N. Engl. J. Med.*, 1996, **335**, 1445–1453.

- 248 R. Nagarajan, *J. Antibiot. (Tokyo)*.
- 249 H. C. Losey, M. W. Peczuh, Z. Chen, U. S. Eggert, S. D. Dong, I. Pelczer, D. Kahne and C. T. Walsh, *Biochemistry*, 2001, **40**, 4745–4755.
- 250 H. C. Losey, J. Jiang, J. B. Biggins, M. Oberthür, X. Y. Ye, S. D. Dong, D. Kahne, J. S. Thorson and C. T. Walsh, *Chem. Biol.*, 2002, **9**, 1305–1314.
- 251 X. Fu, C. Albermann, C. Zhang and J. S. Thorson, *Org. Lett.*, 2005, **7**, 1513–1515.
- 252 R. G. Kruger, W. Lu, M. Oberthür, J. Tao, D. Kahne and C. T. Walsh, *Chem. Biol.*, 2005, **12**, 131–140.
- 253 S. Weist, B. Bister, O. Puk, D. Bischoff, S. Pelzer, G. J. Nicholson, W. Wohlleben, G. Jung and D. S. Roderich, *Angew. Chemie - Int. Ed.*, 2002, **41**, 3383–3385.
- 254 S. Weist, C. Kittel, D. Bischoff, B. Bister, V. Pfeifer, G. J. Nicholson, W. Wohlleben and R. D. Süßmuth, *J. Am. Chem. Soc.*, 2004, **126**, 5942–5943.
- 255 D. L. Higgins, R. Chang, D. V. Debabov, J. Leung, T. Wu, K. M. Krause, E. Sandvik, J. M. Hubbard, K. Kaniga, D. E. Schmidt, Q. Gao, R. T. Cass, D. E. Karr, B. M. Benton and P. P. Humphrey, *Antimicrob. Agents Chemother.*, 2005, **49**, 1127–1134.
- 256 N. E. Allen and T. I. Nicas, *FEMS Microbiol. Rev.*, 2003, **26**, 511–532.
- 257 Y. Nakama, O. Yoshida, M. Yoda, K. Araki, Y. Sawada, J. Nakamura, S. Xu, K. Miura, H. Maki and H. Arimoto, *J. Med. Chem.*, 2010, **53**, 2528–2533.
- 258 J. R. Pinchman and D. L. Boger, *J. Med. Chem.*, 2013, **56**, 4116–4124.
- 259 T. P. Pathak and S. J. Miller, *J. Am. Chem. Soc.*, 2012, **134**, 6120–6123.
- 260 C. M. Crane and D. L. Boger, *J. Med. Chem.*, 2009, **52**, 1471–1476.
- 261 A. Malabarba, T. I. Nicas and R. C. Thompson, *Med. Res. Rev.*, 1997, **17**, 69–137.
- 262 V. Yarlagadda, P. Akkapeddi, G. B. Manjunath and J. Haldar, *J. Med. Chem.*,

- 2014, **57**, 4558–4568.
- 263 B. W. Trotter, J. C. Barrow, T. I. Richardson, J. L. Katz, M. R. Wood and D. A. Evans, *Angew. Chemie Int. Ed.*, 1998, **37**, 2700–2704.
- 264 K. C. Nicolaou, M. Takayanagi, N. F. Jain, S. Natarajan, A. E. Koumbis, T. Bando and J. M. Ramanjulu, *Angew. Chemie - Int. Ed.*, 1998, **37**, 2717–2719.
- 265 D. L. Boger, S. Miyazaki, S. H. Kim, J. H. Wu, S. L. Castle, O. Loiseleur and Q. Jin, *J. Am. Chem. Soc.*, 1999, **121**, 10004–10011.
- 266 C. C. McComas, B. M. Crowley and D. L. Boger, *J. Am. Chem. Soc.*, 2003, **125**, 9314–9315.
- 267 B. M. Crowley and D. L. Boger, *J. Am. Chem. Soc.*, 2006, **128**, 2885–2892.
- 268 J. Xie, J. G. Pierce, R. C. James, A. Okano and D. L. Boger, *J. Am. Chem. Soc.*, 2011, **133**, 13946–13949.
- 269 S. J. Hammond, M. P. Williamson, D. H. Williams, L. D. Boeck and G. G. Marconi, *J. Chem. Soc. Chem. Commun.*, 1983, 1174–1175.
- 270 A. M. van Wageningen, P. N. Kirkpatrick, D. H. Williams, B. R. Harris, J. K. Kershaw, N. J. Lennard, M. Jones, S. Jm and P. J. Solenberg, *Chem. Biol.*, 1998, **5**, 155–62.
- 271 S. Pelzer, R. D. Sussmuth, D. Heckmann, J. Recktenwald, P. Huber, G. Jung and W. Wohlleben, *Antimicrob. Agents Chemother.*, 1999, **43**, 1565–1573.
- 272 V. Pfeifer, G. J. Nicholson, J. Ries, J. Recktenwald, A. B. Schefer, R. M. Shawky, J. Schröder, W. Wohlleben and S. Pelzer, *J. Biol. Chem.*, 2001, **276**, 38370–38377.
- 273 H. Chen, C. C. Tseng, B. K. Hubbard and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 14901–14906.
- 274 T. L. Li, O. W. Choroba, H. Hong, D. H. Williams and J. B. Spencer, *Chem. Commun.*, 2001, 2156–2157.
- 275 A. M. Sandercock, E. H. Charles, W. Scaife, P. N. Kirkpatrick, S. W. O'Brien, E. A. Papageorgiou, J. B. Spencer and D. H. Williams, *Chem.*

- Commun.*, 2001, 1252–1253.
- 276 B. K. Hubbard, M. G. Thomas and C. T. Walsh, *Chem. Biol.*, 2000, **7**, 931–942.
- 277 O. W. Choroba, D. H. Williams and J. B. Spencer, *J. Am. Chem. Soc.*, 2000, **122**, 5389–5390.
- 278 H. Chen and C. T. Walsh, *Chem. Biol.*, 2001, **8**, 301–312.
- 279 M. J. Cryle, A. Meinhart and I. Schlichting, *J. Biol. Chem.*, 2010, **285**, 24562–24574.
- 280 S. Mulyani, E. Egel, C. Kittel, S. Turkanovic, W. Wohlleben, R. D. Süßmuth and K. H. Van Pée, *ChemBioChem*, 2010, **11**, 266–271.
- 281 J. W. Trauger and C. T. Walsh, *Proc. Natl. Acad. Sci.*, 2012, **97**, 3112–3117.
- 282 L. Xu, H. Huang, W. Wei, Y. Zhong, B. Tang, H. Yuan, L. Zhu, W. Huang, M. Ge, S. Yang, H. Zheng, W. Jiang, D. Chen, G. P. Zhao and W. Zhao, *BMC Genomics*, 2014, **15**, 363.
- 283 F. Cava, H. Lam, M. A. De Pedro and M. K. Waldor, *Cell. Mol. Life Sci.*, 2011, **68**, 817–831.
- 284 P. J. Solenberg, P. Matsushima, D. R. Stack, S. C. Wilkie, R. C. Thompson and R. H. Baltz, *Chem. Biol.*, 1997, **4**, 195–202.
- 285 K. Woihte, N. Geib, K. Zerbe, B. L. Dong, M. Heck, S. Fournier-Rousset, O. Meyer, F. Vitali, N. Matoba, K. Abou-Hadeed and J. A. Robinson, *J. Am. Chem. Soc.*, 2007, **129**, 6887–6895.
- 286 M. Peschke, C. Brieke, R. J. A. Goode, R. B. Schittenhelm and M. J. Cryle, *Biochemistry*, 2017, **56**, 1239–1247.
- 287 D. Bischoff, B. Bister, M. Bertazzo, V. Pfeifer, E. Stegmann, G. J. Nicholson, S. Keller, S. Pelzer, W. Wohlleben and R. D. Süßmuth, *ChemBioChem*, 2005, **6**, 267–272.
- 288 O. Puk, P. Huber, D. Bischoff, J. Recktenwald, G. Jung, R. D. Süßmuth, K. H. Van Pée, W. Wohlleben and S. Pelzer, *Chem. Biol.*, 2002, **9**, 225–235.

- 289 O. Puk, D. Bischoff, C. Kittel, S. Pelzer, S. Weist, E. Stegmann, R. D. Süßmuth and W. Wohlleben, *J. Bacteriol.*, 2004, **186**, 6093–6100.
- 290 P. C. Schmartz, K. Zerbe, K. Abou-Hadeed and J. A. Robinson, *Org. Biomol. Chem.*, 2014, **12**, 5574–5577.
- 291 T. Kittilä, C. Kittel, J. Tailhades, D. Butz, M. Schoppet, A. Büttner, R. J. A. Goode, R. B. Schittenhelm, K.-H. van Pee, R. D. Süßmuth, W. Wohlleben, M. J. Cryle and E. Stegmann, *Chem. Sci.*, 2017, **8**, 5992–6004.
- 292 G. G. Smith and T. Sivakua, *J. Org. Chem.*, 1983, **48**, 627–634.
- 293 F. Dettner, A. Hänchen, D. Schols, L. Toti, A. Nußer and R. D. Süßmuth, *Angew. Chemie - Int. Ed.*, 2009, **48**, 1856–1861.
- 294 D. Crich and A. Banerjee, *J. Org. Chem.*, 2006, **71**, 7106–7109.
- 295 C. Djerassi, *Chem. Rev.*, 1948, **43**, 271–317.
- 296 J. M. Daley and R. G. Landolt, *J. Chem. Educ.*, 2005, **82**, 120–121.
- 297 K. Shimamoto and Y. Ohfuné, *Tetrahedron Lett.*, 1988, **29**, 5177–5180.
- 298 H. C. Brown, E. J. Mead and B. C. S. Rao, *J. Am. Chem. Soc.*, 1955, **77**, 6209–6213.
- 299 A. M. Giltrap, F. P. J. Haeckl, K. L. Kurita, R. G. Linington and R. J. Payne, *J. Org. Chem.*, 2018, **83**, 7250–7270.
- 300 J. Béahdy, *Adv. Appl. Microbiol.*, 1974, **18**, 309–406.
- 301 J. Handelsman, M. R. Rondon, S. F. Brady, J. Clardy and R. M. Goodman, *Chem. Biol.*, , DOI:10.1016/S1074-5521(98)90108-9.
- 302 S. D. Bentley, K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O’Neil, E. Rabinowitsch, M.-A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B.

- G. Barrell, J. Parkhill and D. A. Hopwood, *Nature*, 2002, **417**, 141–7.
- 303 R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2017, **44**, 573–588.
- 304 L. Cuthbertson and J. R. Nodwell, *Microbiol. Mol. Biol. Rev.*, 2013, **77**, 440–475.
- 305 C. Kisker, W. Hinrichs, K. Tovar, W. Hillen and W. Saenger, *J. Mol. Biol.*, 1995, **247**, 260–280.
- 306 J. L. Ramos, M. Martínez-Bueno, A. J. Molina-Henares, W. Terán, K. Watanabe, M. T. Gallegos, R. Brennan and R. Tobes, *Microbiol. Mol. Biol. Rev.*, 2005, **69**, 326–356.
- 307 S. O. Rourke, A. Wietzorrek, K. Fowler, C. Corre, G. L. Challis and K. F. Chater, *Mol. Microbiol.*, 2009, **71**, 763–778.
- 308 M. H. Medema, E. Takano and R. Breitling, *Mol. Biol. Evol.*, 2013, **30**, 1218–1223.
- 309 E. L. de los Santos and G. L. Challis, *bioRxiv*, 2017, 4–8.
- 310 P. Marfey, *Carlsberg Res. Commun.*, 1984, **49**, 591–596.
- 311 C. Peng, J.-Y. Pu, L.-Q. Song, X.-H. Jian, M.-C. Tang and G.-L. Tang, *Proc. Natl. Acad. Sci.*, 2012, **109**, 8540–8545.
- 312 J. Li, Z. Xie, M. Wang, G. Ai and Y. Chen, *PLoS One*, 2015, **10**, 1–20.
- 313 A. Surrey, R. W. Attwell and T. Cross, *J. Gen. Appl. Microbiol.*, 1985, **31**, 493–497.
- 314 O. A. Ulanovskaya, A. M. Zuhl and B. F. Cravatt, *Nat. Chem. Biol.*, 2013, **9**, 300–306.
- 315 J. S. Cavanaugh, R. Jou, M. H. Wu, T. Dalton, E. Kurbatova, J. Ershova, J. P. Cegielski, J. Lancaster, R. Odendaal, L. Diem, K. Tan, A. T. Walker, E. Sigman, B. Metchock, M. T. C. Perez, M. T. Gler, C. Bonilla, O. Jave, I. Norvaisha, G. Skenders, I. Sture, V. Riekstina, A. Cirule, S. N. Cho, S. Eum, J. Lee, Y. Cai, I. C. Shamputa, T. Kuznetsova, R. Akksilp, W. Sitti, J. Inyapong, E. V. Kiryanova, I. Degtyareva, E. S. Nemtsova, K. Levina, M.

- Danilovits, T. Kummik, Y. C. Lei, W. L. Huang, V. V. Erokhin, L. N. Chernousova, S. N. Andreevskaya, E. E. Larionova and T. G. Smirnova, *J. Antimicrob. Chemother.*, 2017, **72**, 1678–1687.
- 316 C. Greunke, E. R. Duell, P. M. D’Agostino, A. Glöckle, K. Lamm and T. A. M. Gulder, *Metab. Eng.*, 2018, **47**, 334–345.
- 317 M. N. Thaker, W. Wang, P. Spanogiannopoulos, N. Waglechner, A. M. King, R. Medina and G. D. Wright, *Nat. Biotechnol.*, 2013, **31**, 922–927.
- 318 N. Waglechner, A. G. McArthur and G. D. Wright, *Nat. Microbiol.*, , DOI:10.1038/s41564-019-0531-5.
- 319 G. Byk and C. Gilon, *J. Org. Chem.*, 1992, **57**, 5687–5692.
- 320 I.-H. Kim, C. Morisseau, T. Watanabe and B. D. Hammock, *J. Med. Chem.*, 2004, **47**, 2110–2122.
- 321 C. Winter, R. D. C. Pullin and T. J. Donohoe, *Tetrahedron Lett.*, 2017, **58**, 602–605.
- 322 A. V. Malkov, K. Vranková, M. Černý and P. Kočovský, *J. Org. Chem.*, 2009, **74**, 8425–8427.
- 323 B. P. Bandgar, R. J. Sarangdhar, S. Viswakarma and F. A. Ahamed, *J. Med. Chem.*, 2011, **54**, 1191–1201.
- 324 T. Sawabet, K. Aihara, K. Atsumi and K. Ajito, 2005, **5**.
- 325 L. P. Tardibono, J. Patzner, C. Cesarlo and M. J. Miller, *Org. Lett.*, 2009, **11**, 4076–4079.
- 326 K. Matsuo and M. Shindo, *Org. Lett.*, 2011, **13**, 4406–4409.
- 327 A. J. Ross, H. L. Lang and R. F. W. Jackson, *J. Org. Chem.*, 2010, **75**, 245–248.
- 328 R. Gealageas, A. Devineau, P. P. L. So, C. M. J. Kim, J. Surendradoss, C. Buchwalder, M. Heller, V. Goebeler, E. M. Dullaghan, D. S. Grierson and E. E. Putnins, *J. Med. Chem.*, 2018, **61**, 7043–7064.

- 329 S. Turkyilmaz and C. S. Wilcox, *Tetrahedron Lett.*, 2017, **58**, 2031–2033.
- 330 K. Kawamine, R. Takeuchi, M. Miyashita, H. Irie, K. Shimamoto and Ohfuné, *Chem. Pharm. Bull.*, 1994, **17**, 1460–1462.
- 331 M. Gynther, J. Ropponen, K. Laine, J. Leppänen, P. Haapakoski, L. Peura, T. Järvinen and J. Rautio, *J. Med. Chem.*, 2009, **52**, 3348–3353.
- 332 B. J. W. Barratt, C. J. Easton, D. J. Henry, I. H. W. Li, L. Radom and J. S. Simpson, *J. Am. Chem. Soc.*, 2004, **126**, 13306–13311.
- 333 S. J. Danishefsky, W. H. Pearson and B. E. Segmullert, *J. Am. Chem. Soc.*, 1985, **5**, 1280–1285.
- 334 Z. Li, I. O. Lebedyeva, V. M. Golubovskaya, W. G. Cance, K. A. Alamry, H. M. Faidallah, C. Dennis Hall and A. R. Katritzky, *Bioorg. Med. Chem.*, 2015, **23**, 5056–5060.
- 335 J. F. Soulé, H. Miyamura and S. Kobayashi, *J. Am. Chem. Soc.*, 2011, **133**, 18550–18553.
- 336 L. Biancalana, M. Bortoluzzi, E. Ferretti, M. Hayatifar, F. Marchetti, G. Pampaloni and S. Zacchini, *RSC Adv.*, 2017, **7**, 10158–10174.
- 337 S. M. Voshell, S. J. Lee and M. R. Gagné, *J. Am. Chem. Soc.*, 2006, **128**, 12422–12423.
- 338 M. K. Boukanoun, X. Hou, L. Nikolajev, S. Ratni, D. Olson, A. Claing, S. A. Laporte, S. Chemtob and W. D. Lubell, *Org. Biomol. Chem.*, 2015, **13**, 7750–7761.
- 339 J. Bonnamour, T. X. Métro, J. Martinez and F. Lamaty, *Green Chem.*, 2013, **15**, 1116–1120.
- 340 S. Mori, S. K. Shrestha, J. Fernández, M. Álvarez San Millán, A. Garzan, A. H. Al-Mestarihi, F. Lombó and S. Garneau-Tsodikova, *Biochemistry*, 2017, **56**, 4457–4467.

