Enzymology of nitrogen incorporation in actinobacterial polyketide alkaloid biosynthesis

By

Ufedo Ruby Awodi

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

Doctor of Philosophy in Chemistry

University of Warwick

Department of Chemistry

October 2021

Contents

LIST OF FIGURES

Figure 13 a and b: SDS-PAGE Analysis of **a)** CpkC-ACP-TR (***Lanes L:*** proteins of known molecular weight, ***Lane 1:*** CpkC-ACP-TR) and **b)** (CpkC-TR. (***TP:*** total protein, ***I:*** insoluble, ***S:*** soluble, ***E:*** eluate, E_C***:*** concentrated eluate, E_{C1}***:*** 1µl of concentrated eluate.

Figure 14: Mass spectrum of purified CpkC-ACP-TR didomain showing molecular mass of *ca.* 55 kDa.

Figure 15a: ACP-TR flourimetric assay of enzymatic activity with substrate mimic octanoyl-CoA (**48**) and NADH co-factor.

Figure 15b: ACP-TR flourimetric assay of enzymatic activity with other acyl-CoA substrates: butanoyl-CoA (**47**) and dodecanoyl-CoA (**49**) showing substrate tolerance.

Figure 16: ACP-TR flourimetric assay of enzymatic activity with substrate mimic octanoyl-CoA (**48**) and NADPH co-factor.

Figure 17: Flourimetric assay of TR enzymatic activity showing a decline in the intensity of NADH fluorescence at 462 nm in the presence of an aldehyde: octanal (**a**) and a ketone: 2-octanone (**b**).

Figure 18: Structure of NADH showing the diastereotopic hydrogens.

Figure 19: ^{1}H NMR spectra of NADH and 4-[^{2}H]NADH. Green trace [*4R*-^{2}H]NADH from alcohol dehydrogenase catalysed reaction. Red trace [*4S*-^{2}H]NADH from glucose dehydrogenase catalysed reaction. Blue trace: unlabelled NADH.

Figure 20a and b: HPLC-MS analysis of TR enzymatic reaction for detection of octanoyl-CoA (**48**) consumption **a)** Extracted ion chromatograms at *m/z* = 892.2 top: control reaction without enzyme and bottom: TR reaction containing the active enzyme. **b)** Mass spectrum for $[M-H]^-$ ion of octanoyl-CoA.

Figure 21: HPLC-MS analysis of TR reaction for detection of CoASH (**54**) formation **a)** Extracted ion chromatograms at *m/z* = 766.00 top: control reaction without added enzyme; bottom: TR reaction with active enzyme. **b)** Mass spectrum for $[M-H]^-$ ion of CoASH.

Figure 22a: GC-MS detection of trimethylsilyl-derivatized dodecanal. Top: authentic standard solution of derivatized 1-dodecanol (**57**) middle: ACP-TR with dodecanoyl-CoA bottom: control.

Figure 22b: Extracted ion chromatograms at *m/z* = 187.2 (corresponding to the fragment of the BSA derived octanol (**56**) after neutral loss of a methyl group) of GC-MS analysis of CpkC-TR catalysed reaction of octanoyl-CoA (**48**) with NADH (**18**). Top: authentic standard solution of derivatized 1-octanol; middle: CpkC-TR enzymatic reaction; bottom: control reaction.

Figure 23: Active site model of a typical ω–aminotransferase (GABA-AT from *E. coli*) showing interactions of conserved residues with the **a**) PLP co-factor bound to lysine (internal aldimine) and **b**) PLP co-factor bound to a substrate (external aldimine).

Figure 24: Sequence alignment of CpkG in with other ω–transferases of known (and putative) structure and function showing the strictly conserved active site lysine (K338) and aspartate (D309) residues. The glutamine (Q312) residue interacts with the phenolic oxygen of the co-factor.

Figure 25: In CpkG a conserved glycine (G121) residue and an adjacent alanine (A122) residue are predicted to interact with the phosphate oxygens via the amide nitrogen atoms in the polypeptide chain. A threonine (T368) residue located further downstream is predicted to interact with a phosphate oxygen via hydrogen bonding.

Figure 26: Sequence alignment showing residues predicted to be involved in substrate binding: in CpkG an arginine (R480) would be expected to interact with the α-carboxylate group of the amino acid substrate; a glutamate (E281) residue present only in α–ketoglutarate-specific ω–aminotransferases which forms a salt-bridge with an arginine residue as part of a gateway system; and a basic residue (K204 in CpkG) which is predicted to control substrate specificity towards a preference for linear amines (such as putrescine) that do not contain an α-carboxylate group.

Figure 27: Gel electrophoresis analysis of the *cpk*G PCR product. **Lane 1**: failed PCR product (non-specific primer binding); **Lanes 2** and **3**: PCR products of targeted size.

Figure 28: Restriction sites for *Pvu*II enzyme on the pET151-cpkG clone.

Figure 29: Gel electrophoresis analysis of *Pvu*II restriction digest of pET151-*cpk*G.

Lanes **1** to **4** contain plasmids of single *E. coli* colonies that were selected for further analysis. Positive clones **2** and **4** were confirmed by sequencing and used to transform *E. coli* BL21STR for protein expression.

Figure 30: Analysis of N-terminal His_6-CpkG recombinant protein. **Insert:** SDS-PAGE analysis (***Lane 1:*** His_6-CpkG, ***Lane L:*** proteins of known molecular weight) ; UHPLC-ESI-Q-TOF-MS analysis.

Figure 31: Size exclusion chromatogram of CpkG.

Figure 32: Size exclusion chromatography calibration curve. Standards used: 38 ml Blue dextran (2000 kDa), 49.4 ml β-amylase (200 kDa), 57.6 ml bovine serum albumin (66 kDa), 68.47 ml carbonic anhydrase (29 kDa), 79.74 ml cytochrome c (12.4 kDa).

Figure 33: UV-vis spectrum of CpkG as purified (blue trace) after addition of 0.2mM PLP (pink trace).

Figure 34: Extracted Ion Chromatogram at *m/z* = 162.1 from LC-MS analysis of CpkG corresponding to the product propanal (**69**) derivaised with *D*-cysteine (**79**). CpkG was incubated with 1-propylamine (**75**) as amino group donor and pyruvate (**67**) as amino group acceptor. Top: control reaction with no added enzyme and the bottom: enzyme assay. **b)** Mass spectrum of ethylthiazolidine-4-carboxylic acid (**80**).

Figure 35: Extracted Ion Chromatogram at *m/z* = 232.1 corresponding to $C_{11}H_{22}NO_2S^+$ from LC-MS analysis of CpkG incubated with 1-octylamine (**73**) and pyruvate (**67**) showing *D*-cysteine derivatized octanal (**51**). Standard (top) control reaction with no enzyme (middle) and enzyme assay (bottom).

Figure 36: Extracted Ion Chromatogram of *m/z* = 232.1 corresponding to $C_{11}H_{22}NO_2S^+$ from LC-MS analysis of CpkG incubated with 1-octylamine (**73**) and α-KGA (**68**) showing *D*-cysteine derivatized octanal (**51**). Standard (top) control reaction with no enzyme (middle) and enzyme assay (bottom).

Figure 37: Structure of Cu (II) complex with amino acids with UV λ_{max} of 595 nm.

Figure 38: UV absorbance at 595 nm measuring **a)** glutamate and **b)** alanine Cu^{2+} complex formation.

Figure 39: Fluorescent emission at 462 nm of NADH following incubation of CpkG with **a)** octanal and L-alanine **b)** octanal and D-alanine and **c)** 2-octanone and L-alanine.

Figure 40: Extracted Ion Chromatograph at *m/z* = 130 from the LC-MS analysis of His_6-CpkG catalysed reactions. Top: octylamine (**73**) standard, Middle: control reaction with no added enzyme, bottom: CpkG catalysed reaction.

Figure 41: Representative structures of some polyketide alkaloids of actinobacterial origin.

LIST OF SCHEMES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am extremely grateful to Prof Greg Challis for giving me this great opportunity to glean from his renowned ingenious scientific research skills through his supportive and ever-inspiring guidance throughout the course of the PhD programme; an opportunity I would have otherwise not had access to. He afforded me with the rare chance to work and have access to an excellent research environment in the University of Warwick Chemistry Department; alongside a truly remarkable research team which he built and nurtured with his open-door policy and constant support. I would like to thank all the members of the Challis research team in particular: Paulina, Lona, Lauren, David, Jade, Gideon, Sarah, Joleen, Orestis, Zdenek, among so many others; also Profs. Manuela Tosin, Lijiang Song and Christophe Corre for their valuable advice and guidance. A lot of other researchers across the department which I have not mentioned helped in various ways, I really appreciate each and every one of them. I am thankful for my friends and immediate family, especially my son Emmanuel for doing his very best to keep a smile on my face at all times. Funding from the University of Warwick Chancellor's International scholarship in support of this research is very gratefully acknowledged.

DECLARATION

The experimental work reported in this thesis is original research carried out by the author, unless otherwise stated, in the Department of Chemistry, University of Warwick, between September 2011 and November 2014. No material has been submitted for any other degree, or at any other institution.

Parts of this thesis have been published by the author:

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

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

_____ Date:_____

Ufedo R. Awodi

LIST OF ABBREVIATIONS

ACP	Acyl carrier protein
ADH	Alcohol dehydrogenase
AONS	8-Amino-7-oxononanoate synthase
AT	Acyltransferase (domain)
Bp	Base pairs
CoA	Coenzyme A
DH	Dehydratase
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EIC	Extracted ion chromatogram
ER	Enoyl reductase
ESI	Electrospray ionisation
GDH	Glucose dehydrogenase
FAS	Fatty acid synthase
FPLC	Fast protein liquid chromatography
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
IPTG	Isopropyl-β-D-thiogalactoside
KR	Ketoreductase
KS	Ketosynthase
LB	Luria-Bertani medium
LC-MS	Liquid chromatography-mass spectrometry
LDH	Lactate dehydrogenase

MS	Mass spectrometry
m/z	Mass to charge ratio
NAD+	β-Nicotinamide adenine dinucleotide (oxidized)
NADH	β-Nicotinamide adenine dinucleotide (reduced)
NADP+	β-Nicotinamide adenine dinucleotide 2′-phosphate (oxidized)
NADPH	β-Nicotinamide adenine dinucleotide 2′-phosphate (reduced)
NRPS	Non-ribosomal peptide synthetase
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PLP	Pyridoxal 5′-phosphate
PMP	Pyridoxamine 5′-phosphate
SDR	Short-chain dehydrogenase / reductase
SDS	Sodium dodecyl sulfate
TE	Thioesterase
TEMED	Tetramethylethylenediamine
TMS	Trimethyl silyl
TR	Thioester reductase
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet

Vis	Visible

ABSTRACT

Numerous alkaloids of polyketide origin are known to be produced by Actinobacteria. These polyketide-alkaloids exhibit unusual structural diversity with several examples possessing mono-, di-, tri-and tetracyclic rings. Coelimycin P1 is a polyketide alkaloid which is the product of the cryptic *cpk* polyketide biosynthetic gene cluster in *S. coelicolor* M145. Its structure consists of a unique functionalized 1,5-oxathiocane attached to a piperideine ring with an extended conjugated π system which is responsible for its yellow pigmented colour. Previous bioinformatics analysis showed that the type I modular PKS encoded by the *cpk* gene cluster contains a C-terminal thioester reductase (TR) domain proposed to catalyse the reductive release of the fully assembled polyketide chain, converting it to the corresponding aldehyde. A gene encoding a putative pyridoxal phosphate (PLP)-dependent transaminase (CpkG) was also identified within the *cpk* gene cluster. CpkG was predicted to convert the aldehyde product of the type I PKS to the corresponding amine prior to subsequent transformations leading to the assembly of coelimycin P1.

The TR domain and CpkG were overproduced as recombinant proteins in *E. coli* and purified to investigate their enzymatic activity. The TR domain was shown to catalyse the NADH-dependent reduction of structural analogues of the fully assembled polyketide chain (commercially available short to medium chain acyl-CoAs). It preferentially utilized NADH over NADPH. Recombinant CpkG was shown to catalyse the PLP-dependent transamination of short to medium chain aldehydes to their corresponding amines, utilizing both L-and D-alanine as well as L-glutamate.

Results obtained herein, show that the coelimycin P1 biosynthetic pathway (involving reduction of the thioester bond to the aldehyde followed by reductive amination to yield the amine) is distinct from previously described nitrogen incorporation pathways in alkaloid biosynthesis which typically involve either L-lysine or L-ornithine as precursors; with the entire carbon backbone of the nitrogen heterocycle being derived from the amino acid.

Bioinformatics analysis carried out during subsequent work, on sequenced actinobacterial genomes deposited in the GenBank database, querying for biosynthetic gene clusters harbouring genes encoding for a type I modular PKS with a TR domain as well as a CpkG homologue, identified 22 clusters likely to be responsible for the assembly of polyketide-alkaloids. This suggests that NAD(P)-dependent reductive chain release and PLP-dependent transamination is a conserved feature of polyketide-alkaloid biosynthesis.

CHAPTER ONE

1.0 INTRODUCTION

1.1 Biologically Active Natural Products

Natural products extracted from plants, animals, bacteria and fungi have been found to be of immense economic and therapeutic importance since ancient times.[1] Originally used in their crude form, these secondary metabolites found applications as poisons, euphoriants and for the treatment of disease; they have also been used as pigments and pesticides.[2,3,4] They are known as secondary metabolites because they are thought to be non-essential for growth and development.[3] These chemical substances, however, confer a fitness advantage to the producing organism acting as repellants or attractants, pheromones, growth inhibitors and toxins.[3] They are very often species-specific,[2,5] and their potency, structural diversity and wide-ranging biological activities arose via the progressive development and fine-tuning of strategic biosynthetic mechanisms requiring the products and intermediates to interact specifically with enzymes and other protein targets.[6]

Major classes of natural products include the terpenoids and steroids, non-ribosomal peptides, enzyme co-factors, fatty-acid derived substances, polyketides and the alkaloids.[3] A number of examples of hybrid systems exist such as the non-ribosomal peptide-polyketides and new examples of these combinations continue to emerge.

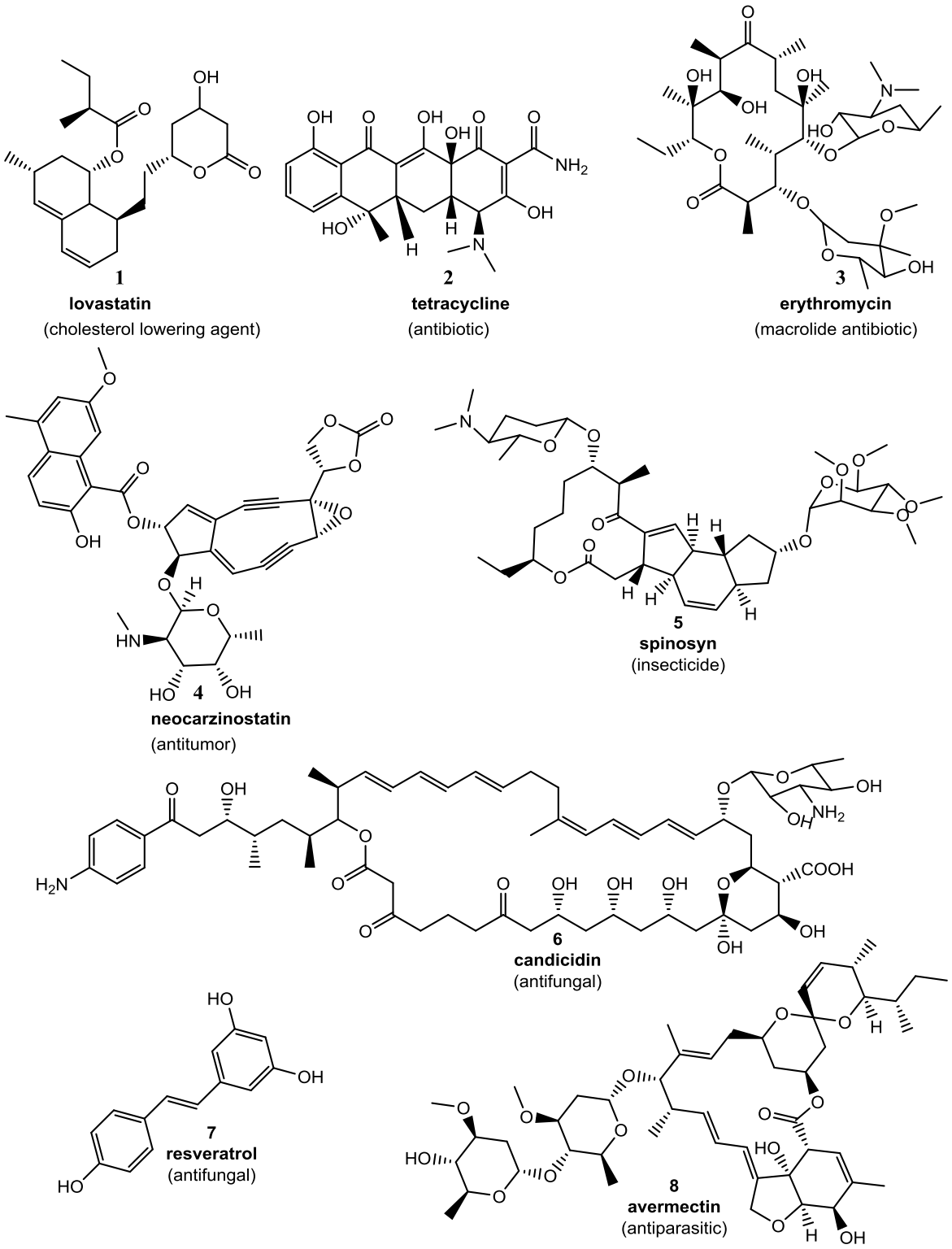


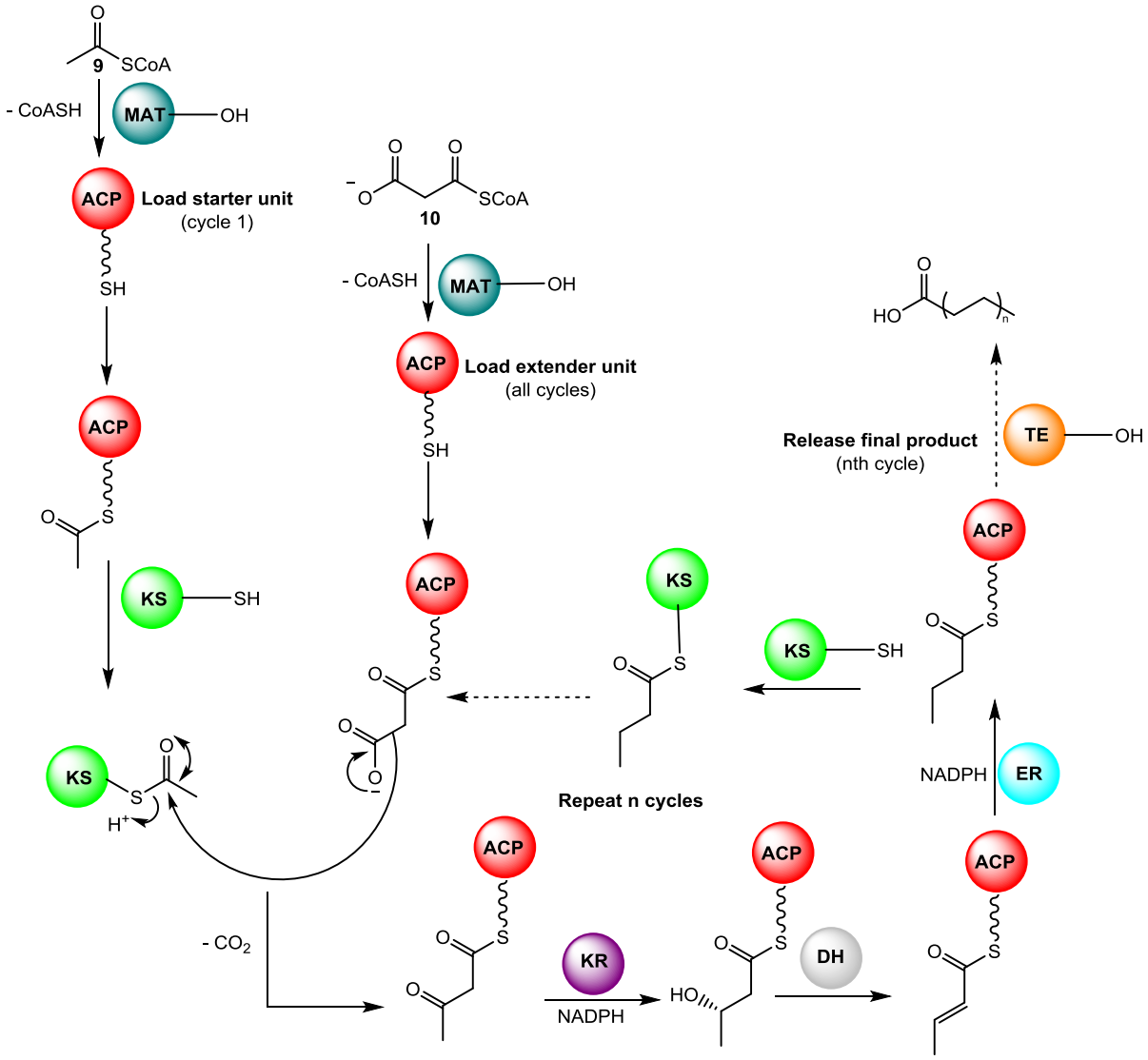
Figure 1: Representative structures of some medicinally and agriculturally important polyketides illustrating their vast structural diversity and density of chirality centres with their primary biological activity in parenthesis.

1.2 Polyketides

Polyketides are a large family of secondary metabolites with diverse biological activity.[1] These intriguing substances isolated from plants, bacteria, fungi and marine animals, have been found to possess anti-bacterial, anti-fungal, insecticidal, immunosuppressant, anti-hypertensive and cytotoxic activity.[1,7.8,9,10,11] Notable examples of industrially relevant polyketides include the cholesterol-lowering agent: lovastatin **1**, the clinically important antibiotics: tetracycline **2** and erythromycin **3**, the anti-tumor agent: neocarzinostatin **4**, the insecticide: spinosyn **5**, the antifungal agents: candicidin **6** and resveratrol **7**, and the antiparasitic agent: avermectin **8**, (**Fig. 1**).[9, 10, 11,12]

1.3 Polyketide Biosynthesis

Polyketides are biosynthesized from simple acyl-CoA precursors (such as acetyl-CoA, propionyl-CoA, methylmalonyl-CoA, butyryl-CoA) by successive condensation reactions catalysed by large multi-modular enzymes with distinct catalytic domains known as polyketide synthases (PKSs).[3,9] Polyketide synthases are similar to the fatty acid synthases (FAS) both structurally and in their biosynthetic mechanisms (**Scheme 1**). For this reason, they have also been classified by a similar convention based on the biosynthetic mechanisms they employ and the extender units they utilize.[9,11,12] Also in common with fatty acids, polyketides are biosynthesized by repeated head-to-tail linkage of carbon-carbon bonds from a small pool of acyl-CoA (acetyl-CoA **9** for example) starter units and malonyl-CoA **10** (and a few of its structural analogues) extender units.[13] All FASs are made up of the following catalytic components: the keto synthase (KS), the malonyl-acetyl transferase (MAT), the acyl carrier protein (ACP), the keto reductase (KR), the dehydratase (DH), the enoyl reductase (ER) and the thioesterase (TE).

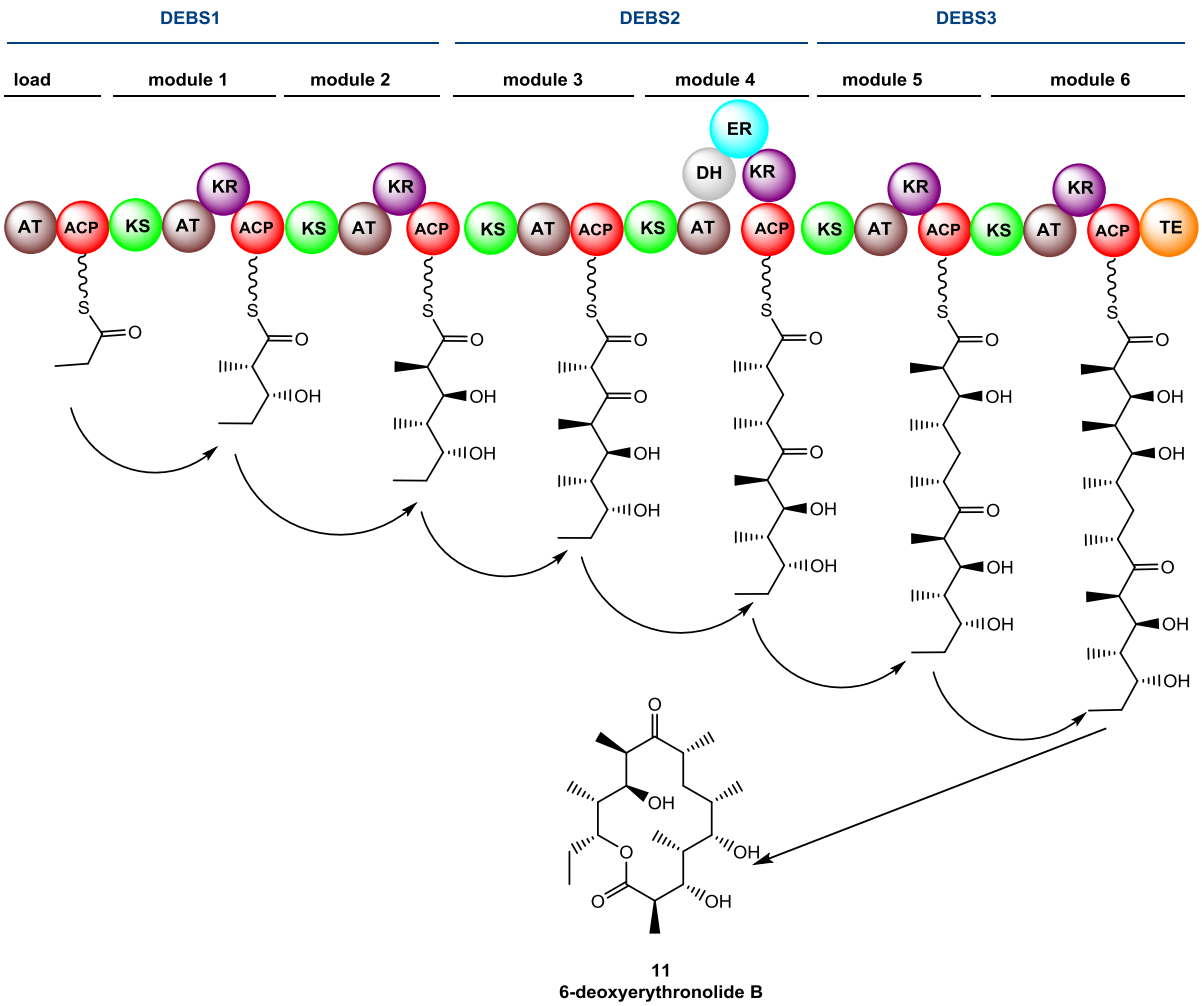


Scheme 1: Fatty acid biosynthesis cycle. MAT (malonyl-acetyl transferase); KS (ketosynthase); KR (β-ketoreductase, NADPH-dependent); DH: dehydratase; ER (enoyl reductase, NADPH-dependent).[13]

The KS catalyses the condensation reaction between the starter and extender units, the MAT transfers the starter and extender acyl-CoAs to the ACP while the KR, DH, and ER domains process the β-carbonyl of the growing thioester-bound intermediate. Identical catalytic domains are found in PKSs and their mechanisms will be discussed in detail in section **1.3.1** below. The chemistry involved is a decarboxylative Claisen-type condensation reaction and the keto group of the resulting β–ketoester may be subsequently reduced to the

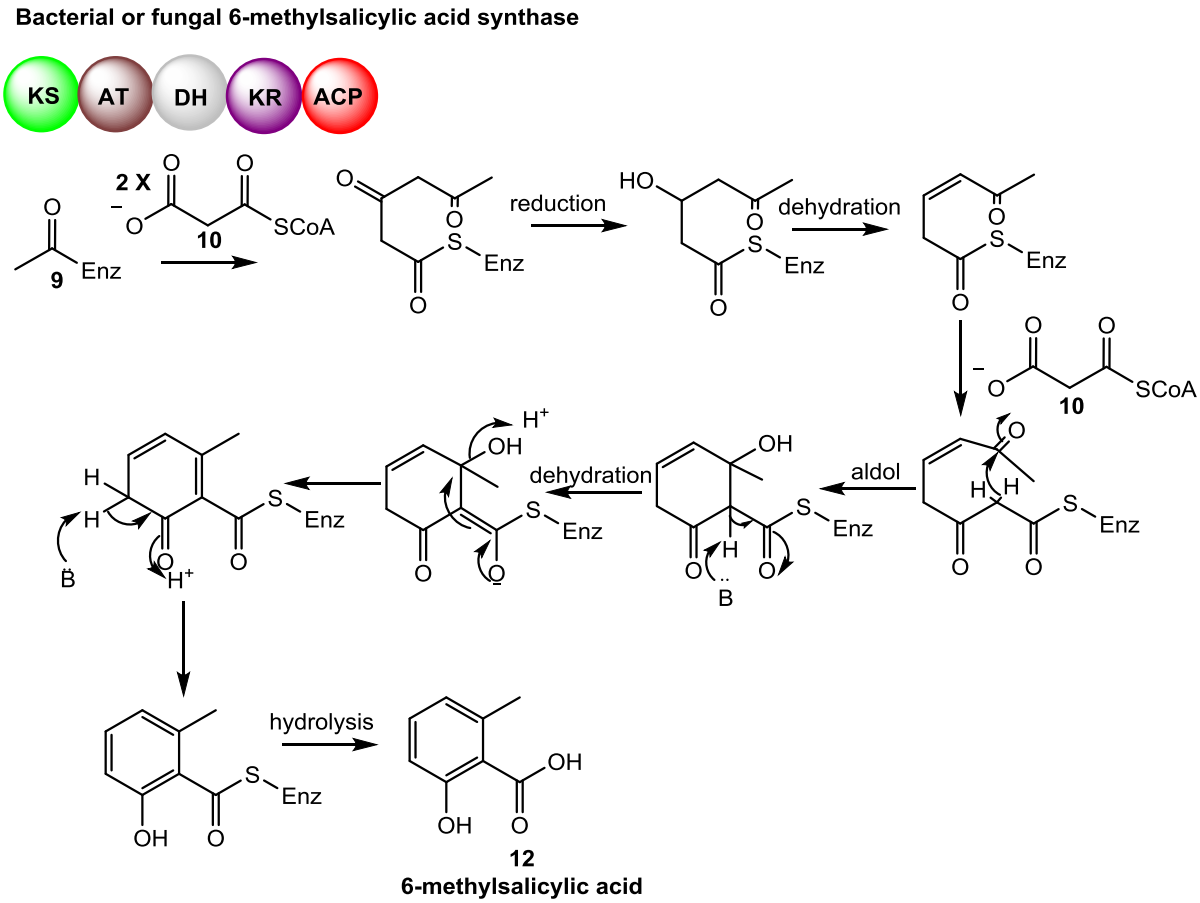
hydroxyl then dehydrated to give an unsaturated double bond which can also be reduced further, to a saturated methylene group.[13]

Polyketide synthases can be broadly classified into types I, II and III.[9,13] Type I polyketide synthases are large multifunctional proteins with catalytically distinct domains covalently linked together. These proteins are similar to the mammalian FAS (discovered and characterized before the polyketides) which are also called type I synthases.[13] The type I PKSs can be further divided into two classes: type I modular and type I iterative PKSs. The catalytic domains present in type I modular PKSs are organized into modules with each module, catalysing a single distinct reaction during the process of polyketide chain elongation. An extensively studied example is the 6-deoxyerythronolide B synthase (DEBS) involved in the biosynthesis of erythromycin (**Scheme 2**), where six rounds of successive chain extension of a propionyl starter unit with methylmalonyl units results in a heptaketide intermediate. Other examples of products of type I PKSs include spinosyn **5** and avermectin **8**.[9]



Scheme 2: Module and domain organization of the type I modular PKS – DEBS (erythromycin **3** biosynthesis).[13]

Type I iterative PKSs also comprise of catalytic domains which are covalently linked together to form a multifunctional enzyme. In this case however, the single module carries out polyketide chain extension by acting iteratively (at least some of the enzymatic domains are used in each successive cycle).[13] Examples of polyketides assembled by a type I iterative mechanism include lovastatin and 6-methylsalycylic acid **12** (6-MSA). One acetate starter unit and three malonate extender units are used for the assembly of 6-MSA **12**. The 6-MSAS PKS comprises of five domains: KS, MAT, DH,KR and ACP (**Scheme 3**).[13]

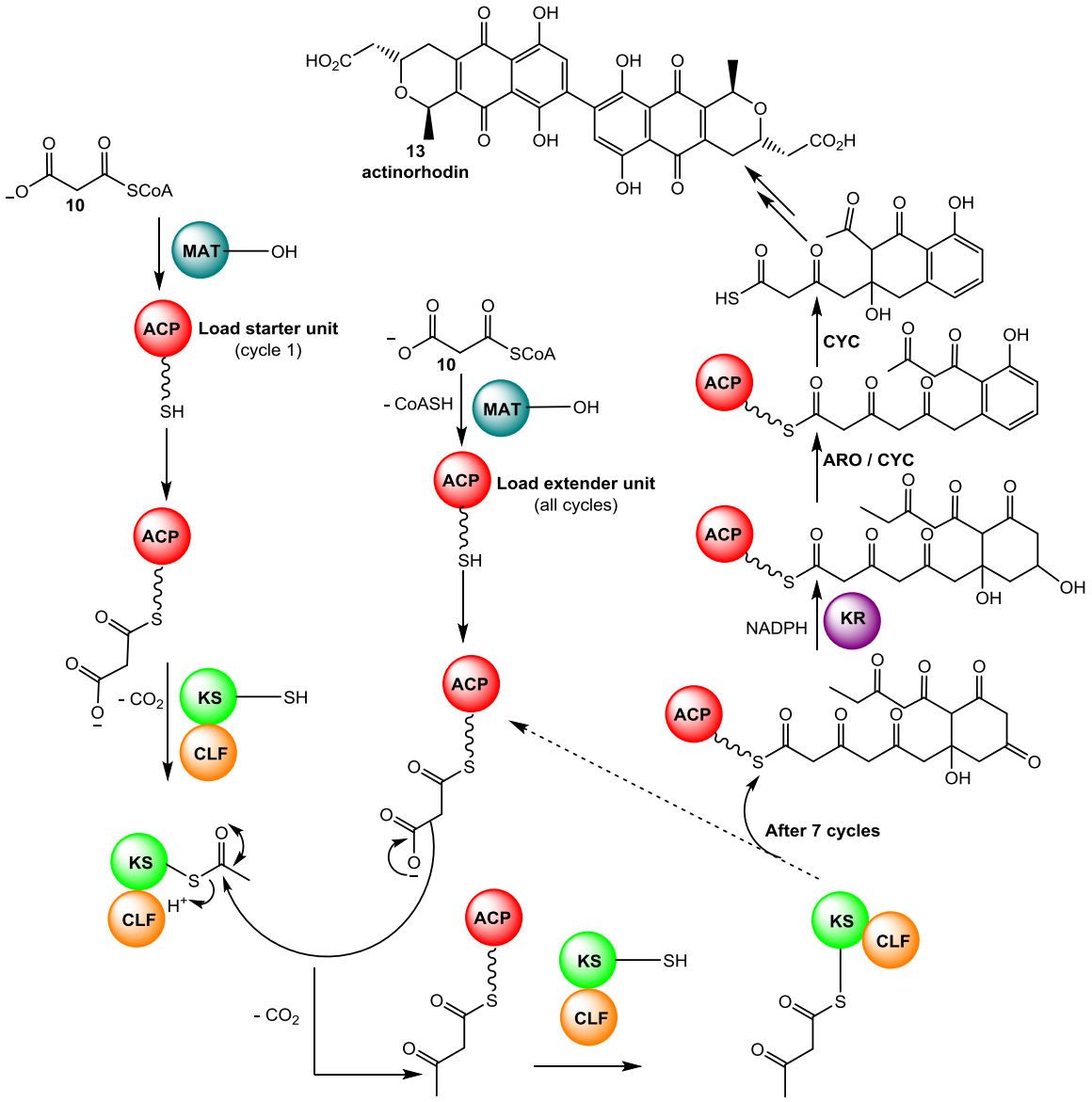


Scheme 3: Polyketide biosynthesis by a type I iterative PKS (6-methylsalicylic acid biosynthesis).[13]

The first extension step does not involve reduction of the β-keto group of the thioester intermediate prior to its reaction with a second malonate unit. The second round of condensation is followed by reduction and then dehydration of the β-keto group. After a third condensation with another malonate unit, an intramolecular aldol condensation results in ring formation and subsequent tailoring reactions lead to aromatization and release of the thioester intermediate to yield 6-MSA **12**.[13]

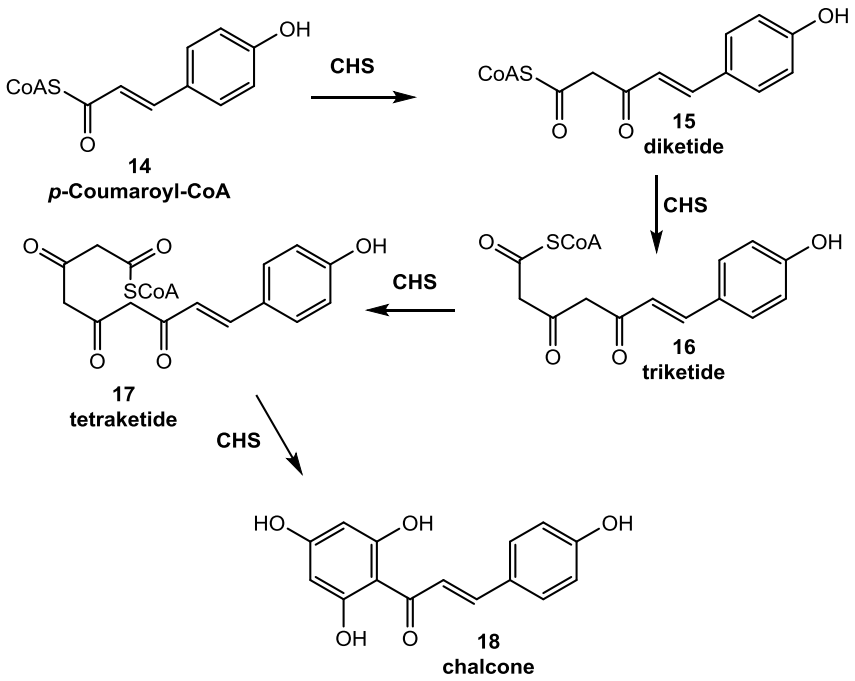
Type II PKSs are involved in the biosynthesis of aromatic polyketides (mostly polycyclic) such as tetracycline **2**. They comprise of an enzyme complex of discrete protein subunits that carry out a set of concerted synthetic steps in an iterative manner.[9] This is

illustrated by the extensively studied actinorhodin biosynthetic system (**Scheme 4**) in *Streptomyces coelicolor*.[14] Actinorhodin is biosynthesized from eight units of malonyl-CoA. The starter unit is transferred to the ACP by a MAT from an endogenous FAS, prior to its decarboxylation and transfer to the KS active site of the KS-CLF complex. Seven successive extensions are then carried out with the same group of enzymes acting iteratively. This is followed by the action of auxiliary PKS subunits: KR, a bifunctional aromatase/ cyclase (ARO/CYC) and a cyclase (CYC).[14]



Scheme 4: Aromatic polyketide biosynthesis by a type II PKS (actinorhodin biosynthesis).[14]

Both type I and type II PKSs are dependent on acyl carrier proteins (ACP) to covalently tether their acyl-CoA derived substrates, while the type III PKSs act directly on the acyl-CoA substrates.[9,13] Type III PKSs (often referred to as chalcone synthases in plants) are also used for the biosynthesis of aromatic polyketides (monocyclic and bicyclic) of the phenylpropanoid family (for example resveratrol **7**) which exhibit antiviral, antioxidant and antitumor activities. Chalcone synthesases provide the biosynthetic precursors for the phenylpropanoids; its biosynthetic mechanism involves the use of *p*-coumaroyl-CoA **14** as a starter unit which undergoes extension with three molecules of malonyl-CoA **10** to yield a tetraketide (**Scheme 5**).[13] The tetraketide intermediate then undergoes a cyclisation via Claisen condensation to yield chalcone. The chalcone synthesases are homodimeric condensing enzymes that also act iteratively like the type II PKSs.[9] Type III PKSs however, possess the unique ability to carry out decarboxylation, condensation, cyclisation and aromatization reactions with the same enzyme active site, unlike their type I and type II counterparts.[13]



Scheme 5: Chalcone biosynthesis by a type III PKS (CHS).[13]

Increasing insight into biosynthetic mechanisms over the years has resulted in the discovery of a significant number of other polyketide sub-classes that may not fit strictly into the above-mentioned classifications. These findings further underscore the immense potential structural diversity of polyketide natural products and justify the need for continuous, active research efforts to better understand their biosynthesis.

1.3.1 Type I Modular Polyketide Synthases

Type I modular polyketide synthases are gigantic multi-enzyme polypeptides with primary protein sequences encoding catalytic domains that are covalently attached by linker regions which contribute to the structural integrity of the biosynthetic system.[13] Each domain catalyses a separate reaction in the assembly process and domains are organized into functional units called modules. Within a particular module, domains catalyse separate but complementary steps in the extension and modification of the growing polyketide chain. Upon completion of one cycle of these reactions, the polyketide intermediate is transferred to the next module which continues the process in an assembly-line fashion.[13] The structural features (and biological activity) of the polyketide its functional groups and carbon chain length are largely dependent on each successive round of chain extension: the number of modules present and the types of domains embedded within these modules.

A typical PKS minimal module contains ketosynthase (KS), acyltransferase (AT) and an acyl-carrier protein (ACP). These domains catalyse the reactions required for one cycle of chain elongation. The ketosynthase (KS) domain catalyses a Claisen-type thioester condensation reaction that results in chain elongation, the acyl transferase (AT) domain selects the acyl-CoA building block from the cellular pool and attaches it to the adjacent acyl carrier protein (ACP) domain which serves as a point of covalent attachment for successive biosynthetic intermediates.[15,16] Auxiliary domains which may be included in a functional module include: ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER).[15,16,17] In

addition to the commonly observed auxiliary domains, a few atypical catalytic domains have been discovered within these biosynthetic systems. Examples include: methyltransferases, transaminases, halogenases and cyclases.[17]

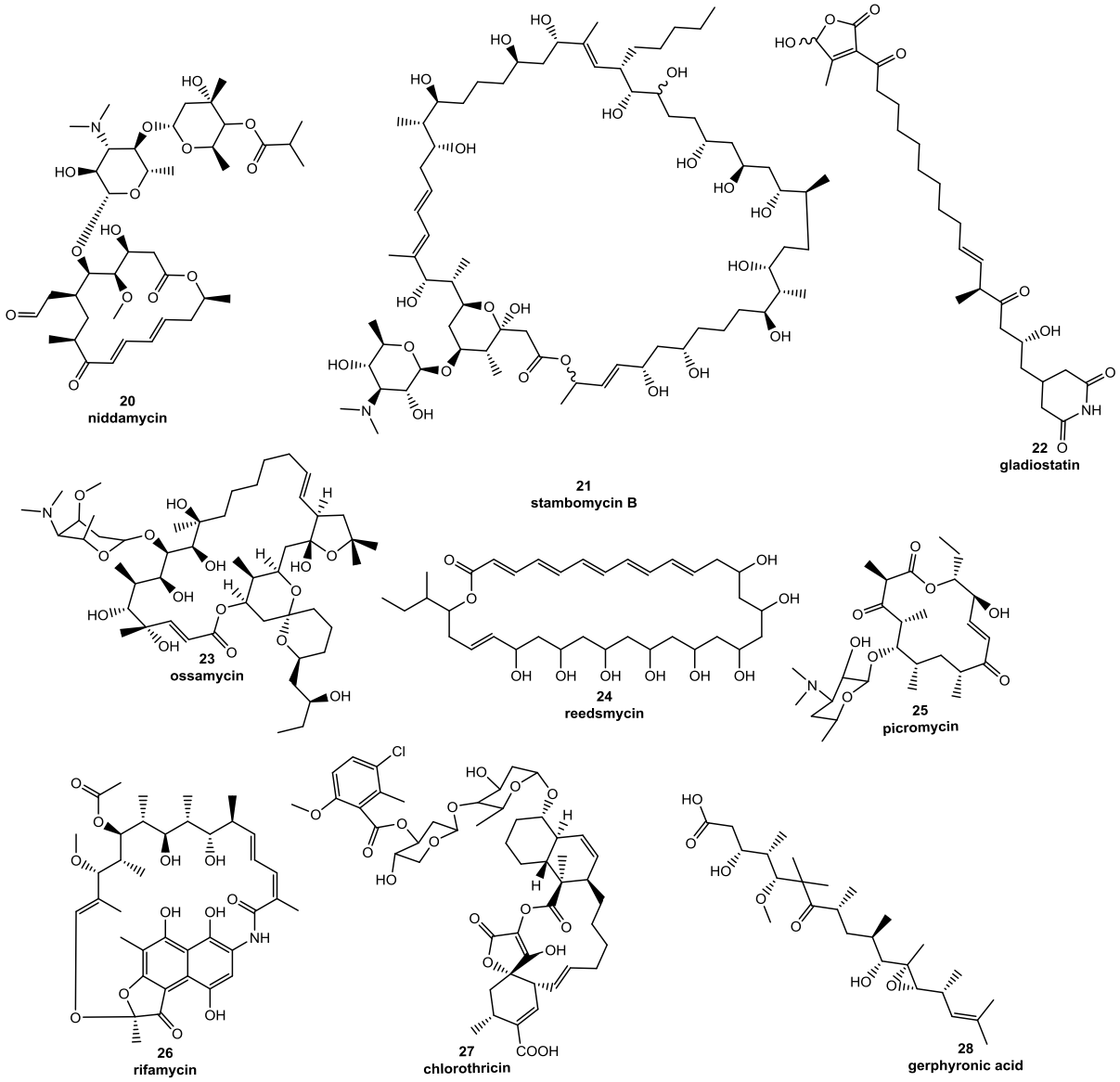


Figure 2: Biosynthetic products of representative type I PKSs illustrating structural diversity.

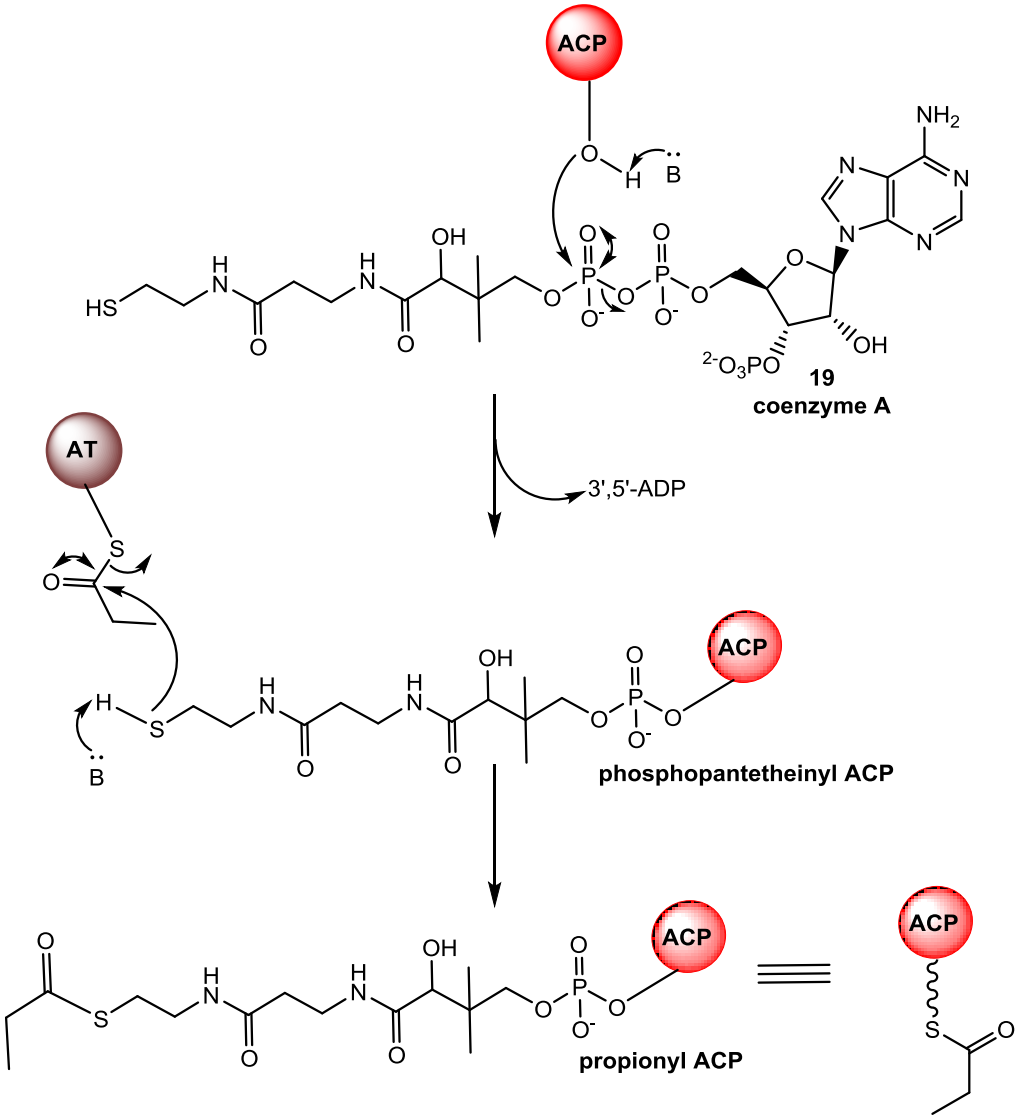
The terminal domain of the best-studied modular PKS DEBS and the majority of the initially discovered modular PKSs, contains a thioesterase (TE) domain. The TE domain catalyses the release of the fully assembled polyketide chain via an addition-elimination reaction using either an external nucleophile (typically water) resulting in the formation of a linear polyketide product (e.g a carboxylic acid); or an internal nucleophile (usually an amino or hydroxyl group) located on the polyketide chain resulting in a macrocycle (a macrolactone or macrolactam).[18] In subsequent years, however, an increasing number of PKSs have been characterized that do not utilize a thioesterase domain for chain termination. These biosynthetic systems adopt various other catalytic mechanisms for release of the fully assembled polyketide chain resulting in the introduction of various functional groups. Examples of some type I PKS systems which employ unusual chain release mechanisms include alpha-oxoamine synthases (undecylprodigiosin biosynthesis), phosphorylated butenolide synthases (gladiostatin biosynthesis) and tetronate synthases (chlorothricin biosynthesis).[8]

1.3.2 Polyketide Chain Initiation

Polyketide biosynthesis is initiated by a loading module that selects the acyl-CoA starter unit. Various structural organizations of loading modules have been characterized (**Scheme 6**). One type of loading module (which is the most common loading module organization) comprises KS^Q (a ketosynthase which is incapable of catalysing a condensation due to a cysteine to glutamine mutation in its active site), AT and ACP domain.[15,17] The AT domain binds the acyl-CoA via an active site serine residue that forms an ester bond with its substrate; these particular AT domains are specific for starter units that contain carboxylic acid groups (such as malonyl and methylmalonyl-CoA) due to the presence of a conserved active site arginine residue which serves to stabilize the free carboxylic acid functional group in the substrate.[15,17] The covalently-tethered acyl group is transferred to the phosphopantetheine thiol group of the ACP domain. The KS^Q domain then

decarboxylates the resulting malonyl or methylmalonyl thioester yielding either an acetyl- or propionyl-thioester respectively.[15,17] An example of a type I PKS which employs this biosynthetic mechanism for chain initiation is the spinosyn **5** PKS from *Saccharopolyspora spinosa*.[18] Other notable examples of type I PKS biosynthetic systems that employ this priming mechanism include (**Figure 2**) niddamycin **20**, stambomycin **21**, ossamycin **23**, reedsmycin **24** and picromycin **25** among others.[17,19,20,21,22,23,24]

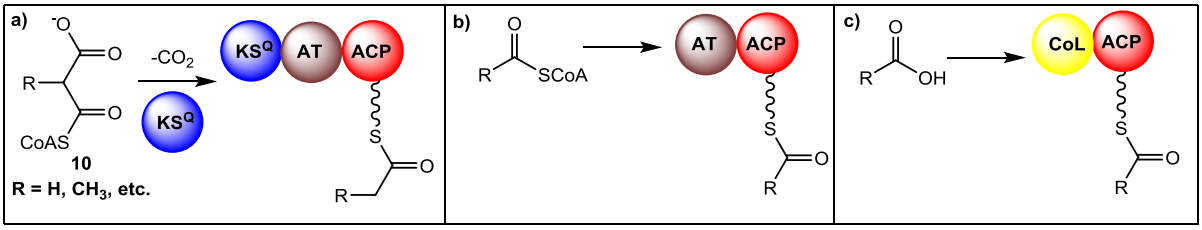
A second type of loading module contains an AT-ACP didomain. These modules typically utilize a much broader substrate range not limited to β–carboxyl-CoA substrates. They may select for propionyl-CoA, butyryl-CoA and 2-methylbutyryl-CoA among many others. The erythromycin **3** and avermectin **8** PKSs employ this chain initiation mechanism.[15]



Scheme 6: Polyketide starter unit priming mechanism in the DEBS type I modular PKS.[3]

A third route for chain initiation involves a CoA-ligase-type domain, which utilizes ATP to activate carboxylic acids and load them onto a dedicated ACP domain.[14] These domains are either fused to the N-terminus of the ACP domain or act *in trans*. Examples of biosynthetic systems where this loading mechanism can be found, include the rifamycin **26** PKS.[15] Polyketide chain initiation is also conducted by loading modules containing a domain showing sequence similarity to general control non-repressive 5 (GCN5) N-acetyl-transferase (GNAT)-like decarboxylases.[25,26] These are selective for substrates containing

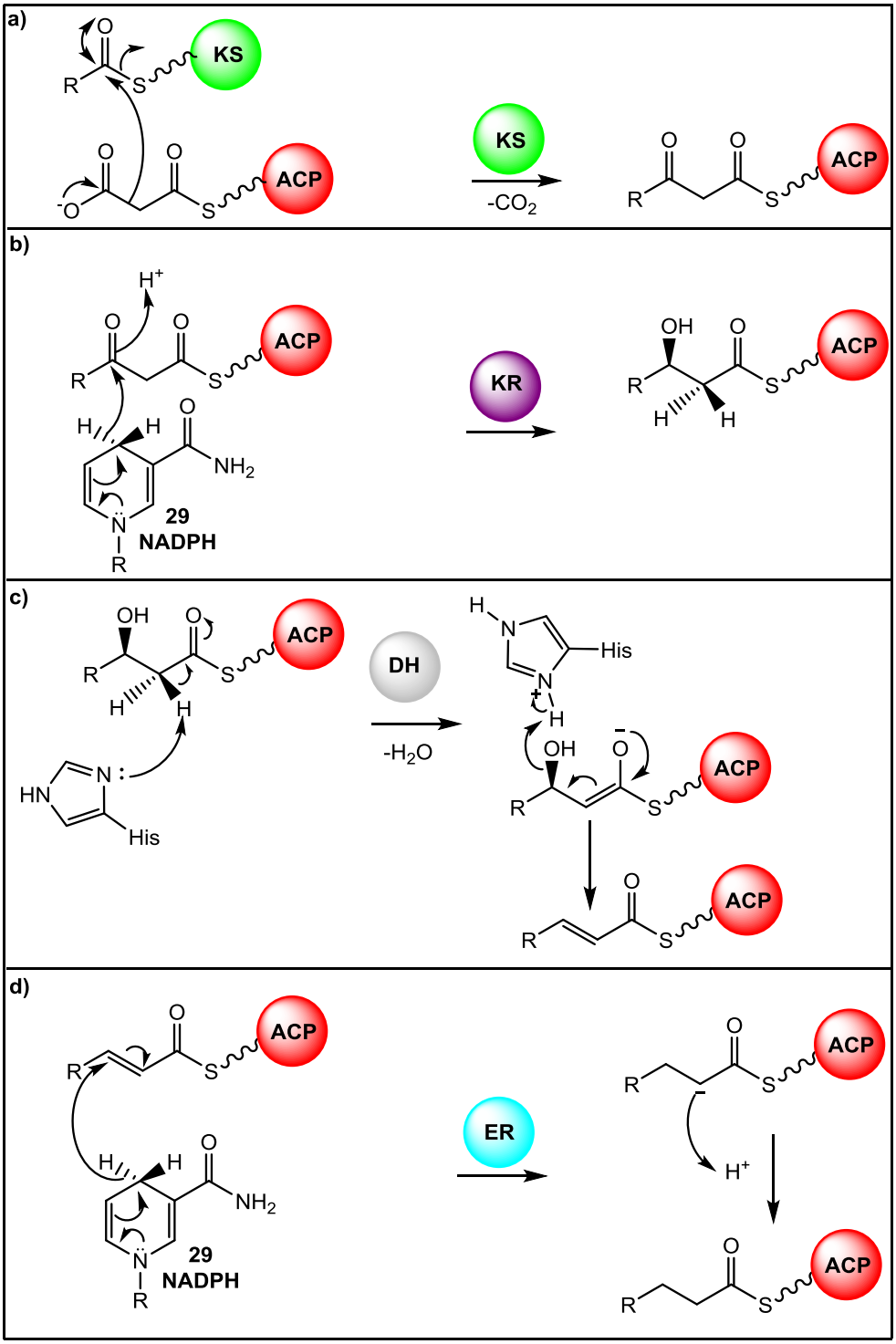
two carboxylic acid functional groups (malonyl-, methylmalonyl- and dimethylmalonyl-CoA for example).[25,26] The product of a type I modular PKS from the myxobacterium *Cystobacter violaceus*: gephyronic acid **28** is an example of a secondary metabolite biosynthesized via this priming mechanism (**Scheme 7**).[25,26]



Scheme 7: Priming of the ACP by three types of PKS loading modules **a)** type A; **b)** type B; and **c)** type C.[15]

1.3.3 Polyketide chain extension and modification

The chain extension process (**Scheme 8**) begins when the priming acyl-group is transferred to the KS domain of the first extension module by the loading module ACP domain.[3] The point of attachment to the KS domain is a thioester bond with a cysteine residue.[3] The AT and ACP domains of the first extension module, select and receive an extender unit (usually malonyl-CoA **10** or structurally related derivatives). Carbon-carbon bond formation is catalysed by the KS domain bearing the priming acyl group. A decarboxylative Claisen-type condensation occurs, where the malonyl group (or its derivative) a molecule of carbon dioxide is eliminated irreversibly to form an incipient β–carbanion nucleophile that attacks the carbonyl group of the upstream priming unit to give an ACP-bound β–keto thioester.[3]



Scheme 8: Mechanisms of polyketide chain extension and modification **a)** KS catalysed chain extension by decarboxylative Claisen condensation **b)** NADPH-dependent reduction of the β-carbonyl catalysed by the KR domain **c)** DH catalysed elimination of water **d)** reduction of the α, β-unsaturated thioester.[16]

Prior to further chain extension, the β–keto thioester intermediate may undergo further processing. This is largely dependent on the presence (or absence) of auxiliary domains within each extension module, or on the action of stand-alone tailoring enzymes. Typical examples of these auxiliary domains include the KR, DH and ER domains. The keto group of the ACP-bound β-keto thioester can be reduced to a hydroxyl group by the NADPH-dependent KR domain.[15] Polyketide synthase KR domains are classified based on the stereo chemical outcome of the reduction reaction. A-type KRs produce L-configured alcohols while B-type KRs produce D-configured alcohols; KR domains can also catalyse epimerisation of α–substituents. A further classification level is therefore, based on the orientation of the α–substituent in the final product; A1 and B1 KRs produce D-configured α–substituents while A2 and B2 KRs produce L-configured α–substituents.[15]

If a DH domain is present, it catalyzes the elimination of water to form an α, β-unsaturated thioester.[15] Most often, the DH domains in multi-modular PKSs produce *trans*-alkenes although, *cis*-alkenes have been characterised occasionally.[27] This is because the stereochemical outcome of the dehydration reaction is dependent on the stereo chemical configuration of the substrate hydroxyl group; and most DH domains follow type B KR domains.

Reduction of the α, β-unsaturated thioester is catalysed by an ER domain which is an NADPH-dependent oxidoreductase. A hydride ion is transferred to the β-carbon of the ACP-bound intermediate, to form an enolate, followed by protonation at the α–carbon to form an alkane group. Epimerization of α-substituents has also been found to be catalysed by ER domains. Enoylreductase domains with a conserved tyrosine residue in the vicinity of the NADPH **29** co-factor produce *L*-α-substituents, while ER domains that do not possess this tyrosine residue produce *D*- α–substituents.[15]

Polyketide chain extension continues with successive addition of two carbon units to the growing polyketide chain with each extension cycle. In typical type I modular PKSs, the number of extension cycles is the same as the number of extension modules within the PKS. However, various examples of biosynthetic systems where module skipping and module iterations are employed have been reported .[28,29,30]

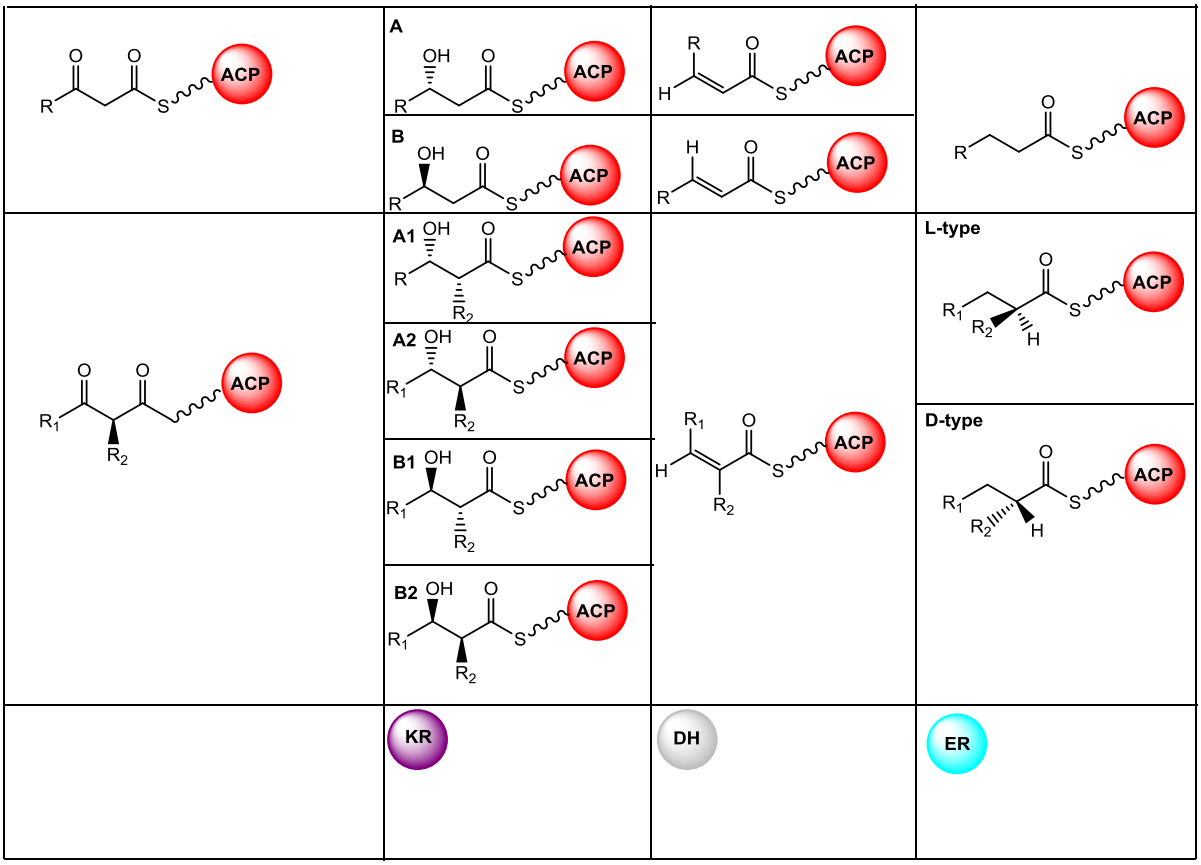
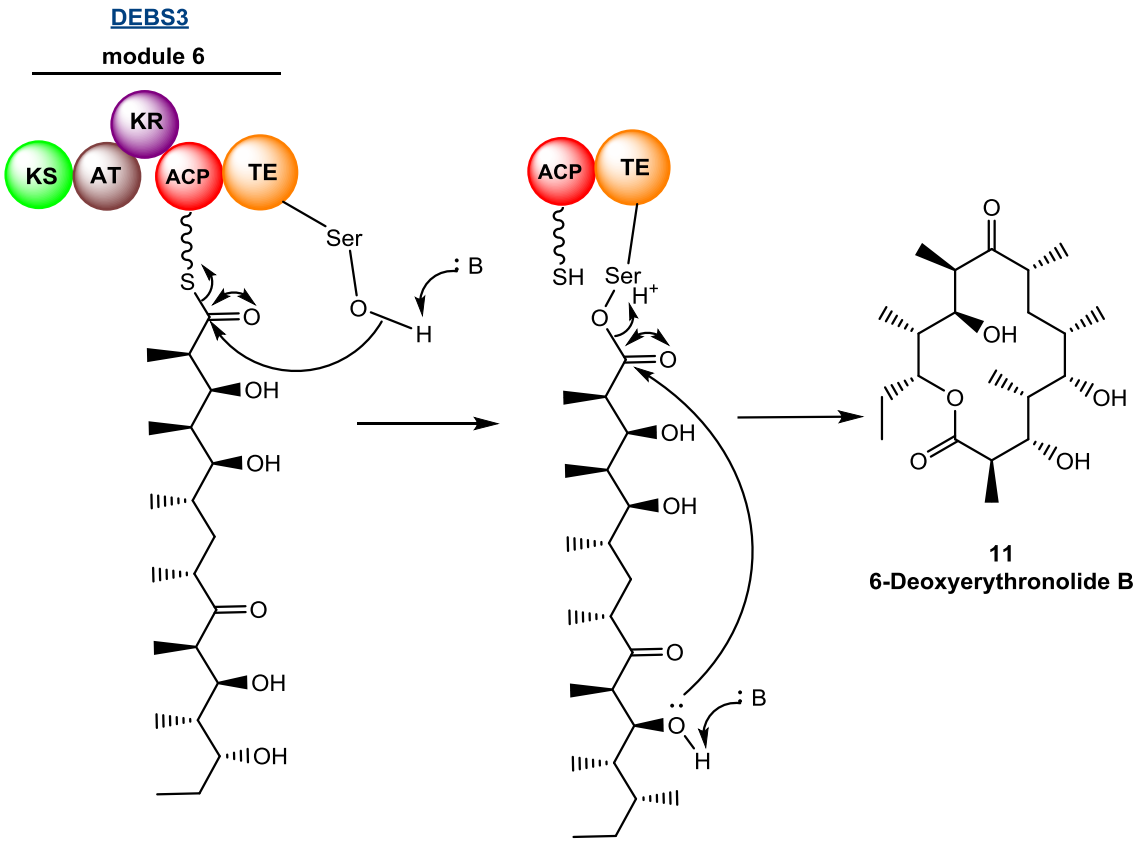


Figure 3: Types of reducing domains (KR, DH and ER) and potential product outcomes.[15]

1.3.4 Polyketide Chain Termination / Release

Different types of mechanisms have been used to release the fully assembled ACP-tethered polyketide chain from PKS and other closely related multienzyme systems that assemble secondary metabolites. The chain termination mechanism is important because, in the absence of this biosynthetic step, the polyketide intermediate will remain covalently

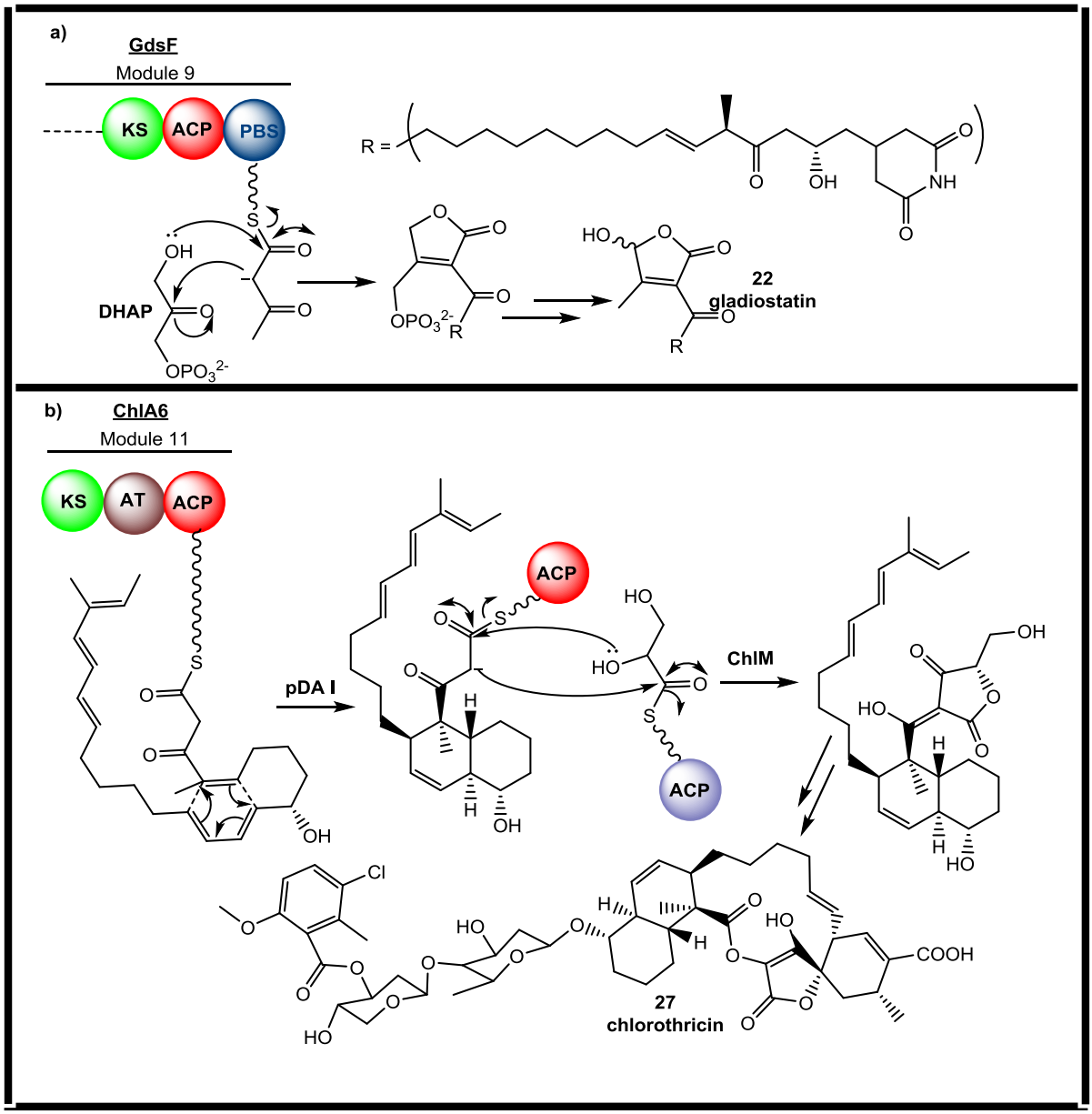
attached to the enzyme via its thioester bond. Also, it provides an additional means for introduction of structural variation to the secondary metabolite. The most common mechanism of chain release employs a thioesterase (TE) domain (**Scheme 9**), which is found in the archetypal DEBS PKS used for the biosynthesis of erythromycin.[31]



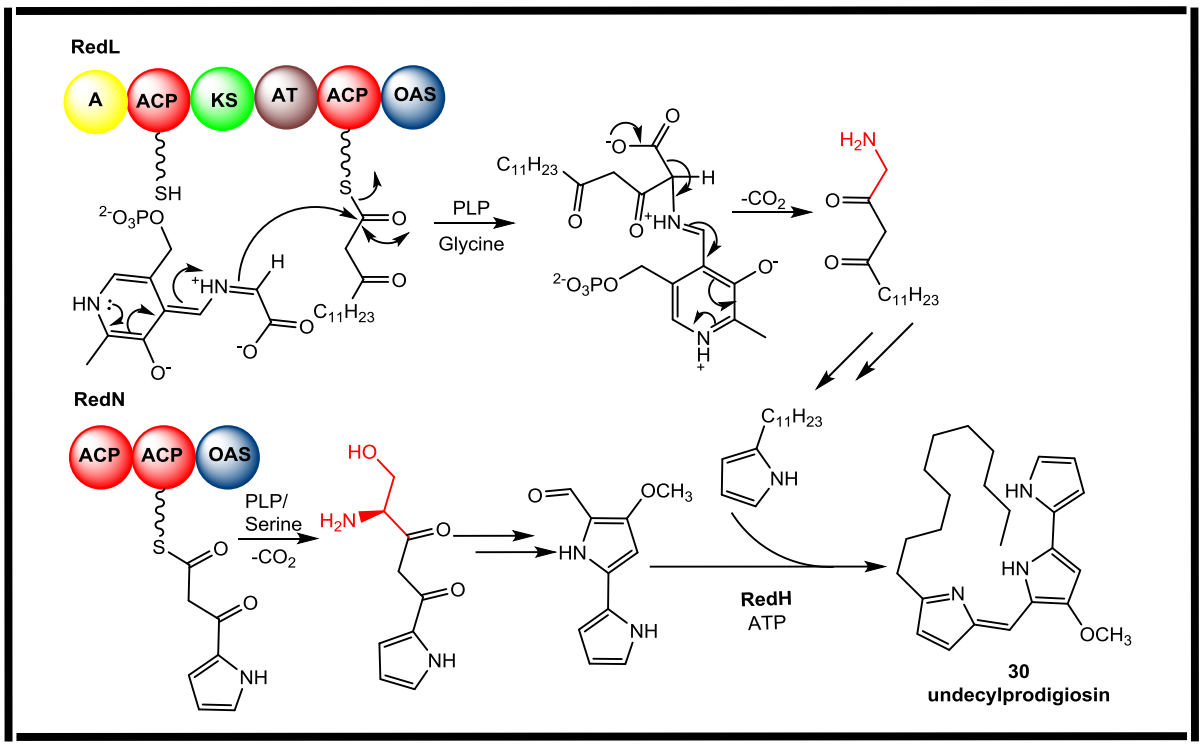
Scheme 9: TE catalysed polyketide chain release (DEBS PKS).[31]

These thioesterases are α/β-hydrolases and are usually situated in the final module of the PKS. They possess a catalytic triad of serine, histidine and aspartate. The histidine residue accepts a proton from the hydroxyl oxygen of the serine residue which attacks the carbonyl group of the ACP-thioester in the fully assembled polyketide chain, replacing the thioester bond with an acyl-*O*-serine oxoester on the TE.[31] The TE-bound oxoester intermediate may then be released via nucleophilic attack from either an external or internal nucleophile. An external nucleophile (water for instance) produces a linear polyketide (a

carboxylic acid) while an internal nucleophile would give a macrocycle (macrolactone, macrolactam or macrothiolactone depending on the nucleophile).[15,18] Stand-alone TEs, classified as type II TEs, have also been identified.[31] Typically, type II TEs perform an editing role by removing biosynthetic intermediates of the wrong structure that are stuck to the PKS along the assembly-line, a role which is distinct from the type I TE which recognizes the mature polyketide chain for release.[18,31] However, in some systems such as in the biosynthesis of the polyethers and enediynes, type II TEs are used to release the final polyketide product.[18]



Scheme 10: Examples of unusual mechanisms of chain release in type I modular PKSs. **a)** phosphobutenolide synthase (gladiostatin **22** biosynthesis)[32] and **b)** spirotetronate synthase (chlorothricin **27** biosynthesis)[33].



Scheme 11: Chain release by an OAS in undecylprodigiosin biosynthesis.[34]

Apart from TE-mediated chain release mechanisms, other catalytic mechanisms are employed by PKSs. Examples (**Schemes 10** and **11**) include the phosphorylated butenolide synthases (gladiostatin **22** biosynthesis)[32] and tetronate synthases (chlorothricin **27** biosynthesis)[33]; as well as the PLP-dependent alpha-oxoamine synthases (undecylprodigiosin **30** biosynthesis);[34] and NAD(P)H-dependent thioester reductases.[15]

In the biosynthesis of gladiostatin **22** and chlorothricin **27** (**Scheme 10**), each of the fully assembled β-ketothioester intermediates undergo condensation with C_3 derivatives of glycerol (intermediates from the glycolysis pathway) resulting in the formation of a butenolide and release of the polyketide intermediate from the ACP.[32,33] While the chain release mechanism for the assembly of undecylprodigiosin **30** (**Scheme 11**), involves two decarboxylative Claisen condensation reactions (for each of the pyrolle-containing biosynthetic intermediates) between either glycine or serine and the respective β-ketothioester intermediates resulting in the formation of a new carbon-carbon bond and the

introduction of an amine group. This chain release mechanism is catalysed by the α-oxoamine synthase (OAS) domain present at the C-terminus.[34]

NAD(P)H-dependent thioester reductases have been identified and characterized within certain other biosynthetic systems including NRPSs and PKS / NRPSs; in addition, they have been predicted to be utilized in some type I modular PKSs. These enzyme domains have been classified as members of the short-chain dehydrogenase / reductase enzyme family and will be discussed further below.[31]

1.4. Short Chain Dehydrogenases / Reductases

The pyridine dinucleotides, nicotinamide adenine dinucleotide (NAD^+ / NADH) and nicotinamide adenine dinucleotide phosphate ($NADP^+$ / NADPH) are essential cofactors (**Figure 4**) which are derivatives of vitamin B_3 **32** (nicotinamide).[3] These cofactors are present in all living systems where they serve as hydride acceptors and donors (dehydrogenases/reductases) during biochemical redox reactions (**Scheme 12**) required for energy metabolism, biosynthesis and detoxification.[2,3] In addition to the vital roles played by these cofactors in the biosynthesis of major cellular components such as DNA and lipids; NADPH and NADH-dependent enzymes are of prime importance for the biosynthesis of numerous secondary metabolites.[35]

The redox-active portion of the co-enzyme is the nicotinamide ring, it is directly involved in hydride transfer during catalysis.[32,36] Enzymes which are NAD(P)H-dependent have often evolved to utilize either NADH or NADPH specifically; despite the fact that there is only a slight structural difference between the two cofactors (the presence of a single phosphate group at the 2' position of the adenosine ribose in NADPH).[36] In general, NADPH is preferentially utilized for biosynthetic (anabolic) processes while NADH is used in catabolic processes.[36] Exceptions, however, exist and this generalization is not always strictly adhered to. Enzymes are also specific for one (or the other) enantiotopic C-4 hydrogen atom of

NAD(P)H. The cofactor binds non-covalently to the enzyme at the beginning of each reduction cycle and is released in its oxidized form at the end of the reaction. Thus, the progress of the reaction can be readily monitored by measuring the change in concentration of NAD(P)H in solution by either UV spectroscopy or fluorimetry.[36]

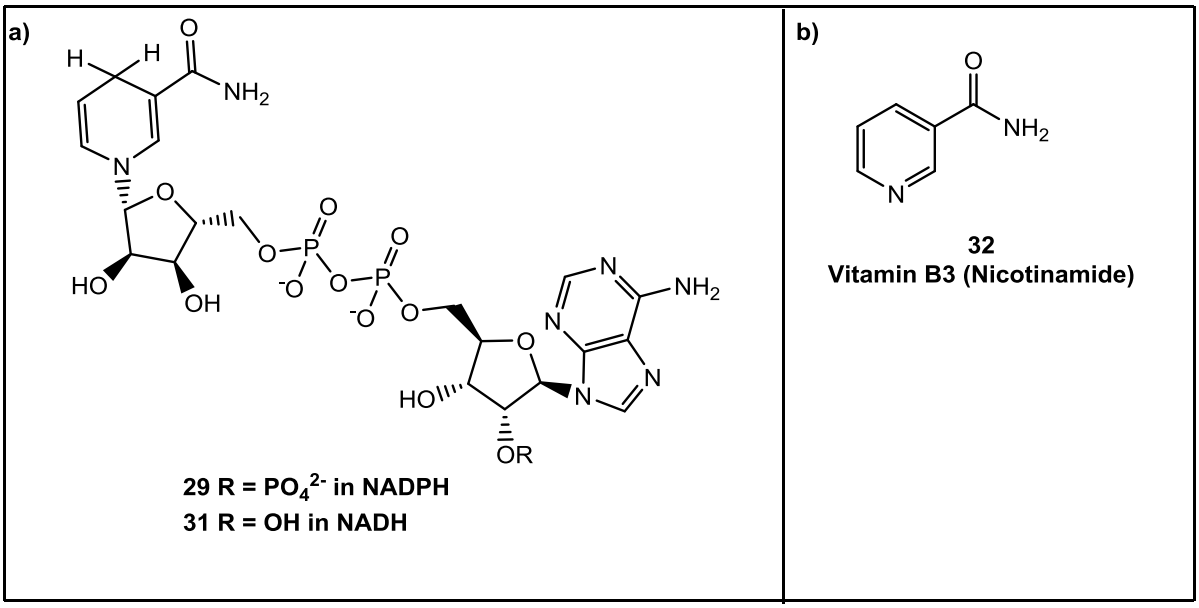
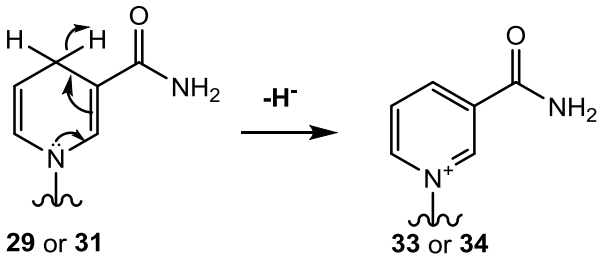


Figure 4: Structures of **a)** NADPH **29**, NADH **31** and b) Vitamin B_3 (Nicotinamide) **32**.



Scheme 12: Oxidation of NADH / NADPH to $NADP^+$ / NAD^+.

Short chain dehydrogenases / reductases are NAD(P)H-dependent enzymes that catalyse a wide variety of reactions such as ketoreduction, isomerization, dehalogenation and dehydration with a broad spectrum of substrates ranging from alcohols, sugars, steroids to aromatic compounds.[36,37] They are typically characterized by a secondary structure with

ca 250 to 350 amino acid residues, and an N-terminal glycine-rich Rossman fold which binds the co-factor, while the C-terminus constitutes the substrate binding part.[37] The α/β Rossman fold generally comprises of six to seven parallel β-sheets surrounded by three or four α–helices.

In thioester reductases, this structural motif is situated at the N-terminus where it serves to bind the NAD(P)H co-factor (the hydride source for reductive release of the ACP-bound thioester intermediate).[37,38] The C-terminus is responsible for substrate recognition and comprises of a flexible hydrophobic structural motif.[38,39] The catalytic mechanism involves reduction via the transfer of a hydride anion from NAD(P)H to the enzyme-bound thioester resulting in the formation of a thio-hemiacetal intermediate which is stabilized by the active site tyrosine residue.[39] The tetrahedral intermediate collapses to release the reduced aldehyde product and free the thiol group on the phosphopantetheinyl arm of the ACP domain. This is a two electron reduction.[36,39] An additional reduction by a second NAD(P)H equivalent may occur resulting in the conversion of the aldehyde to its corresponding alcohol in this case. Two hydride ions are involved, resulting in a four electron reduction.[40] A triad of conserved tyrosine, lysine and serine (or threonine) residues present in the enzyme active site catalyse hydride transfer from the cofactor to the substrate via a mechanism involving a proton relay system proposed to comprise of a number of coordinated water molecules and the 2'-hydroxyl of the nicotinamide ribose.[36,37] Enoylreductases (see above section **1.3.3**) belong to the medium chain NAD(P)H-dependent dehydrogenase/reductase (MDR) superfamily. While ketoreductases (KR) and thioester reductases (see below section **1.5.2**) domains are generally short chain NAD(P)H-dependent dehydrogenases/reductases (SDR).[37]

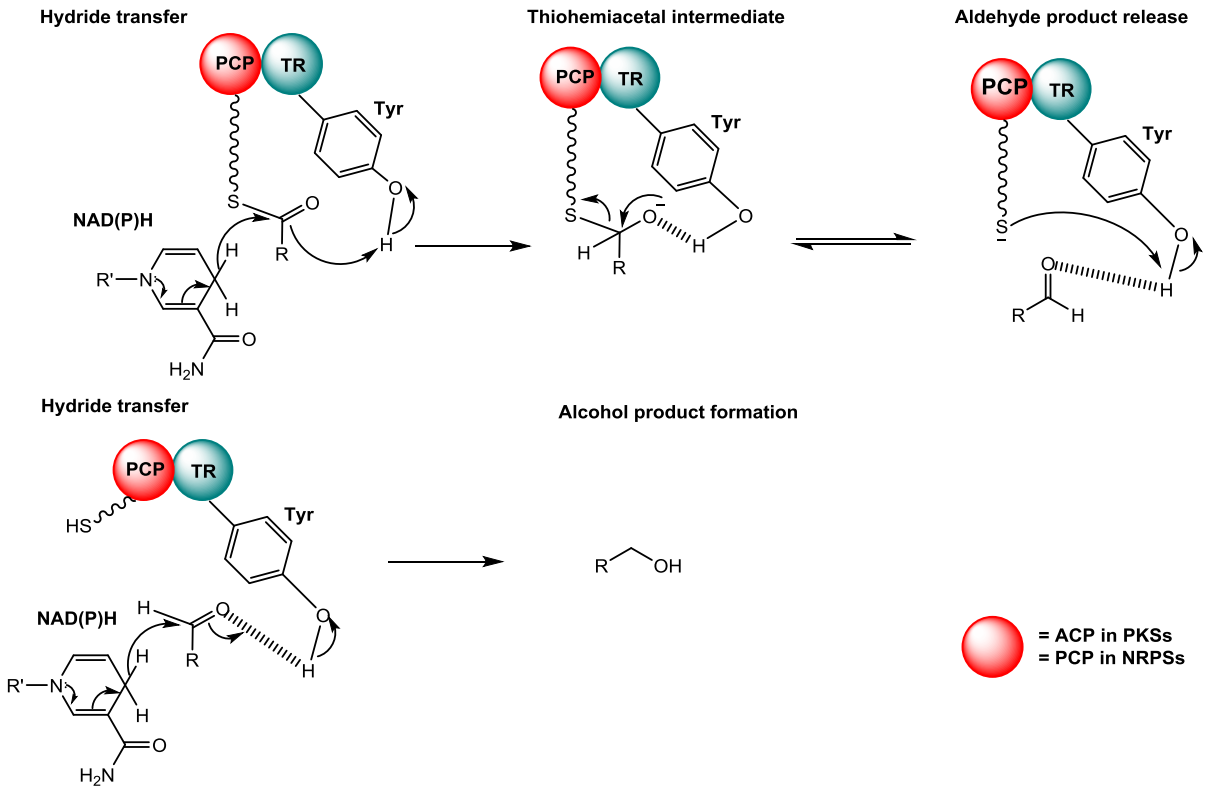
The SDRs have been broadly divided into two large families (classical and extended SDRs) based on the differences in their glycine-rich motifs.[37] Classical SDRs are generally of a shorter chain length (*ca.* 250 amino acids) and contain the characteristic TGxxxGxG

consensus sequence while the extended SDRs contain the TGxxGxxG consensus with a chain length of *ca.* 350 amino acids.[37] Additional sequence-based classification efforts resulted in the identification of five SDR families (classical, extended, intermediate, divergent and complex) and a detailed analysis of the amino acid multiple sequence alignment enabled the sub-classification of the classical and extended SDRs into seven and three distinct subfamilies respectively.[37] The subfamilies were defined based on subtle differences in their sequences related to the cofactor preference (NADH vs NADPH) of each enzyme.[37]

Reductase domains play a central role in the control and introduction of chemical diversity of the final PKS product. They are important for introducing stereogenic centres and introduction of functional groups that may serve as a chemical handle for further diversification. Since structural diversity often translates to biological activity in secondary metabolites, where even the slightest structural modifications may result in significant changes in activity. These enzymes are currently under active study for their potential use in biosynthetic engineering approaches for rational redesign of polyketides produced by these modular biosynthetic systems.

1.5 Thioester Reductases

Thioester reductase (TR) domains catalyse the NAD(P)H-dependent reduction of the thioester bond between the ACP-bound fully assembled polyketide chain in the final module of the multi-enzyme system (**Scheme 13**). This reduction results in the termination of chain assembly and the release of the fully assembled product as an aldehyde which may be further processed by the same TR (to form an alcohol) or be modified by other accessory enzymes (**Scheme 13**).[18,31]



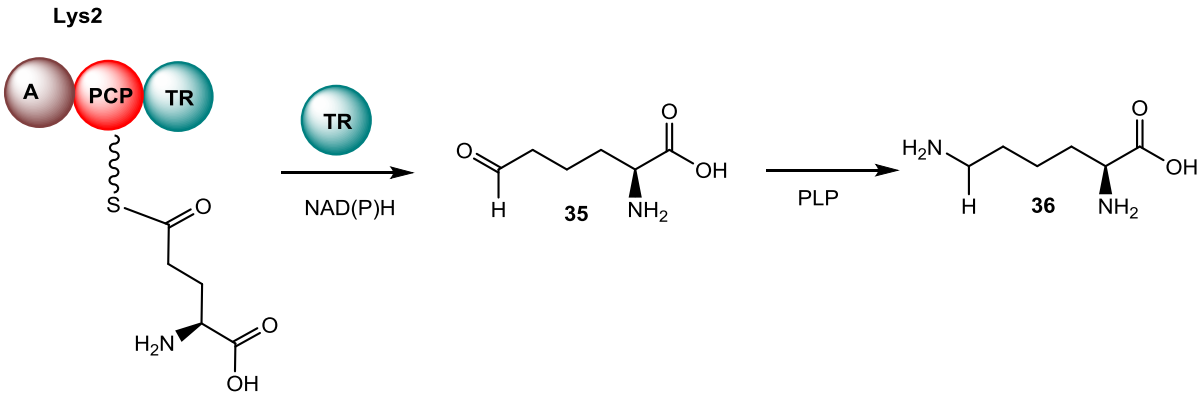
Scheme 13: Mechanism for thioester reductase (TR) mediated NAD(P)H-dependent chain release.[35]

TR domains within PKS / NRPS systems belong to the extended SDR family.[33] A number of reductase domains do not catalyse redox chemistry and have been shown to release the mature polyketide via a Dieckman condensation; examples include the non-redox R domains (R*) involved in the biosynthesis of tetramic acid containing secondary metabolites.[31] Biochemical analyses showed that these R* chain-terminating reductase domains do not bind the co-factor. Comparison of the primary sequence of these enzymatic domains revealed that the catalytic tyrosine residue is not strictly conserved across the entire class of reductases and may be occasionally replaced by a phenylalanine residue. This, together with their inability to bind the reducing co-factor might contribute to the redox inactivity and altered catalytic action observed in R* domains.[31]

Reductase domains have been found embedded within NRPS modules. In certain cases, a stand-alone reductase may be utilized or it may act in concert with an integrated thioester reductase domain to yield the desired final product.[31,35,40] The potential for structural variation in the final products of these secondary metabolite biosynthetic pathways is enormous as the aldehyde or alcohol functional group may be retained in the final product or further processed to yield a range of other functional groups including amines, imines, enol ethers, hemiaminals and a variety of heterocyclic compounds.[31,35] Examples of biosynthetic systems containing TR domains include NRPSs, hybrid NRPS-PKSs and hybrid FAS-NRPS/PKS systems.[35,39,40,41] Prior to commencing this work, TR domains within type I PKS systems had not been biochemically characterized; their presence had however, been detected and their function inferred indirectly by bioinformatics analyses and retrobiosynthetic approaches.

1.5.1 Thioester Reductase in a Primary Metabolic Pathway (Lysine Biosynthesis)

One of the earliest examples where the use of the TR domain was identified in the biosynthesis of lysine (**Scheme 14**) in *Saccharomyces cerevesiae*.[42,43] Studies carried out by Walsh and co-workers showed that the predicted biosynthetic intermediate α–aminoadipate is first of all activated by phosphopantetheinylation and then loaded onto a peptidyl carrier protein (PCP) domain that is part of a multimodular enzyme system.[42] The PCP is adjacent to a reductase domain which catalyzes the reductive cleavage of the α–aminoadipoyl thioester upon addition of NADPH to the enzymatic reaction. The product of reductive cleavage is the α–aminoadipoyl semialdehyde **35**. This intermediate is converted to lysine **36** by reductive amination of the terminal aldehyde catalysed by an aminotransferase.[42,43]



Scheme 14: A TR domain is used for lysine biosynthesis in *S. cerevisiae*.[31]

1.5.2 Thioester Reductases in Non-Ribosomal Peptide Biosynthetic Synthetases

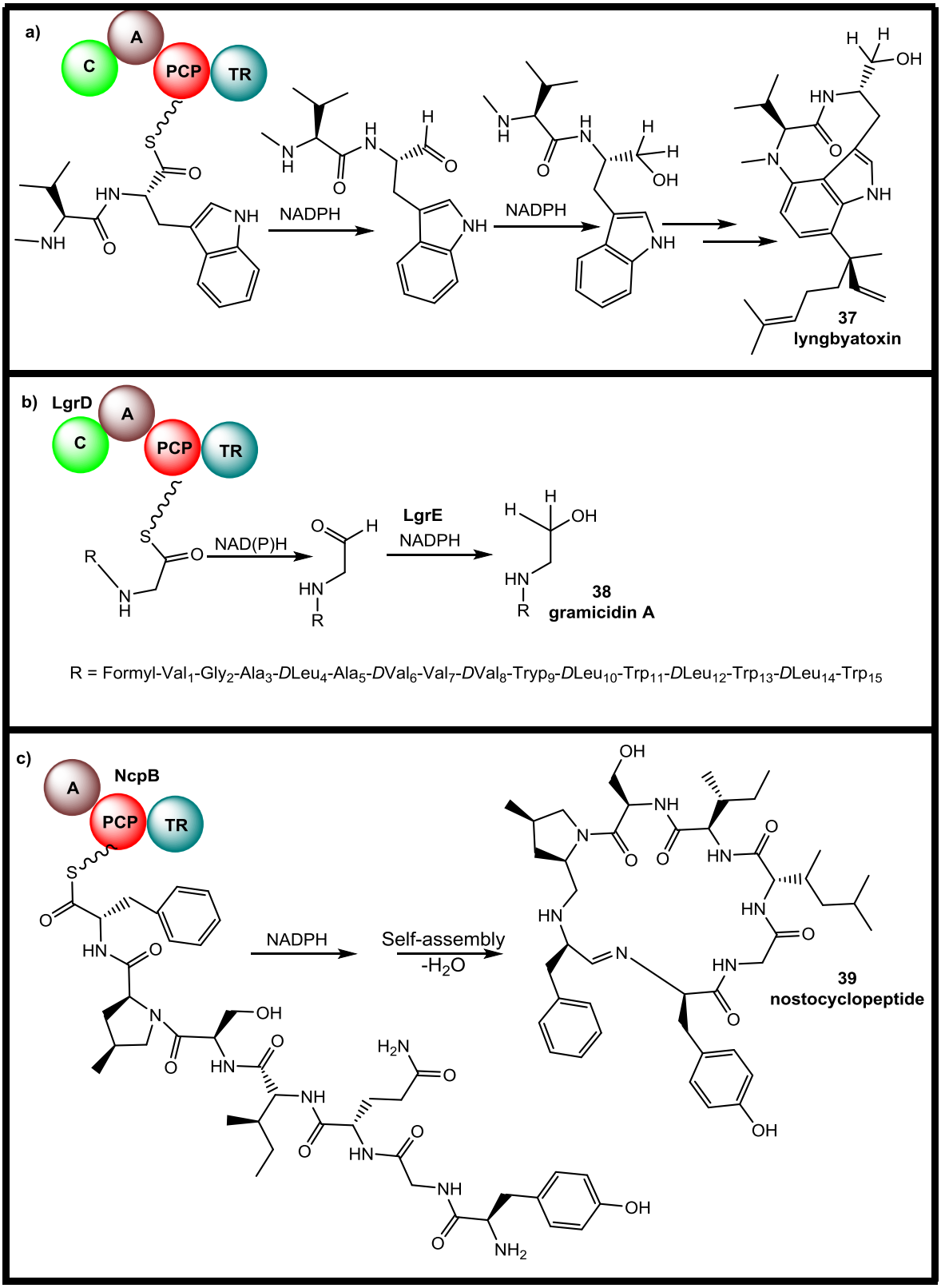
Representative examples of secondary metabolites produced by NRPSs (**Scheme 15**) harbouring a TR domain include the lyngbyatoxins **37**, linear gramicidin **38** and nostocyclopeptide **39**.[44,45,46]

The lyngbyatoxins **37** are potent cyanotoxins produced by the cyanobacterium *Lyngbya majascula.* Their structures comprise of the amino acids N-methyl-*L*-valine and *L*-tryptophan condensed into a dipeptide.[44] A terminal reductase domain (which has been biochemically characterised *in vitro*) catalyses the NADPH-dependent reductive cleavage of the dipeptide intermediate bound to a PCP domain to yield the alcohol.[44] Further tailoring modifications involving a cytochrome P450 and a prenyltransferase yield the final product.[44]

A slightly different mechanism is employed for the biosynthesis of the commercially available antibiotic linear gramicidin **38** (produced by the soil dwelling bacterium *Bacillus brevis*). An NRPS is used to synthesize the linear polypeptide containing 15 hydrophobic amino acids with alternating *L*- and *D*- configurations.[45] The linear gramicidin **38** NRPS (encoded by *lgr*ABCD) contains a TR domain (**Scheme 15b**) appended to the C-terminus of the final module which has been shown to catalyse reductive cleavage of the polypeptide thioester intermediate yielding the aldehyde.[45] An additional reduction step catalysed by a

stand-alone reductase, converts the aldehyde to the corresponding alcohol.[45] The stand-alone reductase is specific for NADPH **29** while the *cis*-acting thioester reductase domain embedded within the gramicidin NRPS, accepts both NADH **31** and NADPH **29**.[45]

Another polypeptide biosynthesized by an NRPS employing a reductive chain releasing mechanism, is the cytotoxin nostocyclopeptide **39** produced by the cyanobacterium *Nostoc* sp. ATCC53789.[42] This NRP is particularly notable because of the macrocyclization strategy employed during its biosynthesis. In the nostocyclopeptide NRPS (*ncp*), however, the reductive release of the NRP as an aldehyde, results in the spontaneous macrocyclization by condensation with an internal primary amine group to form a stable imine bond.[42] The *ncp* TR domain specifically utilizes NADPH **29** and does not accept NADH **31**.[46]



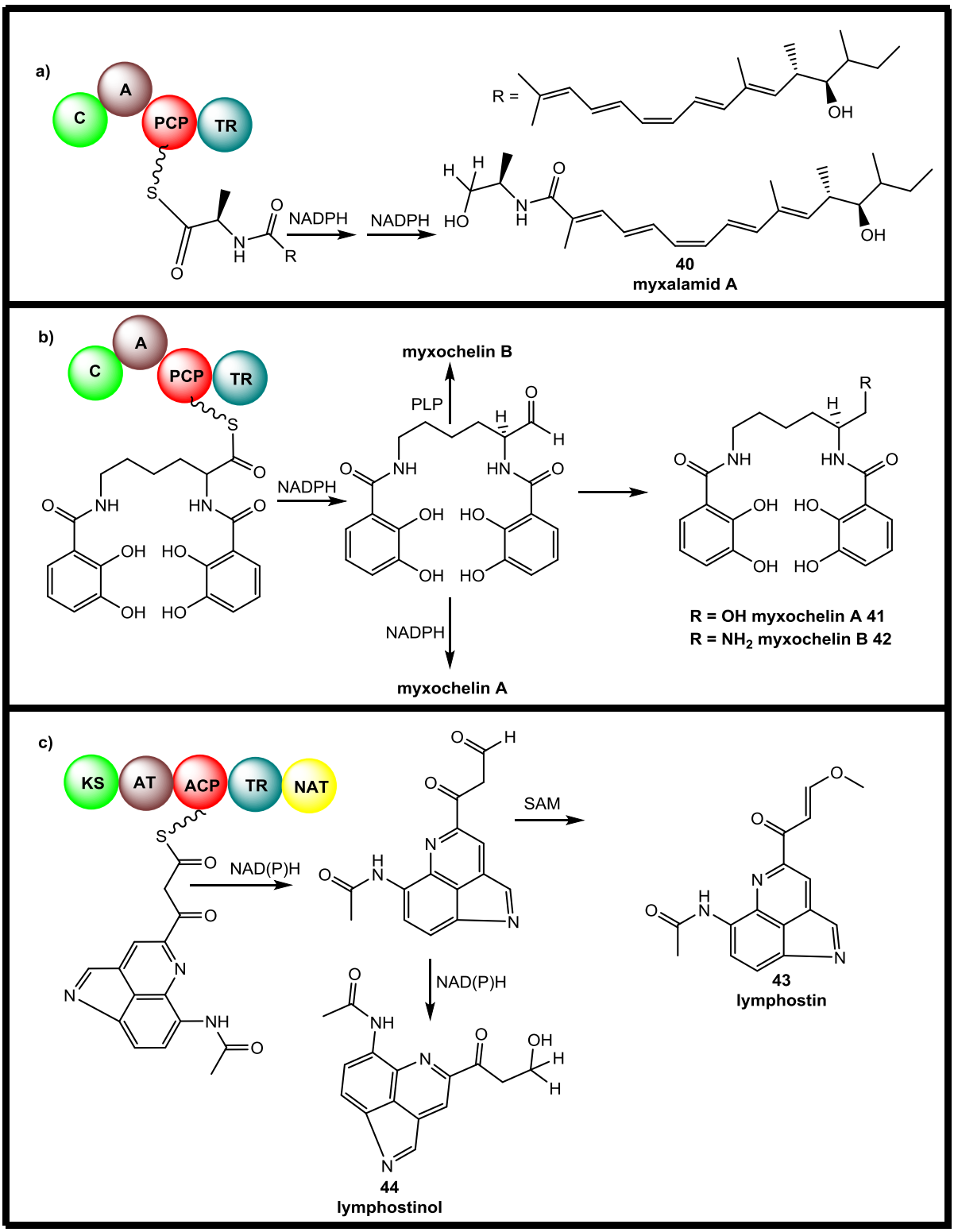
Scheme 15: TR domain-mediated chain release in NRPS biosynthetic systems **a)** lyngbyatoxin **37 b)** linear gramicidin **38 c)** nostocyclopeptide **39**.[44,45,46]

1.5.3 Thioester Reductases in Hybrid NRPS/PKS Biosynthetic Systems

The myxobacterium *Stigmatella aurantiaca* produces both myxalamids and myxochelins.[38,48] Myxalamids are produced by an octamodular hybrid type I PKS / NRPS with a reductase domain located at the C-terminus which utilizes 2-methylbutyrate (or isobutyrate or propionate or acetate) as a starter unit and methylmalonate (or malonate) for chain extension.[38,48] The fully assembled polyketide thioester intermediate is transferred to the terminating NRPS module which extends it by the attachment of an alanine residue prior to reductive cleavage. The TR domain in this PKS / NRPS utilizes NADPH to catalyse a two-step reduction to yield myxalamid A **40** bearing a primary alcohol at its terminus (**Scheme 16**).[38]

The myxochelin biosynthetic gene cluster contains genes encoding the biosynthesis of the 2,3-dihydroxybenzoic acid (DHBA) intermediate which is covalently attached to the aryl carrier protein (ArCP); the NRPS condensation domain catalyzes two successive condensations with the two amino groups of lysine (**Scheme 16b**).[49] The TR domain catalyzes reductive cleavage to yield the aldehyde which may be reduced further to the corresponding alcohol myxochelin A **41** or reductively aminated yielding the corresponding amine myxochelin B **42**.[49]

Lymphostin **43**, an immunosuppressant produced by the marine actinomycete *Salinispora arenicola* is assembled by a bimodular heptadomain NRPS / PKS. The NRPS module incorporates a tryptophan-derived starter unit predicted to contain the pyrrolo [4.3.2-*de*] core.[41] The PKS module (**Scheme 16c**) extends the carbon chain with a malonate prior to release of the resulting intermediate by a TR domain to yield the aldehyde.[41] An N-acetyltransferase (NAT) is predicted to catalyse an acetylation reaction at some point during chain assembly.[41] The aldehyde is converted to lymphostin **43** by a methyltransferase or reduced further to yield lymphostinol **44**.[41]

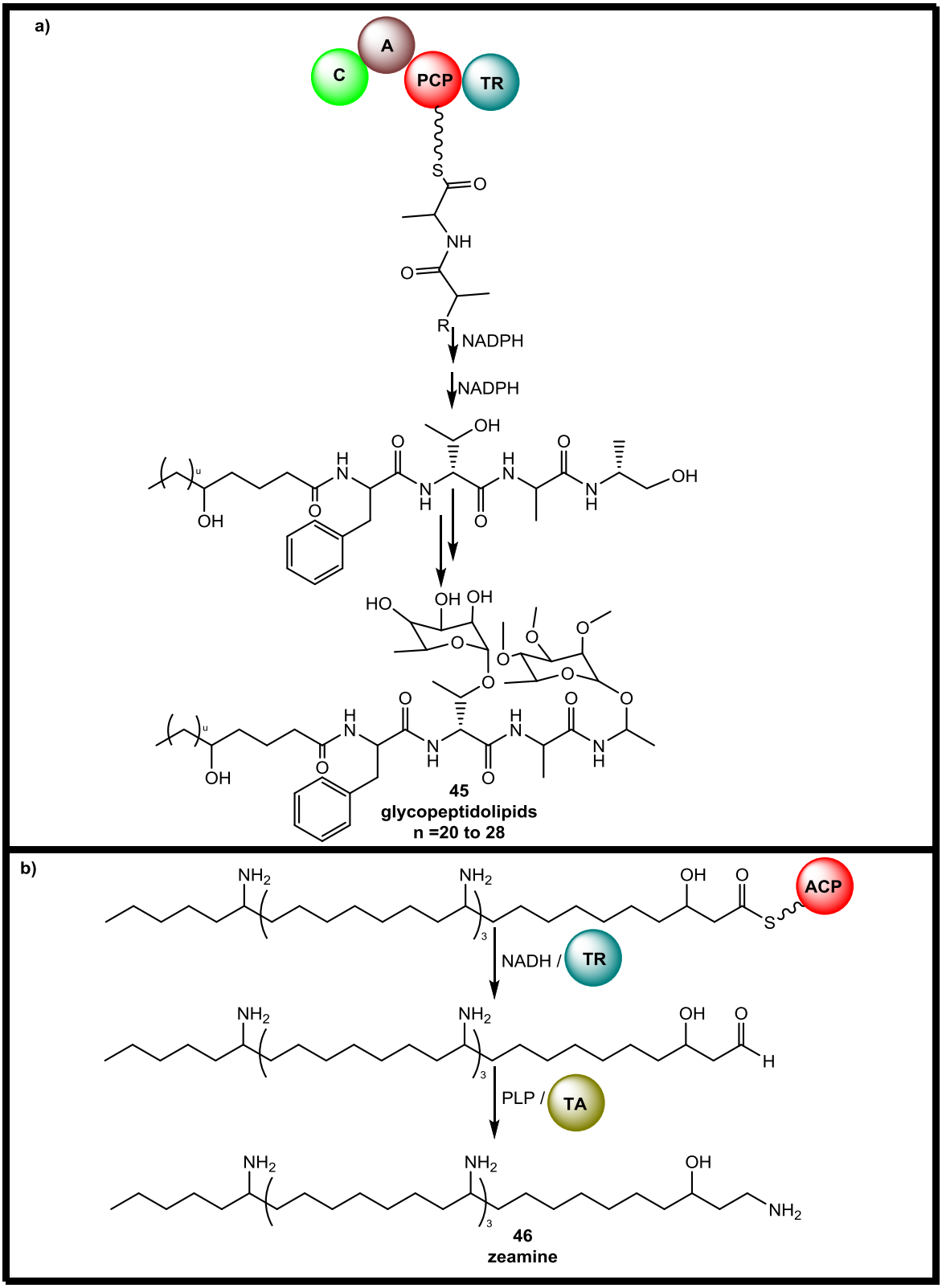


Scheme 16: TR domain mediated chain release in hybrid PKS / NRPS biosynthetic systems **a)** myxalamid A **40 b)** myxochelins **41, 42 c)** lymphostin **43** (and lymphostinol **44**).[38,41,49]

1.5.4 Thioester Reductases in Hybrid FAS-NRPS/PKS Biosynthetic Systems

Glycopeptidolipids **45** are modified peptides attached to a long fatty acyl chain and carbohydrate residues that are vital for mycobacterial survival and pathogenesis.[39] The biosynthetic gene clusters for glycopeptidolipids (GPLs) in *M. smegmatis* and *M. tuberculosis* comprise of an NRPS which creates the tetrapeptide (phenylalanine, threonine, and two alanines) and condenses it with a 3-hydroxy-fatty acyl chain. This NRPS contains a terminal TR domain which cleaves the thioester intermediate to yield the corresponding alcohol (**Scheme 17**).[39] The resulting hydroxyl group is the point of covalent attachment of one of the carbohydrate residues by post-NRPS tailoring enzymes.[39,50] This TR domain utilizes NADPH as a reducing co-factor.[39]

The zeamines **46** are broad spectrum antibiotics and phytotoxins produced by *Serratia plymuthica* RVH1. Possessing a unique chemical scaffold of a polyaminoalcohol chain attached to a polyketide chain linked to a valine residue.[40] A combination of three different classes of biosynthetic machinery are used for the assembly of the zeamines: an iterative type I polyunsaturated fatty acid synthase (PUFAS); and a hybrid NRPS / PKS system.[40] The products of the PUFAS and the PKS / NRPS are linked together by the action of a stand-alone condensation domain catalysing the formation of an amide bond between the terminal primary amine produced by the PUFAS and the thioester-bound PK / NRP intermediate.[40] Further processing involving the cleavage of the pentapeptide terminus yields the final product.[40] A stand-alone TR domain cleaves the thioester intermediate bound to the PUFAS (**Scheme 17**) to yield the corresponding aldehyde.[40] This biosynthetic intermediate is further processed to the corresponding primary amine.[40] The stand-alone TR has been biochemically characterized, and shown to be specific for NADH **31** in contrast to the gramicidin stand-alone thioester reductase which is specific for NADPH **29** (discussed in section **1.4.2.2** above).[40,45]

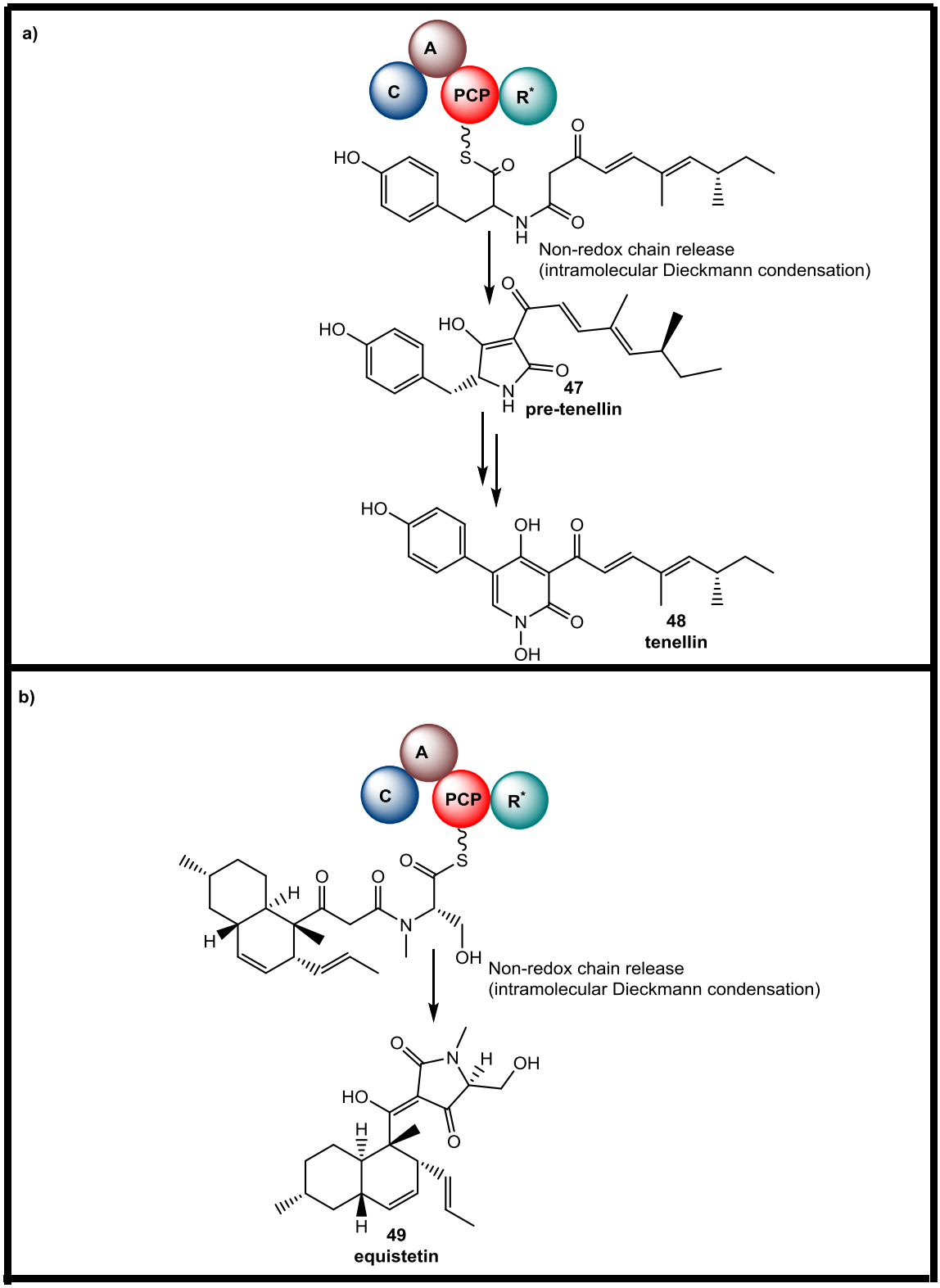


Scheme 17: TR domains in Hybrid FAS-NRPS/PKS Systems **a)** glycopeptidolipds, **45 b)** zeamine **46**.[39,40]

1.5.5 Thioester Reductases in NAD(P)H-Independent Non-Redox Biosynthetic Systems

The tetramic acid-containing (pyrrolidine-2,4-dione) secondary metabolites tenellin **48** and equistetin **49** are the products of hybrid PKS/ NRPS biosynthetic systems that utilize unusual TR-type domains (R*) which catalyse non-redox chain release of the thioester intermediate. Tenellin **48** is a cytotoxic and antifungal agent isolated from the fungal insect pathogen *Beaveria bassiana*.[31] It is produced by a hybrid iterative type I PKS / NRPS (TENS) with an R* domain fused to its C-terminus (**Scheme 18**).[31,51] The NRPS module incorporates a tyrosine or phenylalanine residue.[31,35,51] The R* domain has been shown to catalyse a Dieckmann condensation by *in vivo* studies; resulting in the formation of the 5-membered tetramic acid ring in pretenellin **47** prior to further tailoring steps to yield the final product: tenellin **48**.[31,35,51]

Equistetin **49** is produced by the fungal plant pathogen *Fusarium equiseti*. It inhibits HIV-1 integrase and has been used as a model during the development of the drug raltegavin.[52] The equistetin synthase (EqiS) is also a hybrid iterative type I PKS/NRPS with an R* domain fused to the C-terminus of the NRPS. Unlike the tenellin synthase, the EqiS NRPS incorporates a serine residue.[31,35,52] The EqiS-R* domain was shown to catalyzes a Dieckmann condensation to yield the corresponding tetramic acid.[31,35,52] Its enzymatic activity was analysed using synthetic substrate analogues revealing that EqiS-R* does not carry out redox chemistry; instead, the reaction was shown to be NADPH-independent.[31,35,52]



Scheme 18: NAD(P)H-Independent chain release catalysed by R* domains in **a)** tenellin **48** **b)** equistetin **49** biosynthesis.[51,52]

1.6 Tailoring Steps in Polyketide Biosynthesis

Several possibilities exist for further structural diversification following the release of the fully assembled polyketide chain from the PKS. The genes encoding these enzymes are often found within the vicinity of the PKS genes and are usually expressed at the same time. This makes it possible to predict the final product of the PKS (following post-PKS) modifications in certain cases.

Post-PKS tailoring enzymes have been discovered that possess numerous different catalytic activities; including: cyclases, enzymes for deoxysugar biosynthesis and attachment, isomerases, halogenases, oxido-reductases and group transferases (such as methyl transferases and pyridoxal-5'-phosphate-dependent aminotransferases).[53]

1.6.1 Pyridoxal-5'-Phosphate – Dependent Enzymes

Vitamin B_6 is essential to organisms in all four domains of life. Pyridoxal 5'-phosphate **50** is the main metabolically active form of this vitamin; it mediates a wide-range of reactions both in its free state in solution, and in association with various enzymes.[54,55,56] Examples of these biochemical transformations include transamination, decarboxylation, racemization, eliminations and substitutions.[55] Consequently, enzymes that utilize this co-factor (**Fig. 5**) have been found to play major roles in both primary and secondary metabolism of all living organisms.[55,56]

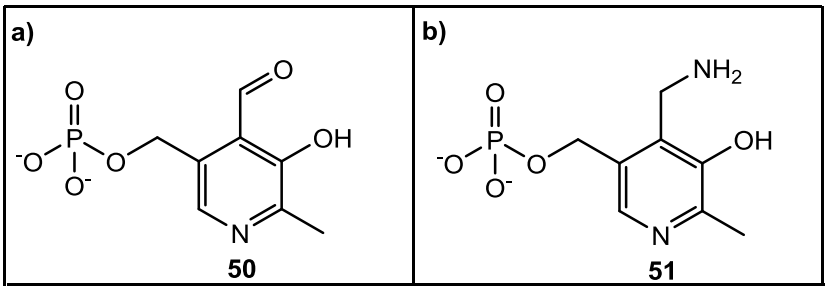


Figure 5: Structures of **a)** pyridoxal-5'-phosphate (PLP) **50** and **b)** pyridoxamine-5'-phosphate **51**.[55]

1.6.1.1 Key structural features of PLP

There are five functional groups present in PLP **50**; a methyl group, a pyridine nitrogen atom, a phosphate group, a phenolic hydroxyl group and an aldehyde group (**Figure 5**). The aldehyde group is the point of covalent attachment to the enzyme; its position on the aromatic ring of the co-factor is also central to the catalytic mechanism of PLP.[56] The pyridine nitrogen is also required for catalysis. Variations in the amino acid residues in the active sites of PLP-dependent enzymes, interacting with the pyridine nitrogen atom determine its protonation state and consequently, the reactions catalysed by these different enzymes.[57,58,59,60] A conserved aspartate (or glutamate) residue in aminotransferases interacts with the pyridine nitrogen in order to maintain it in its protonated state. While, an arginine residue in racemases, interacts with the pyridine nitrogen leaving it essentially unprotonated and changing the course of the reaction.[59] The phenolic group interacts with enzyme residues (via hydrogen bonding) to control the reactivity of PLP **50**.[57] In non-enzymatic transamination and elimination, conversion of the phenolic hydroxyl group to its *O*-methyl equivalent, suppresses the reaction, showing that it is essential for catalysis.[56] The phosphate and methyl groups are non-essential for catalysis. However, an additional point of interaction with the enzyme is provided by the phosphate group which

binds via electrostatic interactions with enzyme residues in the active site that make up the phosphate binding cup[55,61]

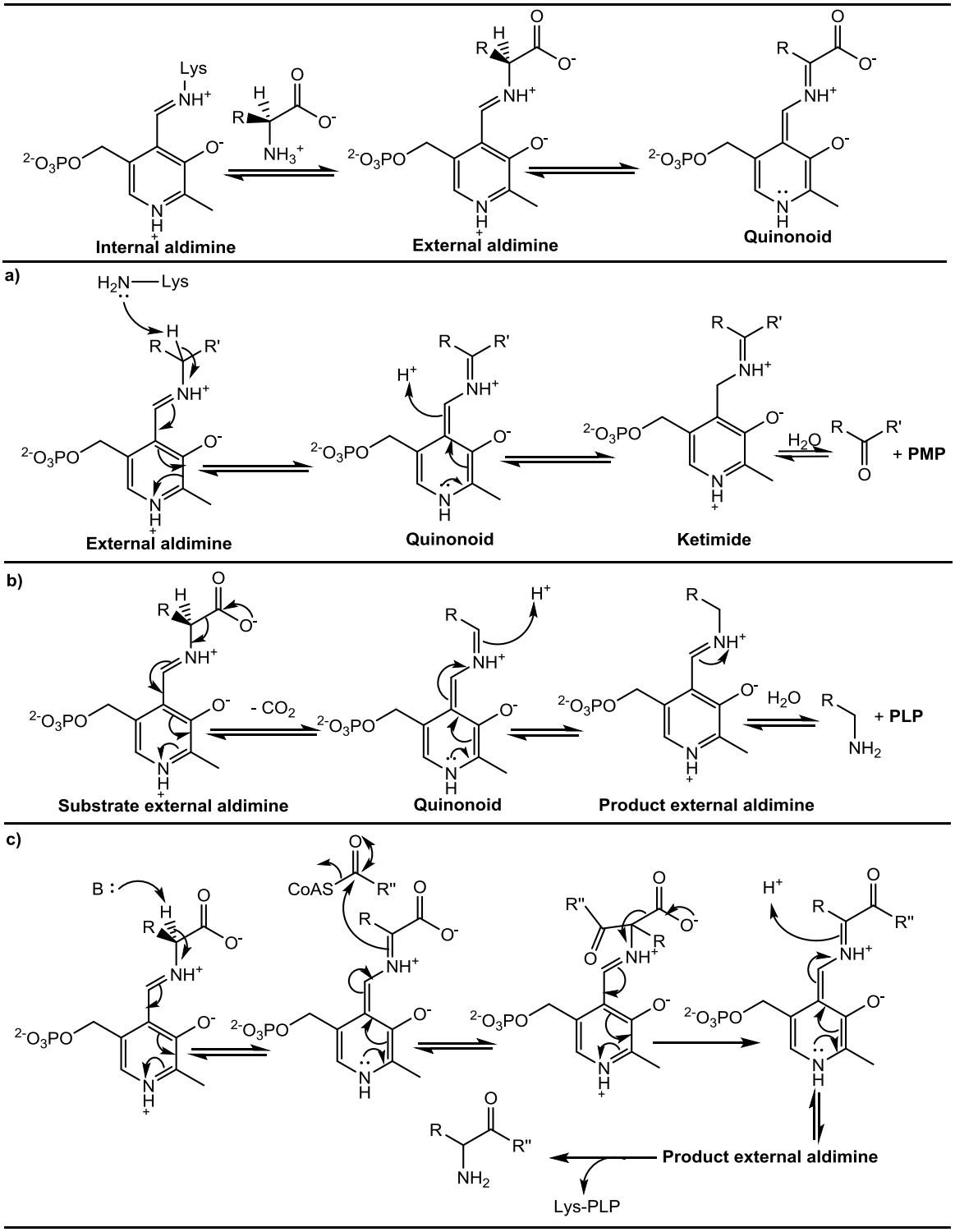
1.6.1.2 Mechanism of catalysis

Most of the enzymatic reactions involving PLP as a cofactor take place at the α, β, or γ positions in the amino acid substrate.[54] Reactions occurring at the α-position include transamination, decarboxylation, racemization, elimination and substitution; while at the β and γ positions, the major reactions are elimination and substitution.[54]

In its resting state, the enzyme is bound to the cofactor by forming an imine linkage between the aldehyde group of the cofactor and the Ɛ-amino group of a strictly conserved lysine residue, the enzyme-cofactor complex is generally known as the internal aldimine.[54,55] The substrate amino donor displaces the lysine residue by transimination to form the external aldimine intermediate and this is the common point from which the various possible reactions proceed.[54]

In the case of transamination (**Scheme 19a**), removal of a proton from the α position by a catalytic base drives the formation of the quinonoid intermediate. While in decarboxylation, the reaction begins with loss of CO_2 from the external aldimine which also results in the formation of a quinonoid intermediate.[54] Each enzyme strictly determines the reaction outcome (**Scheme 19**) by controlling the initial bond breaking step via structural conformational control and intricate interactions with active site residues that govern electron flow and proton transfers.

Electron flow in the reverse direction (beginning from the nitrogen lone pair in the quinonoid intermediate) along with reprotonation at the C-4' position (in transamination) results in the formation of the ketimine intermediate which undergoes hydrolysis to yield the product.[54]

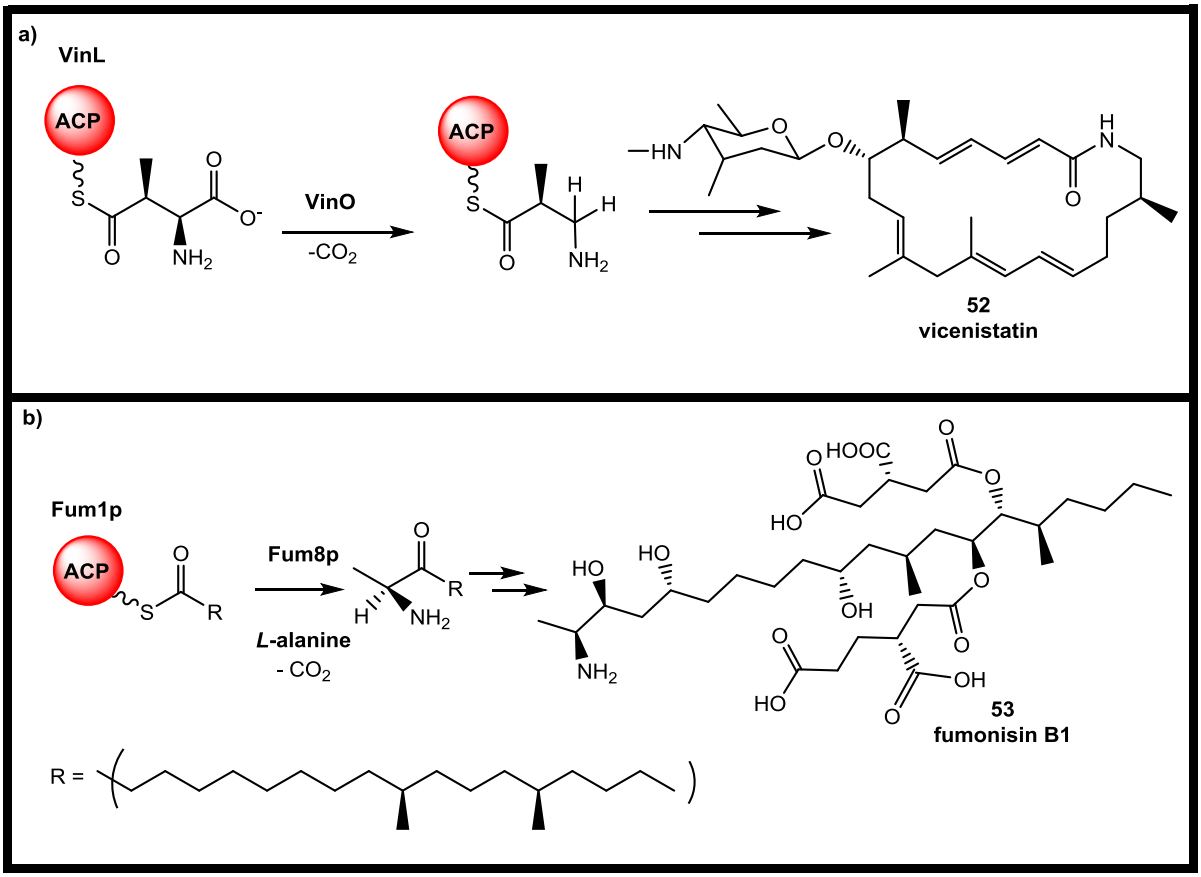


Scheme 19: General mechanism of catalysis of PLP-dependent enzymes showing possible routes for reactions at the α-position a) transamination b) elimination (decarboxylation) c) decarboxylative Claisen condensation.

1.6.1.3 Pyridoxal Phosphate-Dependent Transformations in Polyketide and Nonribosomal Peptide Biosynthesis

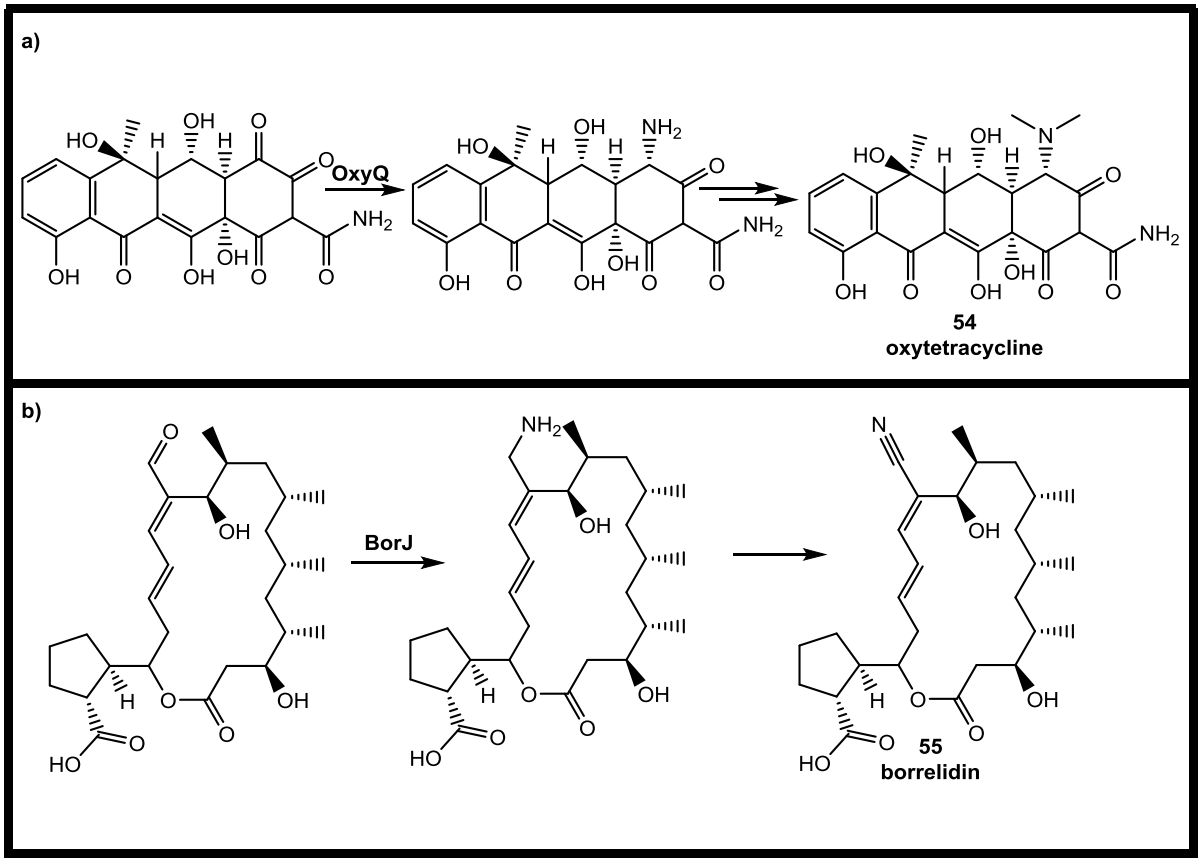
Within the context of the polyketide and nonribosomal peptide biosynthetic systems, several PLP-dependent enzymes and enzymatic domains have been identified and characterised. Some of the more common reactions catalysed by these include decarboxylation, carbon-carbon bond formation and transamination. A few examples of enzymes catalysing PLP-mediated decarboxylation have been identified within hybrid polyketide-nonribosomal peptide biosynthetic pathways. For example in the biosynthesis of the macrolactam vicenistatin **52** shown in **Scheme 20**.[62] A β-methyl-*iso*-aspartyl thioester is decarboxylated during an early step in the biosynthetic pathway yielding the corresponding α-methyl β-alaninyl thioester which serves as the starter unit for polyketide chain assembly and ultimately enable chain release via macrolactamisation.[62,63]

The biosynthesis of the mycotoxin fumonisin **53** produced by the plant pathogen *Fusarium verticilliodes*; features an unusual thioester chain release mechanism involving a decarboxylative condensation between a thioester and *L*-alanine catalysed by an α-oxoamine synthase; this results in termination of the polyketide chain (**Scheme 20**) and formation of a new carbon-carbon bond (introduction of an aminomethyl group).[31] A similar chain release mechanism has been described in **section 1.3.4** for the biosynthesis of undecylprodigiosin (**Scheme 11**).[34]



Scheme 20: PLP-catalyzed decarboxylation in **a)** vicenistatin **52** and PLP-dependent chain release via decarboxylative Claisen condensation catalysed by an AONS in **b)** fumonisin **53** biosynthesis.[31,62]

PLP-mediated transamination in polyketide assembly features in the biosynthesis of oxytetracycline **54**.[59] A ketone group in a biosynthetic intermediate is converted into a secondary amine (**Scheme 21a)** prior to dimethylation which results in the formation of oxytetracycline **54**.[64]



Scheme 21: PLP-mediated transamination in **a)** oxytetracycline **54** and **b)** borrelidin **55** biosynthesis.[64,65]

The biosynthesis of the polyketide macrolide borrelidin **55**, also involves a PLP-dependent transaminase (**Scheme 21b**); BorJ converts an aldehyde-containing biosynthetic intermediate to the corresponding amine which is oxidized to give the nitrile group in borrelidin.[65]

Other examples of PLP-mediated transamination reactions in secondary metabolite biosynthesis have been discussed briefly above, e.g. myxochelin **41** and the zeamines **46** (sections **1.4.3.3** and **1.4.4.2** respectively).[40,49]

More recently, transaminases have been implicated in polyketide alkaloid biosynthesis. Several bioactive mono-, di-, tri- and tetracyclic alkaloids produced by

microbes (very often isolated from Actinobacteria), have been reported to be of polyketide origin. However, direct biochemical evidence for the involvement of such enzymes in nitrogen incorporation was lacking at the point this work was initiated.

Nitrogen heterocycles are key structural features of most alkaloids.[66] Many of the widely known piperidine and pyrollidine alkaloids including nicotine, coniine and cocaine contain piperidine and pyrollidine rings which are biosynthetically derived from *L*-lysine and *L*-ornithine respectively.[66] In these cases, the entire carbon backbone plus the nitrogen atom of an amino acid precursor is incorporated into the heterocycle. However, as is frequently observed in nature, different biosynthetic routes have evolved to achieve similar outcomes. An alternative route to the piperidine and pyrollidine rings in alkaloids of polyketide origin isolated from Actinobacteria involves transamination, where an amino acid contributes only its nitrogen and not its carbon skeleton to formation of the heterocycle.

1.7 Study of Secondary Metabolite Biosynthesis in *Streptomyces coelicolor*

Actinobacteria are known to be prolific producers of most known structural classes of secondary metabolites found in Nature.[67] The *Streptomyces* genus in particular has been studied in great detail due to the significant number of industrially-relevant metabolites it produces. For over half a century, *Streptomyces coelicolor* A3(2) has been identified and used as a model organism for the characterization of biosynthetic mechanisms used to assemble and control the production of various secondary metabolites.[67,68] The name ‘coelicolor’ was given to this *Streptomyces* species because of its characteristic blue colour which it derives from the actinorhodin complex of pigmented antibiotics. The first example of rational biosynthetic engineering to produce novel secondary metabolites was reported by Hopwood and co-workers during their pioneering research on the molecular genetics for the biosynthesis of actinorhodin **13**.[68,69]

The *S. coelicolor* M145 strain is a derivative of the A3(2) strain lacking its two plasmids - the giant linear and small circular plasmids: SCP1 and SCP2 respectively.[68] Upon completion of the genome sequence of *Streptomyces coelicolor* M145 in 2002, it became apparent that this model actinomycete is capable of producing at least 17 different structural classes of secondary metabolites.[68] Detailed analysis of the genome sequence led to the rational design of targeted experiments for the isolation and characterization of many of these metabolites (including novel compounds), and the discovery of previously undescribed biosynthetic mechanisms (in desferrioxamine **65** and coelichelin **69** biosynthesis).[68] Prior to genome sequencing studies *S. coelicolor* had been known to produce the antibiotic methylenomycin A **57** (encoded by the linear plasmid, SCP1); the calcium dependent antibiotics **60** (CDA2a and CDA2b); desferrioxamine E **65** (a siderophore); and hopanoids (e.g aminotrihydroxybacteriohopane **68**). In addition to the blue pigment (actinorhodin **13**), a red pigment (comprising of undecylprodigiosin **30** and other unidentified compounds) and a yellow pigment (of unidentified composition) had also been reported.[70]

The availability of the complete genome sequence enabled the discovery of various other compounds belonging to different structural classes as metabolites of *S. coelicolor* including polyketides (flaviolin **64** and coelimycin P1 **61**), non-ribosomal peptides (e.g coelichelin **69**) and terpenoids (albaflavenone **63**, and 2-methylisoborneol **58**), and carotenoids (isorenieratene **70**).[68] The second component of the red pigment complex was also identified as streptorubin B **56** a carbocyclic derivative of undecylprodigiosin **30**. During the course of time, considerable efforts have been invested into isolating these metabolites and studying the mechanisms by which they are biosynthesized resulting in significant insight into the chemistry and the biological roles played by these intriguing compounds.[68] The molecular structures and biosynthetic mechanisms for the blue and red pigments are now well characterized but equivalent characterisation of the yellow pigment remained elusive for years.[7]

1.8 Coelimycin

1.8.1 Isolation and Characterisation of Coelimycin P1

In the late 1970s, *S. coelicolor* was reported to produce a diffusible yellow-orange pigment with antibiotic activity.[7] Genetic mapping during this study indicated that the genes associated with the production of this compound were grouped in a cluster subsequently named of the *cpk* gene cluster (fully identified upon completion of the genome sequence in 2002).[7,67] Studies in 2010 also reported the production of a yellow pigment as a product of the *cpk* gene cluster.[7,70]

In 2012, a rational genetic engineering strategy was employed to produce and structurally characterise a yellow metabolic product of the *cpk* gene cluster.

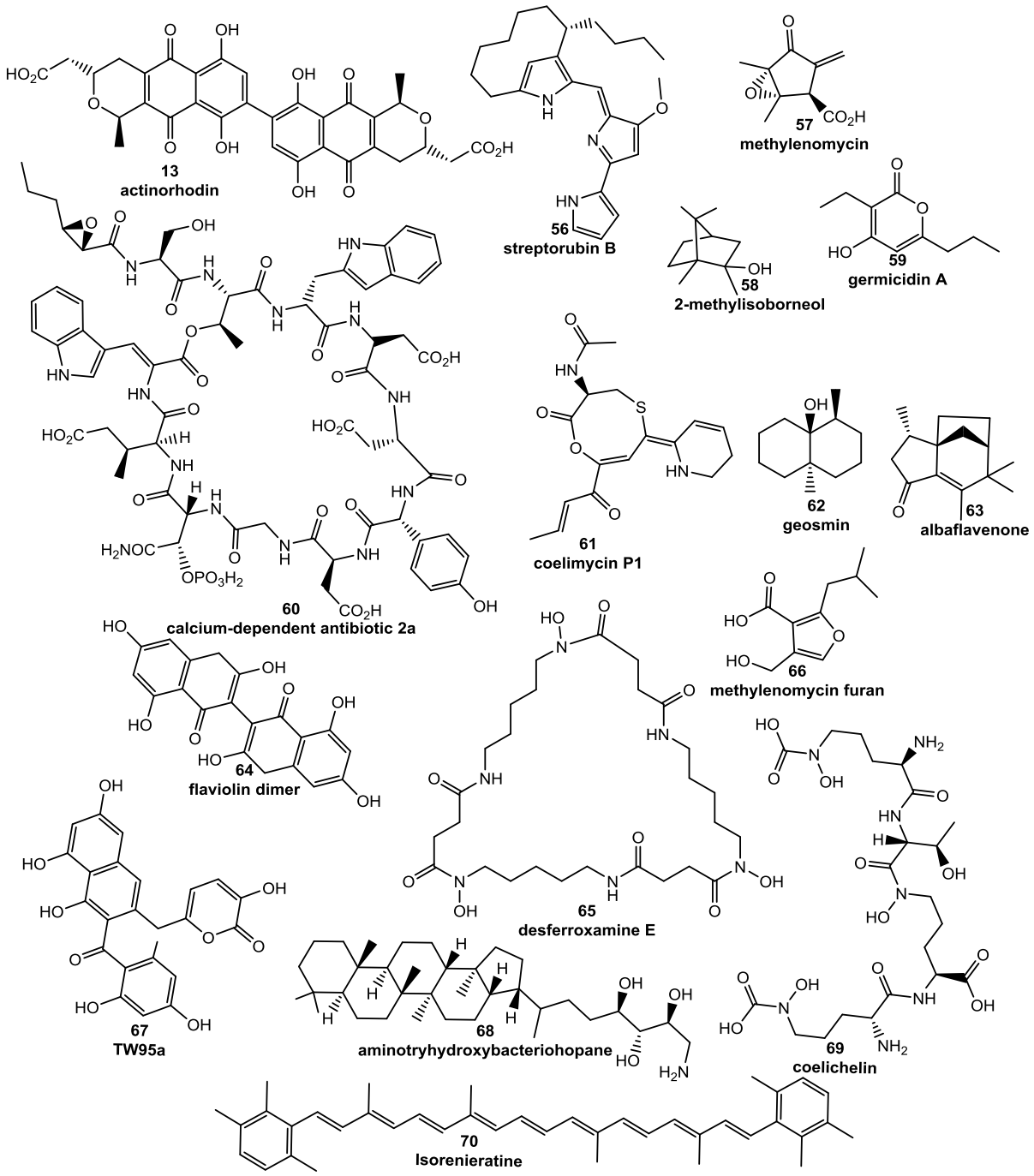


Figure 6: Examples of metabolites isolated from *Streptomyces coelicolor*.[68]

The strain *S. coelicolor* M1148 was constructed from the sequenced M145 strain by deleting genes responsible for the production of other secondary metabolites (actinorhodin, prodiginines and calcium-dependent antibiotics) to reduce competition for common biosynthetic precursors and to increase metabolic flux through the biosynthetic pathway of

interest; and provide a cleaner metabolic background for easier analysis[7] (this strain also does not produce methylenomycin due to the absence of the SCP1 plasmid). A specific mutation into the β-subunit of the RNA polymerase gene *rpoB* was introduced (H437R) by selecting for spontaneous rifampicin resistance; this mutation had been previously reported to increase levels of secondary metabolite production.[7] This strain was then further modified to delete a section of the *cpk* cluster and the production profile of the two strains was compared, leading to the identification of coelimycin P1 **61**.[7] Coelimycin P1 **61** is a unique functionalized 1,5-oxathiocane (**Figure 6**), corresponding to the N-acetylcysteamine adduct of the predicted metabolic product of the *cpk* gene cluster.[7]

1.8.2 Coelimycin Biosynthetic Gene Cluster

The coelimycin biosynthetic gene cluster (**Figure 7**), consists of 24 genes spanning *ca.* 58 kb.[7] The first group of genes encode proteins involved in transcriptional regulation (*scbR, scbA, scbB, cpkM, cpkO, scbR2* and *cpkN*). These include enzymes involved in γ-butyrolactone (GBL) biosynthesis (*scbA* and *scbB*); a TetR-like GBL receptors (*scbR*); a repressor protein (*scbR2*) a two-component histidine kinase (*cpkM*); and *Streptomyces* antibiotic regulatory protein (SARP)-family transcriptional activators *cpkO* and c*pkN*.[7]

Four genes encode enzymes predicted to be involved in precursor supply: α-ketoacid-dependent ferredoxin reductase α and β subunits (c*pkPα* and *cpkPβ*) and acyl-CoA carboxylase α and β subunits (*accA1* and *cpkK*).[7]

A further four genes encode components of a type I modular PKS (*cpkA, cpkB* and *cpkC* encode a loading module and chain extension modules 1-5), *scoT* encodes a type II TE. Several genes encode putative tailoring enzymes including flavin-dependent epoxidases / dehydrogenases (*scF, cpkD, cpkH*), an isomerase (*cpkE*), PLP-dependent aminotransferase (*cpkG*) and a nicotinamide-dependent dehydrogenase (*cpkI*). One gene

encodes a transmembrane efflux protein (*cpkF*) and two other genes *cpkJ* and *cpkL* comprise of hypothetical proteins of unknown function.[7]

- ScbR (butyrolactone-responsive repressor protein).
- ScbA (butenolide synthase).
- ScbB (reductase).
- CpkM (two-component system hystidine kinase).
- CpkO (SARP-family transcriptional activator).
- ScbR2 (Repressor protein).
- CpkN (SARP-family transcriptional activator).

- CpkPα (α-ketoacid-dependent ferredoxin reductase α-subunit).
- CpkPβ (α-ketoacid-dependent ferredoxin reductase β-subunit).
- AccA1 (acyl-CoA carboxylase α-subunit).
- CpkK (acyl-CoA carboxylase β-subunit).

- ScF (secreted flavin-dependent epoxidase / dehydrogenase).
- CpkD (secreted flavin-dependent epoxidase / dehydrogenase).
- CpkE (isomerise α,β-hydrolase fold).
- CpkG (pyridoxal-dependent aminotransferase).
- CpkH (secreted flavin-dependent epoxidase / dehydrogenase).
- CpkI (Nicotinamide-dependent dehydrogenase).

- CpkA (polyketide synthase loading module, module1 and module 2).
- CpkB (polyketide synthase module 3 and module 4).
- CpkC (polyketide synthase module 5).
- ScoT (type II thioesterase).

- CpkF (transmembrane efflux protein).

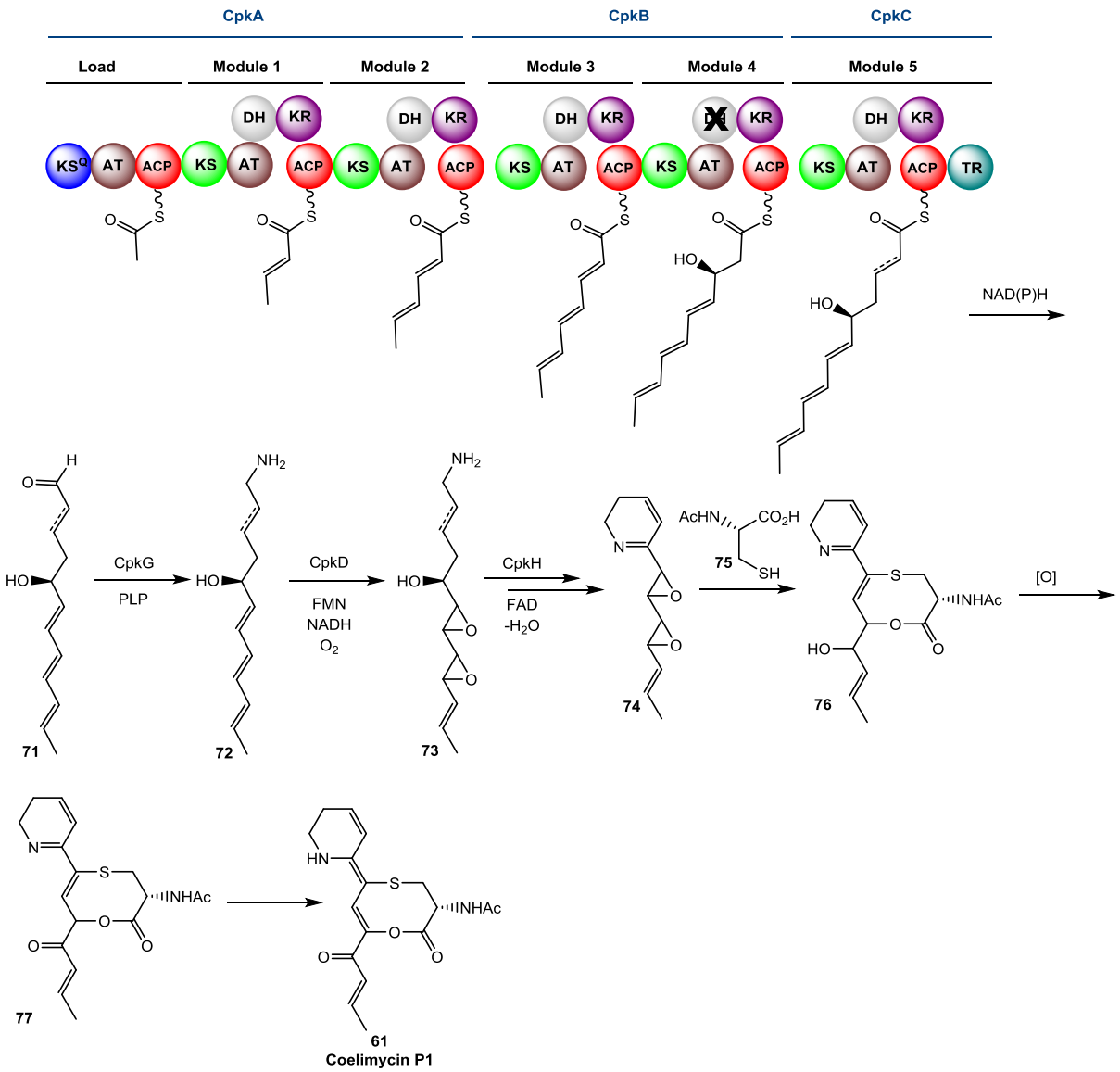
- CpkJ (NmrA-family protein).
- CpkL (hypothetical protein of unknown function).

Figure 7: Organization of the *cpk* gene cluster. Colour codes classify encoded enzymes by their putative function based on bioinformatics analysis.[7]

1.8.3 Proposed Biosynthesis of Coelimycin

The carbon skeleton of the western portion of coelimycin P1 **61** is assembled by a hexamodular type I PKS (CpkABC) (**Scheme 22**). The loading module comprises of a KS^{Q}-AT-ACP tridomain, which loads and decarboxylates a malonyl group resulting in an acetyl starter unit for chain assembly.[7] The starter unit is predicted to undergo five rounds of chain extension via Claisen-type condensations with 5 malonyl extender units. The first three rounds of chain extension are each followed by ketoreduction and dehydration due to the presence of KR and DH domains but not an ER domain in modules 1-3.[7] The DH domain in the fourth extension module appears to be inactive, hence the presence of the hydroxyl group at the C5 position of the fully assembled polyketide chain.[7] The fully assembled polyketide chain is predicted to contain a *trans*-2,3 double bond.

Reduction of the thioester by the NADH-dependent TR domain fused to the C-terminus of module 5 is proposed to yield the corresponding aldehyde **71**. A PLP-dependent aminotransferase (CpkG) is predicted to convert the aldehyde to the corresponding amine **72** with glutamate as a probable nitrogen donor.[7] CpkD a flavin-dependent epoxidase is hypothesized to catalyse the double epoxidation of the intermediate **72** yielding **73**.[71] A flavin-dependent dehydrogenase CpkH is proposed to catalyse the oxidation of the C5 hydroxyl, resulting in spontaneous intramolecular imine formation, yielding the piperideine **74**. Tandem nucleophilic addition of N-acetylcysteine to the epoxides is proposed to lead to the formation of the oxathiocane **76**.[7] A final oxidation of a hydroxyl group catalysed by an unidentified dehydrogenase, is proposed to precede tautomerization of the double bonds to form the conjugated ene-triene-one system in coelimycin P1 **61**.[7,71]



Scheme 22: Proposed biosynthetic pathway for coelimycin P1 showing the organization of the CpkABC type I modular PKS bearing each of its ACP-bound biosynthetic intermediates and post-PKS tailoring reactions.[7,71]

1.9 Objectives and Scope of Work

This project aimed to decipher the post-PKS biosynthetic steps of coelimycin P1 biosynthesis. Of particular interest was the proposed mechanism for release of the fully assembled polyketide chain from the PKS and the subsequent incorporation of a nitrogen atom leading ultimately to formation of a piperidine.

- The first objective was to clone and overproduce the region of *cpk*C encoding the thioester reductase (TR) domain; and to investigate the activity of the resulting purified recombinant protein with various analogues of the natural substrate.
- The second objective was to determine the co-factor preference (either NADH or NADPH) of the TR domain and determine the stereospecificity of hydride transfer from the reduced nicotinamide.
- The final objective was to overproduce, purify and characterize the PLP-dependent aminotransferase CpkG, to provide insight into the mechanism for nitrogen incorporation into polyketide-alkaloids.

CHAPTER TWO

RESULTS AND DISCUSSION I

2.0 NADH-DEPENDENT THIOESTER REDUCTASE MEDIATED CHAIN RELEASE

2.1 Sequence Alignment

The catalytic domain boundaries within the CpkC module were determined by sequence analysis with PKS / NRPS analysis software in order to identify the region corresponding to the TR domain.[72] A sequence alignment of the coelimycin thioester reductase (CpkC-TR) and other reductase domains involved in processing the carrier protein-bound thioester intermediate of NRPSs (LymA-R[41], LtxA-R[44], NcpB-R[46]) a hybrid PKS / NRPS (MyxaA-R)[38], a catecholate siderophore synthase / NRPS (MxCG-R)[49] a glycopeptidolipid (GPL-R[39]) and NAD(P)H-independent reductases (TenS-R[51] and EquiS-R[52]) was performed in Clustal Omega.[73] The results are shown in **Figures 8** and **9**.

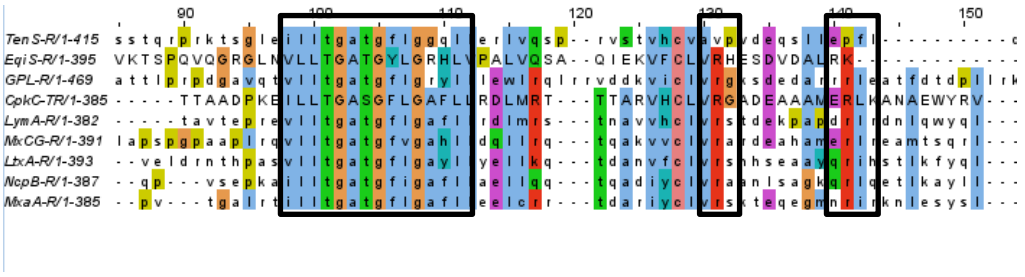


Figure 8: Clustal Omega sequence alignment showing conserved glycine-rich nucleotide-binding motif (TGxxGxxG); *ca.* 20 and 30 residues downstream of the glycine-rich motif are the two conserved arginine (R) residues which interact with the adenosine-2'-phosphate oxygen.[73]

CpkC-TR contains the glycine-rich nucleotide binding motif (**Figure 8**) with the consensus sequence TGxxGxxG with three glycine residue (G31, G34 and G37) used to identify reductases belonging to the 'extended' SDR family. In a thioester reductase of known structure the glycopeptidolipid reductase (GPL-R) the glycine residues are also present (G59, G62 and G65). A chain length of 381 amino acid residues is within the range expected for extended SDRs (*ca* 350 residues).[37]

A set of residues downstream of the nucleotide-binding motif (**Figure 8**) are involved in determining the specificity of SDRs for either NADH or NADPH. NADPH-dependent enzymes are identified by the presence of two basic residues in Cpkc-TR, these are R57 and R67 which bind to the 2'-phosphate. The presence of an acidic residue (aspartate, D or glutamate, E) adjacent to the second basic residue indicates a preference for NADH. A glutamate residue (E66) is present in CpkC-TR. Enzymes with a preference for NADPH generally do not have this acidic residue. The GPL-R has an arginine (R96) at this position and it utilizes NADPH. The other arginine residues are present (R87 and R97) and can be seen clearly in the alignment. The alignments (**Figures 8** and **9)** show that five of the enzymes used for the analysis have this acidic residue (D or E) at this position: TenS-R, CpkC-TR, LymA-R, MyxcG-R. The TenS-R reductase does not catalyze redox-chemistry.[51] Myxochelin biosynthesis has been reconstituted *in vitro* using the purified enzymes in the presence of NADPH and MyxcG-R. The biochemical assay was carried out using NADPH not NADH. It is not clear whether MyxG-R utilizes NADH as well as NADPH as has been shown in at least one case of thioester reductase domains, the linear gramicidin thioester reductase (discussed in section **1.4.2.2** above).[41,45] The lymphostin reductase LymA-R has yet to be biochemically characterised to show its co-factor preference.[41]

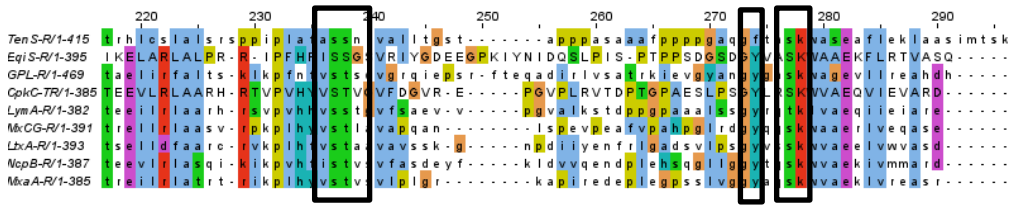


Figure 9: Sequence alignment showing conserved active site residues: threonine (or serine, S), tyrosine (Y), serine (S) or threonine (T) and lysine (K).

The prediction of co-factor preference can, therefore, not be made entirely based on primary sequence analysis alone. As seen in this study, additional *in vitro* experiments need to be carried out in order to determine which co-factor(s) will be utilized.

The catalytic threonine (T) and tyrosine (Y) in **Figure 9;** are generally conserved across all sequences, with the exception that threonine was substituted with a serine (S) in five of the sequences. This substitution is not expected to result in a significant change in activity since threonine and serine are nearly identical in chemical structure and activity. In CpkC-TR, the catalytic triad T159, Y189 and K193 were identified by comparison with the corresponding locations in the GPL-R enzyme (T193, Y228 and K232). Additionally, in the TenS-R the tyrosine is replaced by a phenylalanine (F). These data support *in vivo* studies showing that TenS-R does not carry out redox catalysis; it therefore, does not need the tyrosine residue in the active site for its enzymatic activity.[51]

Studies of the mechanism of catalysis carried out by SDRs suggest that the active site residues together with the co-factor and water molecules present in the active site participate in a proton relay system.[37,25] The proposed proton relay would involve the hydroxyl group of tyrosine (Y), the 2'-hydroxyl of the nicotinamide ribose and the amine group of lysine (K) while the hydroxyl group of the active site threonine (T) forms a hydrogen bond with the carbonyl containing substrate.[37,25]

2.2 Overproduction and Characterisation of CpkC-TR Domain and CpkC-ACP-TR Di-domain

2.2.1 Preparation of an Expression Vector for Overproduction of the CpkC-TR Domain and CpkC-ACP-TR Di-domain

Upon optimization of PCR conditions, both amplicons (*ca* 1100 and 1400 bp for *cpk*C-ACP-TR and *cpk*C-TR) were obtained (**Figure 10**), purified and analysed by DNA sequencing. Both PCR products were cloned into the pET151 (invitrogen) vector (**Figures 11 a** and **b**) which incorporates a hexahistidine tag to enable recombinant protein purification. Positive clones were further analysed by restriction digest (**Figure 12**) and then verified by sequencing to ensure there were no replication errors.

The pET151 vector system is a commercially available system used for the expression of heterologous genes in *E. coli* at high yields using the highly efficient and specific bacteriophage T7 promoter. Blunt ended PCR products are used for directional cloning by the introduction of a short vector recognition sequence into the forward primer; this eliminates the need for restriction enzymes and ligase. The cloning strategy utilizes the ability of topoisomerase I from *Vaccinia* virus to cleave double stranded DNA at a particular restriction site forming a phosphodiester bond between a thymine DNA residue and a tyrosyl residue of topoisomerase I; this bond can subsequently be broken to release the topoisomerase, by the attack of the 5'-hydroxyl of the original cleaved DNA strand, essentially reversing the reaction.[74,75] The pET151 vector is available in a linearized form with topoisomerase I attached to double stranded DNA with a four nucleotide overhang sequence (GTGG). Introduction of the complementary sequence (CACC) to the forward primer used for PCR will result in annealing of the PCR product with the vector overhang sequence (when incubated with the linearized vector) and release of the topoisomerase I. The cloning reaction is used to transform chemically competent *E. coli* cells which are used

for stable propagation and maintenance of the recombinant plasmids. The pET151 vector contains ampicillin resistance genes used for selection of positive clones. Protein expression is carried out in BL21 Star^{TM}(DE3) *E. coli* cells which contain the T7 RNA polymerase that recognizes the T7 promoter sequence on the pET151 vector. Expression of RNA polymerase in BL21 Star^{TM}(DE3) cells is induced by the addition of isopropyl β–D-thiogalactoside (IPTG) to the medium because IPTG binds to the lac repressor protein causing a conformational change which prevents it from binding to the lac operator sequence upstream to the T7 promoter.[74,75]

The two proteins CpkC-ACP-TR and CpkC-TR comprising the ACP-TR di-domain and the stand-alone TR domain respectively were over-produced in *E. coli* and purified to homogeneity by nickel affinity chromatography. Their molecular weight and purity were determined by SDS-PAGE (**Figure 13**). Typical yields of recombinant protein were *ca.* 1mg/L.

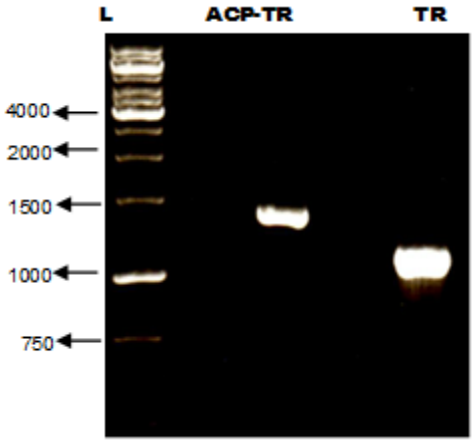
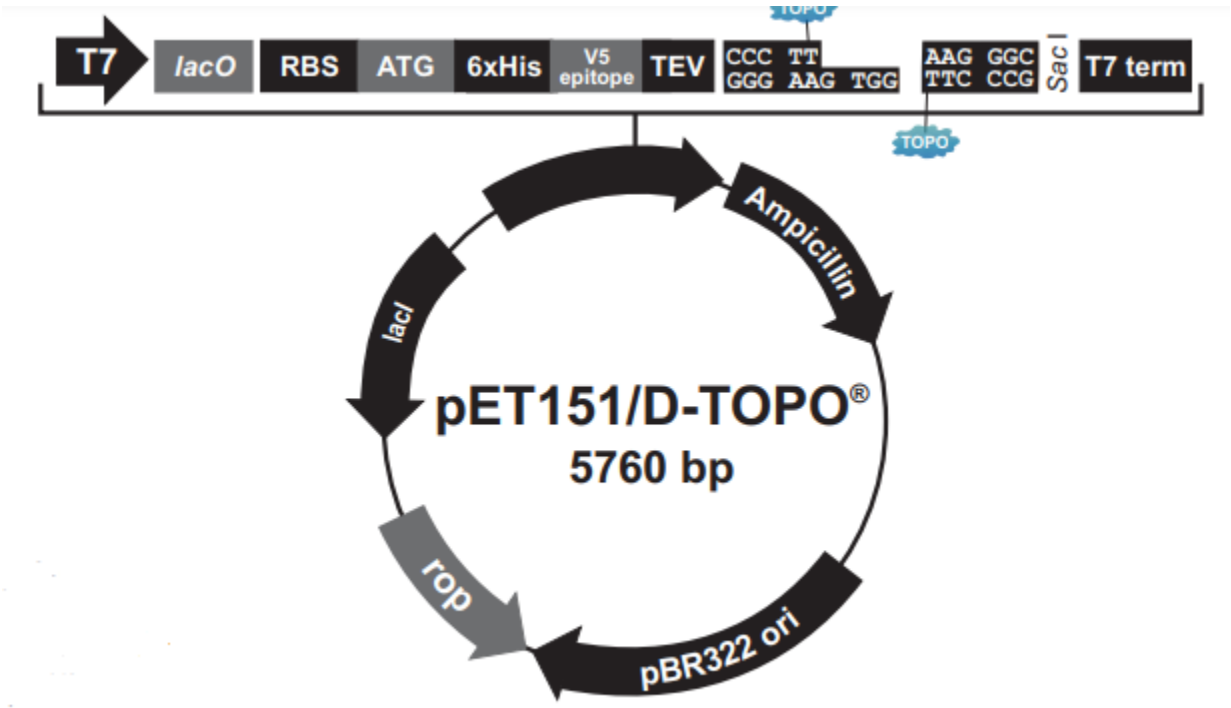


Figure 10: Gel electrophoresis analysis of PCR products obtained for the DNA fragments encoding the TR domain (*ca.* 1100 bp) and ACP-TR di-domain (*ca.* 1400 bp). ***Lane L:*** DNA fragments of known chain lengths (bp).



Comments for pET151/D-TOPO®
5760 nucleotides

T7 promoter: bases 209-225
T7 promoter priming site: bases 209-228
lac operator (lacO): bases 228-252
Ribosome binding site (RBS): bases 282-289
Initiation ATG: bases 297-299
Polyhistidine (6xHis) region: bases 300-317
V5 epitope: bases 318-359
TEV recognition site: bases 360-380
TOPO® cloning site (directional): bases 387-400
T7 reverse priming site: bases 455-474
T7 transcription termination region: bases 416-544
bla promoter: bases 849-947
Ampicillin (*bla*) resistance gene: bases 948-1808
pBR322 origin: bases 1953-2626
ROP ORF: bases 2997-3188 (complementary strand)
lacI ORF: bases 4500-5612 (complementary strand)

Figure 11a: pET151 complete vector map.

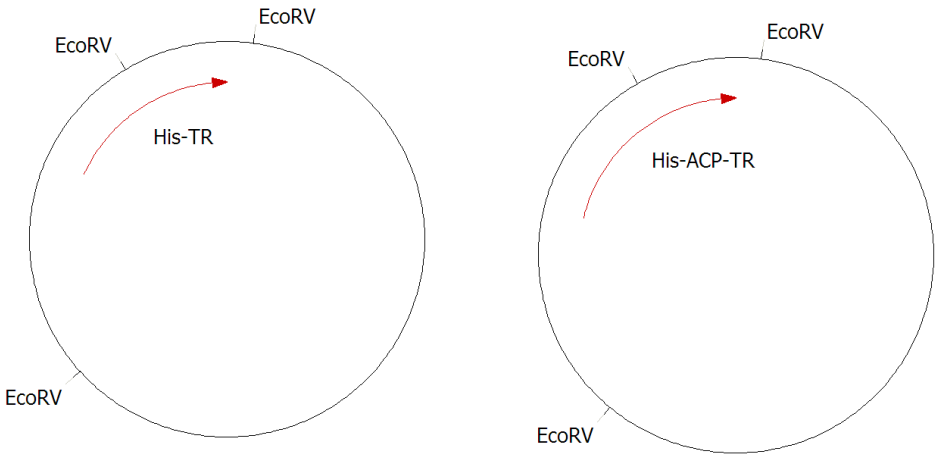


Figure 11 b: Plasmid containing pET151 ACP-TR and TR inserts showing the restriction sites for *Eco*RV.

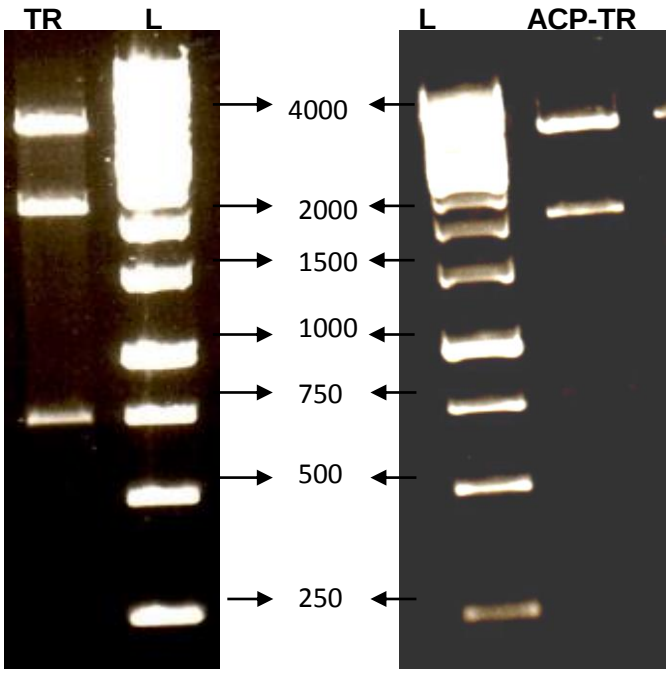


Figure 12: Gel electrophoresis analysis of *Eco*RV restriction digest fragments of pET151ACP-TR plasmids constructs (expected sizes *cpk*C-TR: 742, 1939, 4237 and *cpk*C-ACP-TR: 595, 2203 and 4237bp). **Lane L:** DNA fragments of known chain lengths (bp).

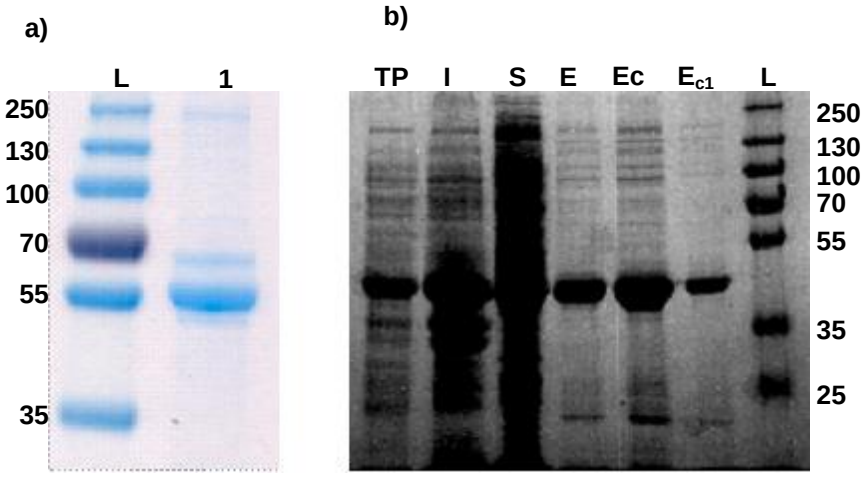


Figure 13 a and b: SDS-PAGE Analysis of **a)** CpkC-ACP-TR (***Lanes L:*** proteins of known molecular weight, ***Lane 1:*** CpkC-ACP-TR) and **b)** (CpkC-TR. (***TP:*** total protein, ***I:*** insoluble, ***S:*** soluble, ***E:*** eluate, E_C***:*** concentrated eluate, E_{C1}***:*** 1µl of concentrated eluate.

2.2.2 Overproduction and Characterisation of His_6-CpkC-TR and His_6-CpkC-ACP-TR

The molecular mass of both purified proteins agreed with the calculated molecular weight of the hexahistidine-tagged recombinant protein ACP-TR: calculated molecular weight: 55158 *Da*, observed: 55156 *Da* (**Figure 14**).[76]

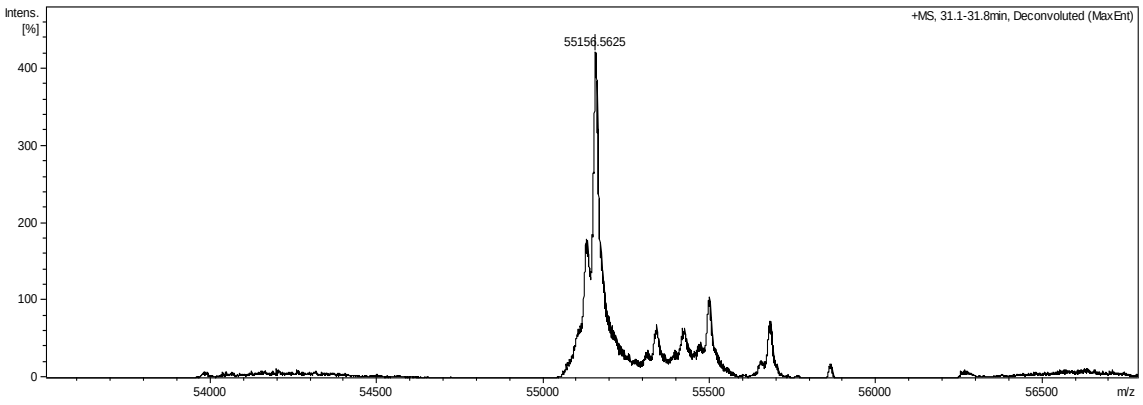
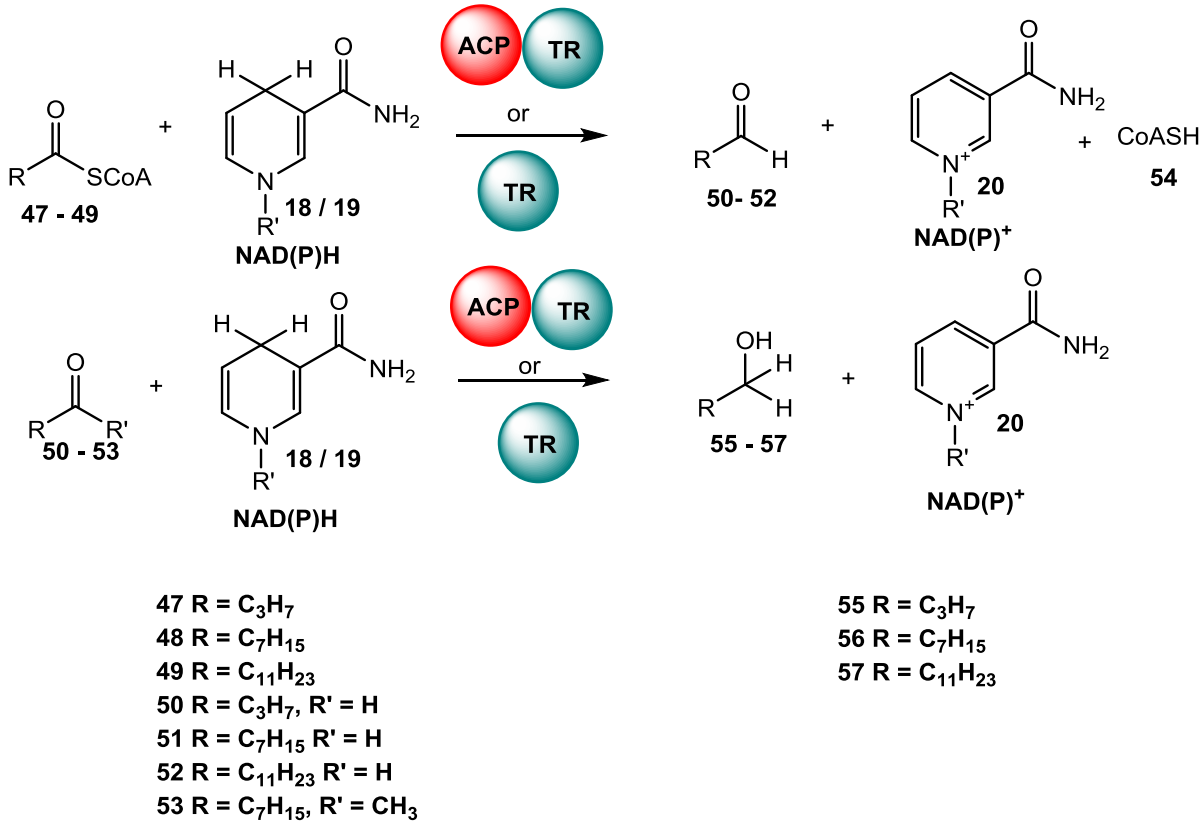


Figure 14: Mass spectrum of purified CpkC-ACP-TR didomain showing molecular mass of *ca.* 55 kDa.

2.3 Enzymatic Activity of the CpkC-TR Domain

2.3.1. Reduction of the Octanoyl-CoA Thioester Bond

Scheme 23 shows the reactions catalysed by the thioester reductase resulting in the reduction of substrate analogues of the native coelimycin biosynthetic intermediates used during the course of this study to probe enzymatic activity. The CpkC-ACP-TR didomain was assessed initially using a fluorimetric assay (**Figure 15**) that detects the change in intensity of fluorescent emission of NAD(P)H at 462 nm. The substrate analogues used were a range of commercially available short to medium chain acyl-CoAs: butanoyl (**47**), octanoyl (**48**) and dodecanoyl (**49**).



Scheme 23: Thioester reductase enzymatic assay using short to medium chain length acyl-CoA and aldehyde substrate analogues.

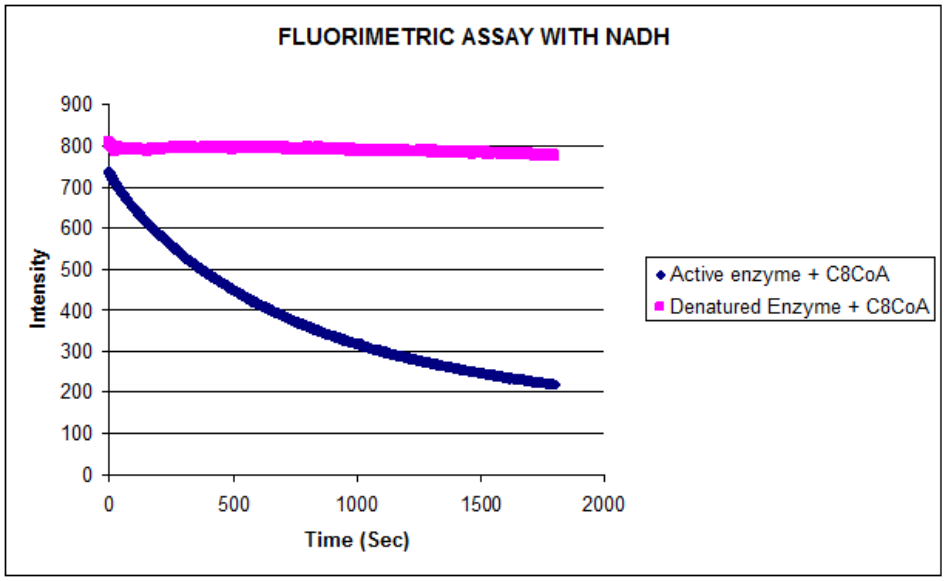


Figure 15a: ACP-TR fluorimetric assay of enzymatic activity with substrate mimic octanoyl-CoA (**48**) and NADH co-factor. Wavelength for absorption: 340 nm; and emission 462nm.

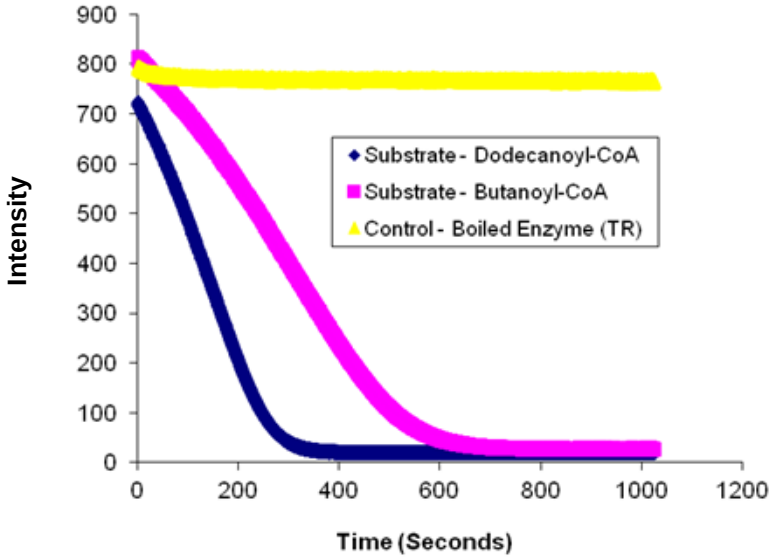


Figure 15b: ACP-TR flourimetric assay of enzymatic activity with other acyl-CoA substrates: butanoyl-CoA (**47**) and dodecanoyl-CoA (**49**) showing substrate tolerance. Wavelength for absorption: 340 nm; and emission 462nm. Enzyme concentration: 20 μM, NADH, butanoyl-CoA and dodecanoyl-CoA concentration: 200 μM.

In similar studies of terminal reductase domains, it has previously been shown that enzymatic activity could be detected when utilizing acyl-CoA thioester analogues of the ACP (or PCP)-bound substrate of the cognate reductase.[39] Terminal reductases generally exhibit relaxed substrate specificity and tend to accept a range of substrate analogues that are reasonably similar to the native biosynthetic intermediate. This is probably because they play a role in maintaining metabolic flux through the biosynthetic assembly line.[39]

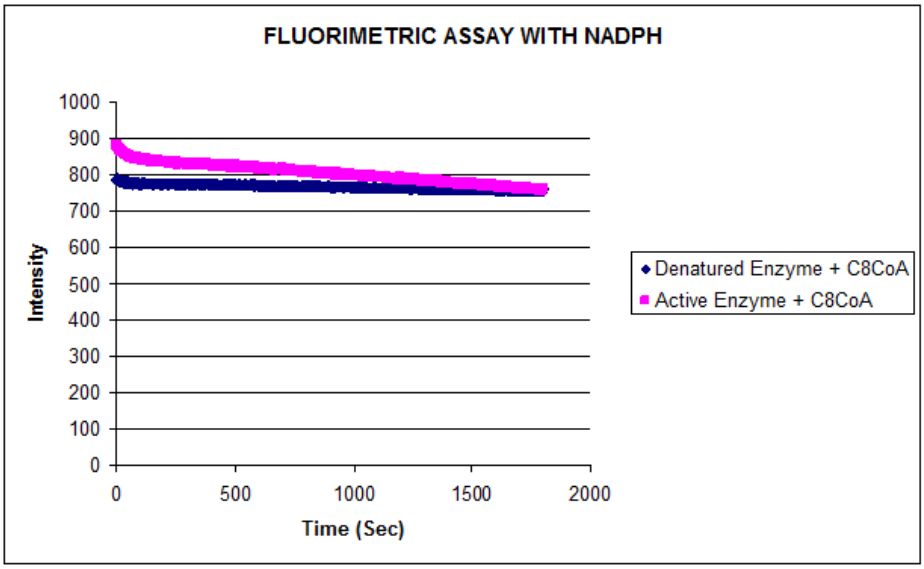


Figure 16: ACP-TR fluorimetric assay of enzymatic activity with substrate mimic octanoyl-CoA (**48**) and NADPH co-factor. Wavelength for absorption: 340 nm; and emission 462nm. Enzyme concentration: 20 μM, NADPH, octanoyl-CoA concentration: 200 μM.

In this study, it was shown that the CpkC-TR domain preferentially utilizes NADH over NADPH (**Figures 15 a** and **b**). A visible decrease in the emission of NADH in the enzymatic reaction was observed over a period of 30 min. When the co-factor was changed to NADPH, there was little or no decrease in the intensity (**Figure 16**) of the emission over

the same period of time under identical reaction conditions. Most of the other biochemically characterized reductases utilized NADPH except in the case of a stand-alone reductase involved in the biosynthesis of the zeamine antibiotics which also has a marked preference for NADH over NADPH.[40]

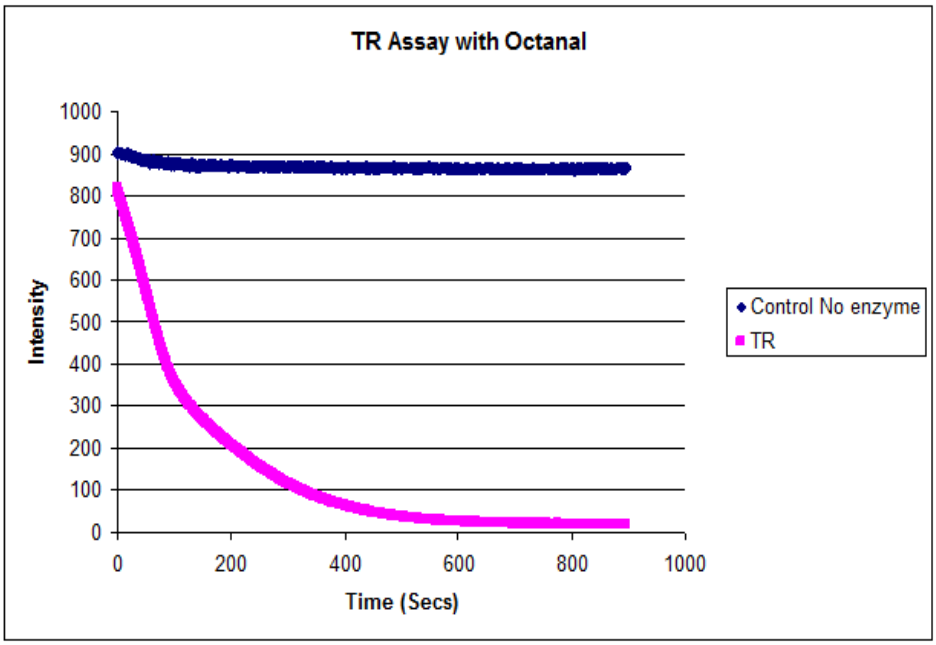
2.3.2 Substrate tolerance

Using the same fluorescence emission assay, octanal (**51**) and 2-octanone (**53**) were also tested as substrates (**Figures 17 a** and **b**). As discussed in section **1.4** several thioester reductases in other PKS and NRPS biosynthetic contexts have been shown to perform an additional reduction step (following the initial reduction of the phosphopantethienyl-bound thioester intermediate to give an aldehyde) leading to the formation of an alcohol. This experiment was carried out, to determine if CpkC-ACP-TR accepted aldehydes and ketones as substrates in addition to thioesters.

The fluorescence emission assay, showed a decrease in intensity of emission at 462nm with octanal (**51**) as a substrate (**Figure 17 a**); suggesting the CpkC-TR domain is capable of further reducing the initial thioester-bound intermediate to an alcohol.

NADH-dependent dehydrogenases have also been studied in attempts for their potential use in the synthesis of chiral alcohols.[77] Significant efforts have been directed towards developing biocatalytic routes to these and other important industrial chemical intermediates aiming to achieve higher yields, fewer by-products, greater production efficiency and safer reaction conditions. Octan-2-one (**53**) was therefore tested as a substrate for the TR domain. Fluorimetric analysis showed a decrease in intensity at 462nm (**Figure 17 b**) indicative of a productive reaction.

a



b

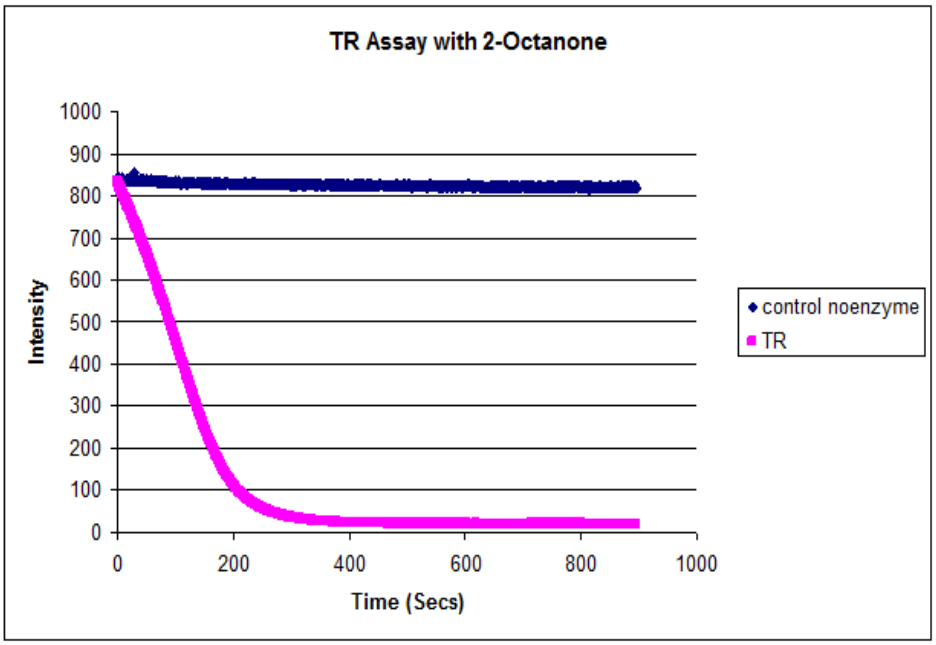


Figure 17: Fluorimetric assay of TR enzymatic activity showing a decline in the intensity of NADH fluorescence at 462 nm in the presence of an aldehyde: octanal (**a**) and a ketone: 2-octanone (**b**). Wavelength for absorption: 340 nm; and emission 462nm. Enzyme concentration: 20 µM, NADH, octanal and 2-octanone concentration: 200 µM.

The stereochemistry of the secondary alcohol produced may be determined by a Mosher's ester analysis.[78] This is an NMR-based method where the absolute configuration of a stereogenic centre (either a secondary alcohol or a secondary amine) is determined by derivatizing the hydroxyl (or amino) functional group with the two enantiomers of a chiral carboxylic acid (α-methoxy-α-trifluoromethyl-phenylacetic acid, is commonly used) to form the two corresponding diastereomeric (R and S) Mosher's esters (or an amides).[78] The derivatized products are then analysed by proton NMR spectroscopy. The configuration of the stereogenic centre can be assigned because the methine proton, the ester carbonyl and the trifluoromethyl functional groups have been shown to be on the same plane in the most populated solution-phase conformer.[78] It would therefore be expected that in the (*R)*-ester the signal for the carbonyl proton would appear upfield relative to the signal of the (*S)*-ester.

Additional experiments to determine the identity and stereochemistry of the product reaction would be necessary for detailed characterization of the TR domain as a useful catalyst for asymmetric reduction.

2.3.3 Stereospecificity of Hydride Transfer

2.3.3.1 Enzymatic Synthesis of Stereospecifically Deuterated NADH

A hallmark of enzymatic catalysis is the stereoselectivity of their reactions. NAD(P)H-dependent dehydrogenases have been shown to exhibit absolute stereoselectivity for either of the two diastereotopic hydrogens at the C4 position (**Figure 18**), which are designated *pro*-S and *pro*-R.[36] These two protons have different NMR chemical shifts because the NAD(P)H adopts a folded conformation in solution, causing the *pro*-S hydrogen to be shielded by the ring current of the adenine ring.[79,80] Therefore, its chemical shift is slightly upfield relative to the *pro*-R hydrogen.[79] The *pro*-S hydrogen also shows greater splitting than the *pro*-R hydrogen because it is more closely aligned with the proton at the C5

position. (**Figure 18**).[79] The two C4 deuterated diastereomers of NADH were therefore prepared to probe the stereospecificity of hydride transfer of the CpkC-ACP-TR domain.

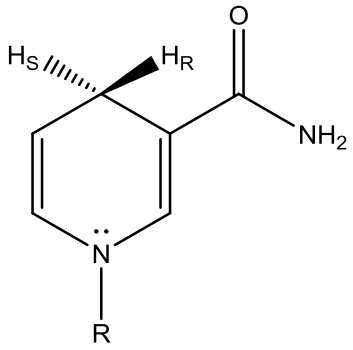
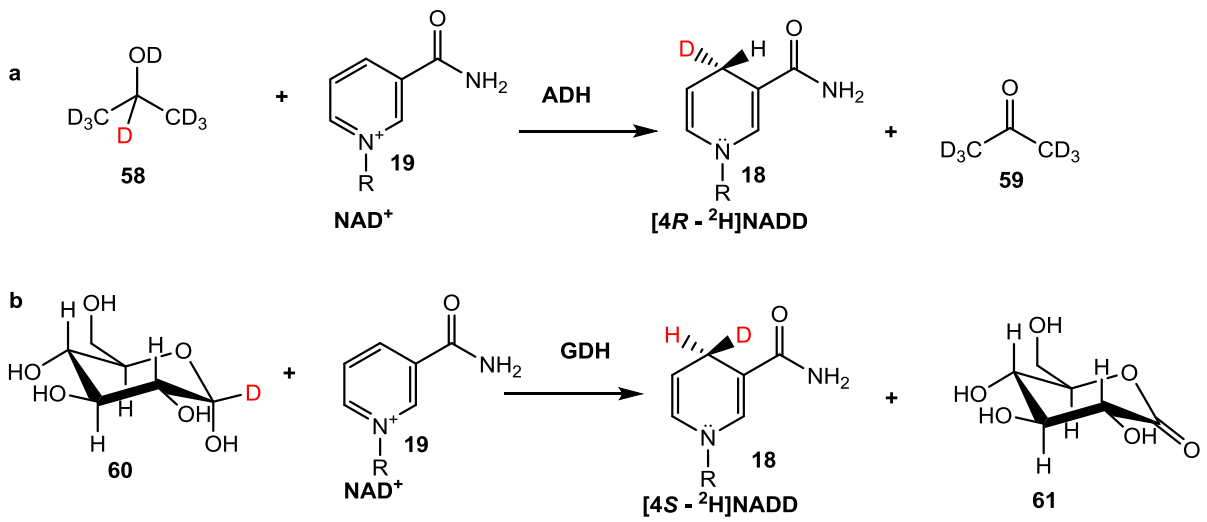


Figure 18: Structure of NADH showing the diastereotopic hydrogens.[36]

NAD^+ was separately incubated with alcohol dehydrogenase (ADH) and d_8-2-propanol (**58**) or glucose dehydrogenase (GDH) and 1-*d*-glucose (**60**) to (**Scheme 24**). The deuterated substrates were oxidized to form d_6-2-propanone (**59**) and *D*-glucono-1,5-lactone (**61**), though these compounds were not analysed. The ADH catalysed reaction produced [4R-^{2}H]NADH in a 71% yield and the GDH catalysed reaction produced [4S-^{2}H]NADH at a 50% yieldy.[79,80] ADH catalyzes the reduction of NAD^+ by the transfer of the hydride ion to the *pro*-S position.[79] In the experiment, the reduced co-factor (NADH) is deuterated at the *pro*-S position; therefore, the *pro*-R hydrogen is observed in the NMR spectrum. The exact reverse is the case for GDH.

The two diastereomers can be distinguished by measurement of proton NMR spectra.[79,80] The *pro-S* hydrogen appears upfield relative to the *pro-R* hydrogen and the *pro-S* hydrogen shows greater chemical coupling to the proton on C5 (**Figure 19**).



Scheme 24: a) Alcohol dehydrogenase catalyses the reduction of NAD^+ with *pro-R* specificity. b) Glucose dehydrogenase catalyses the reduction of NAD^+ with *pro-S* specificity.

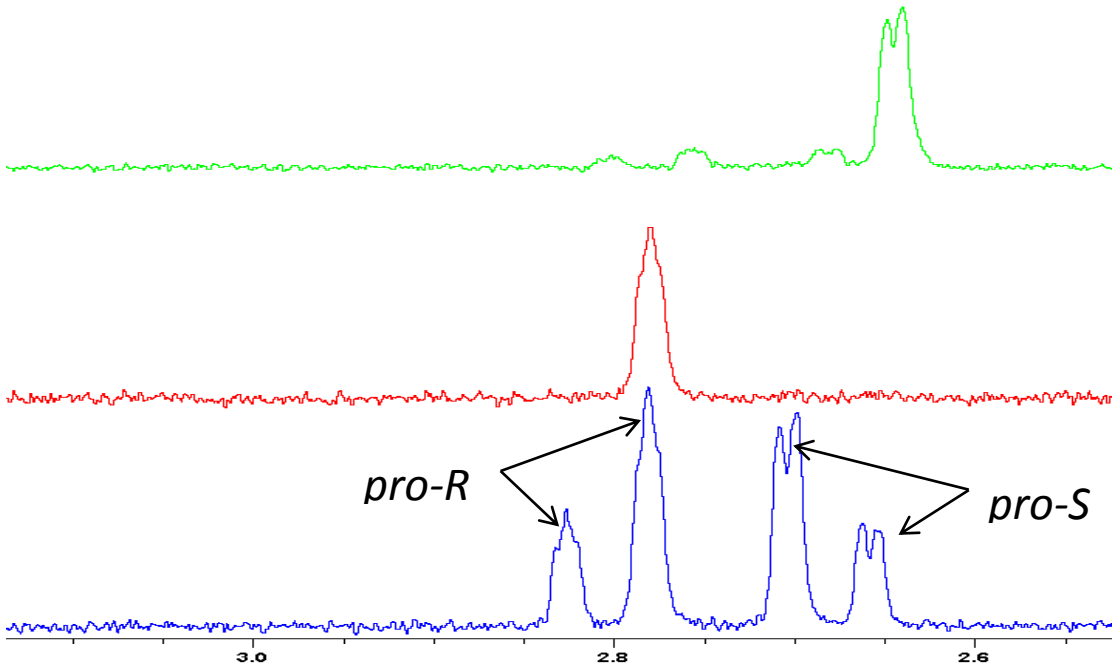


Figure 19: ^1H NMR spectra of NADH and 4-[^2H]NADH. Green trace [*4R*-^2H]NADH from alcohol dehydrogenase catalysed reaction. Red trace [*4S*-^2H]NADH from glucose dehydrogenase catalysed reaction. Blue trace: unlabelled NADH.

Incubating the TR domain with the two separate deuterated NADH co-factors with either octanoyl-CoA (**48**), octanal (**51**) or 2-octanone (**53**) followed by either LC-MS or GC-MS analysis should reveal the stereospecificity of hydride transfer. However, a method for detecting the products of the reaction would be required to enable this.

2.3.4 Substrate Consumption

Having established the enzymatic activity of the TR domain and its preference for the co-factor NADH (see above), an additional experiment to detect consumption of the substrate octanoyl-CoA (**48**) was carried out.

Octanoyl-CoA (**48**) was incubated with CpkC-TR and NADH (**18**) for 12 hours, alongside a control reaction in which the enzyme was omitted and the reactions were analysed by HPLC-MS (**Figure 20**). A peak with a retention time 13.5 mins giving rise to a species with *m/z* = 892.20 was observed in both the control and enzymatic reactions. This peak corresponded to the negatively charged $[M-H]^-$ ion for octanoyl-CoA $C_{29}H_{49}N_7O_{17}P_3S^-$ (calculated *m/z* = 892.2118, observed *m/z* = 892.2231). Its reduced intensity in the reaction containing CpkC-TR compared to the control reaction indicated that the acyl-CoA substrate (octanoyl-CoA) is being consumed by in the enzymatic reaction.

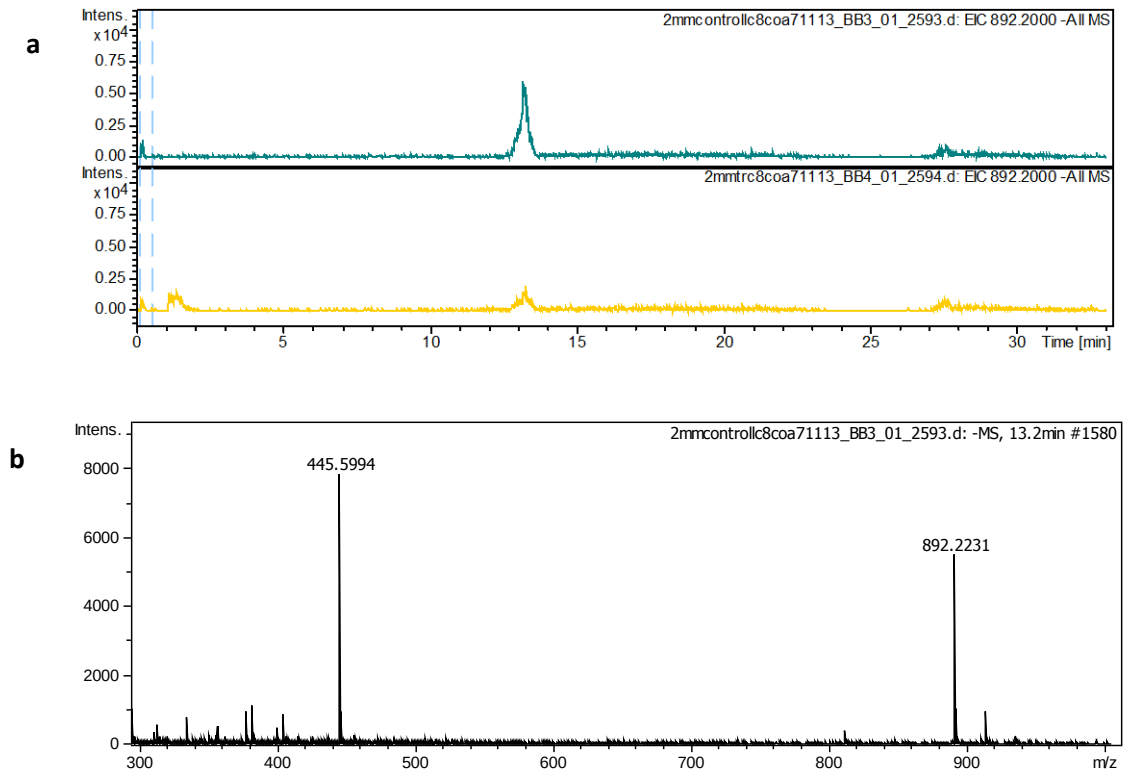


Figure 20a and b: HPLC-MS analysis of TR enzymatic reaction for detection of octanoyl-CoA (**48**) consumption **a)** Extracted ion chromatograms at *m/z* = 892.2 top: control reaction without enzyme and bottom: TR reaction containing the active enzyme. **b)** Mass spectrum for $[M-H]^-$ ion of octanoyl-CoA.

2.3.5 Product Detection

The expected products for the TR catalysed reaction (**Scheme 4**) of octanoyl-CoA (**48**) with NADH (**18**) are octanal (**51**) and octanol (**56**), CoASH (**54**) and NAD^+ (**19**).

2.3.5.1 Detection of Co-enzyme A

The same TR reaction mixture analysed in section **2.3.4** above (for octanoyl-CoA consumption), was analysed for the formation of the co-product co-enzyme A (CoASH).

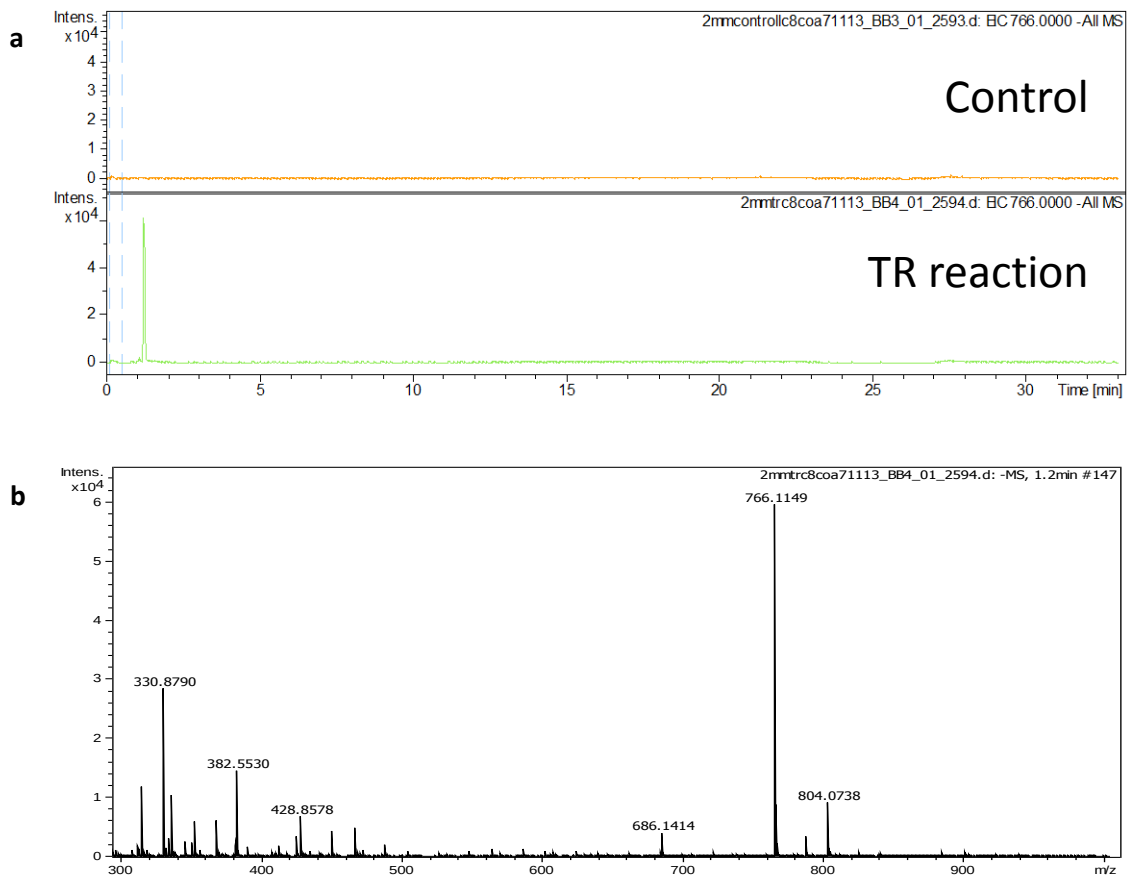


Figure 21: HPLC-MS analysis of TR reaction for detection of CoASH (**54**) formation **a)** Extracted ion chromatograms at *m/z* = 766.00 top: control reaction without added enzyme; bottom: TR reaction with active enzyme. **b)** Mass spectrum for $[M-H]^-$ ion of CoASH.

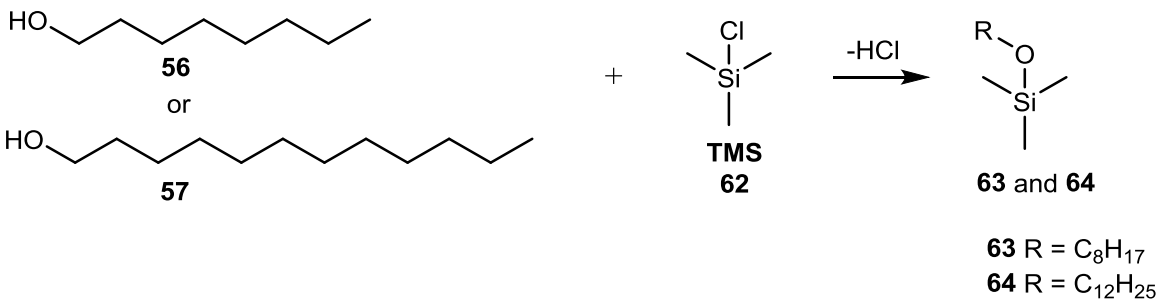
A peak at retention time 1.5 mins (**Figure 21**) with *m/z* = 766.00 was observed in the TR enzymatic reactions and was absent in the control reaction where the enzyme was omitted. This peak corresponded to the negatively charged ion of $[M-H]^-$ ion for CoASH $C_{21}H_{35}N_7O_{16}P_3S^-$ (calculated *m/z* = 766.1074, observed *m/z* = 766.1149). Its presence in the reaction containing CpkC-TR indicated that the co-product CoASH is being formed.

2.3.5.2 Detection of the Alcohol

Several attempts at detection of octanal (**51**) or octanol (**56**) directly by both GC-MS and HPLC-MS analysis were unsuccessful as no reproducible results could be obtained. Previous studies had shown that in cases where the TR catalysed a further reduction of the aldehyde product to the corresponding alcohol, the alcohol product was more readily detected than the aldehyde.[39] This is probably due to the significantly higher rate of enzymatic reduction of the aldehyde to the corresponding alcohol compared to the reduction of the thioester. This would account for the insufficient accumulation of the aldehyde in the reaction mixture making its detection challenging.

Organic extraction of the enzymatic reaction mixture followed by derivatization with *N,O*-bis(trimethylsilyl)acetamide to convert the alcohol to the trimethylsilyl ether (**Scheme 25**), was carried out in order to enhance product detection by GC-MS analysis (**Figure 22a** and **b**).

In the first instance dodecanoyl-CoA was used as a substrate for TR since it was a closer mimic to the natural substrate (having the same carbon chain length) and was expected to be more readily converted than other commercially available acyl-CoAs. The dodecanoyl-CoA TR reaction contained a peak (**Figure 22 a**) with a retention time of = 8 mins and a *m/z*= 243.0 corresponding to the fragment ion resulting from a neutral loss of a methyl group from trimethylsilyl-1-dodecyl ether. An identical peak was observed in a sample of an authentic standard of the 1-dodecanol (**57**) which was derivatized to its trimethylsilyl ether. This peak was absent in the control sample.



Scheme 25: Derivatization of alcohol (product of TR enzymatic reaction) with trimethylsilyl chloride **62** for GC-MS detection.

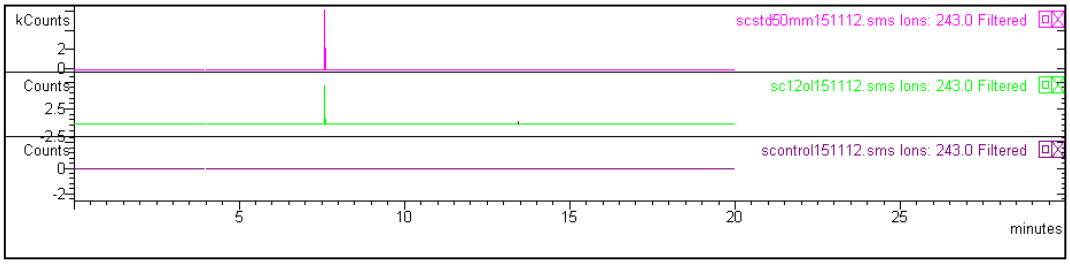


Figure 22a: GC-MS detection of trimethylsilyl-derivatized dodecanal. Top: authentic standard solution of derivatized 1-dodecanol (**57**) middle: ACP-TR with dodecanoyl-CoA bottom: control.

To confirm the products of the reactions in the experiments discussed in sections **2.3.1** and **2.3.2** (showing that TR accepted octanoyl-CoA and octanal as substrates), the above reaction for alcohol detection was repeated in separate reactions using octanoyl-CoA and octanal respectively. Identical results showing the formation of octanol (**Figure 22 b**) were observed. This provides evidence that the CpkC-TR domain catalyses the reduction of a thioester to its corresponding aldehyde which is further reduced to the alcohol *in vitro*.

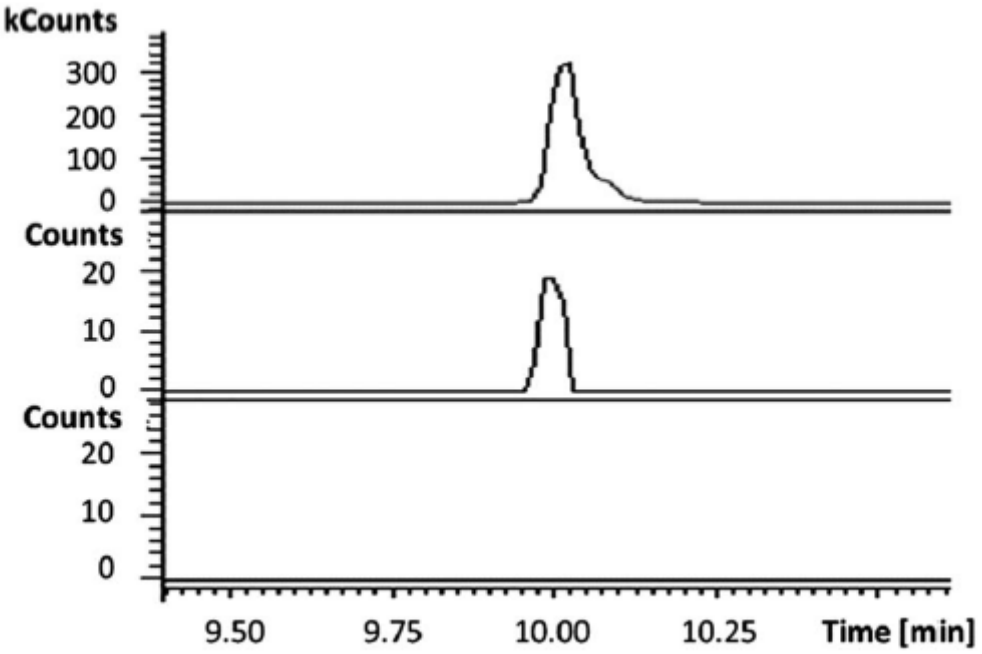


Figure 22b: Extracted ion chromatograms at *m/z* = 187.2 (corresponding to the fragment of the BSA derived octanol (**56**) after neutral loss of a methyl group) of GC-MS analysis of CpkC-TR catalysed reaction of octanoyl-CoA (**48**) with NADH (**18**). Top: authentic standard solution of derivatized 1-octanol; middle: CpkC-TR enzymatic reaction; bottom: control reaction.

2.3.6 Conclusions and future work

The results presented here, show that the TR domain in the coelimycin P1 modular PKS, catalyses release of the ACP-bound fully assembled polyketide chain via NADH-dependent reduction of the thioester. The enzyme preferentially utilizes NADH over NADPH and exhibits a relaxed substrate specificity being able to catalyse the reduction of aldehydes and ketones (in addition to thioesters). Co-factor and substrate consumption, as well as product accumulation was detected. However, the aldehyde intermediate in formation of the alcohol could not be detected directly possibly as a result of its transient accumulation under

the conditions employed. The detection of the corresponding alcohol resulting from the reduction of the aldehyde and the ability of the TR domain to accept aldehydes as substrates for reduction, provide indirect evidence that the aldehyde is an intermediate in coelimycin biosynthesis.

Stereospecifically deuterium-labelled NADH was prepared for the purpose of examining the stereospecificity of hydride transfer, but due to circumstances beyond my control, it was not possible to complete these experiments.

Further experiments can be carried out to determine the stereospecificity of hydride transfer from NADH catalysed by CpkC-TR using the deuterium-labelled NADH prepared, as co-substrates with a pro-chiral ketone (2-octanone for example). The incorporation of deuterium into the alcohol product can then be detected by GC-MS. In addition, the stereochemistry of the chiral alcohol product can be determined by Mosher's ester derivatization and analysis. This will provide additional data towards the use of CpkC-TR as an industrial biocatalyst for the synthesis of chiral alcohols.

CHAPTER THREE

RESULTS AND DISCUSSION II

3.0 Nitrogen Incorporation by PLP Dependent Transamination

3.1 Sequence Alignment

Within the biochemical context of the coelimycin pathway, the proposed substrate for the putative enzyme CpkG would be the aldehyde derived from reductive cleavage of the thioester intermediate produced by the CpkABC type I modular PKS.[7] The enzymatic function of this protein was predicted by carrying out a BLAST search of the amino acid sequence of the protein encoded by the *cpkG* gene and the results suggest the enzyme would most likely possess aminotransferase activity with the top 100 hits for the CpkG query being aminotransferases with a minimum sequence identity of 52% and a maximum of 99%.[81] Aminotransferases that utilize aldehydes, ketones (or primary and secondary amines as amino donors in the reverse reaction) or other compounds with the carbonyl group at a position distal to the carboxylate group (where a carboxylate group is present) as amino acceptors are called ω-aminotransferases.[58] A Clustal Omega multiple sequence alignment was used to analyse the primary amino acid sequence of the CpkG protein by comparing it with the sequences of known ω–aminotransferases (some of which, have had their X-ray crystal structures solved) to determine the presence of conserved amino acid residues present in PLP-dependent ω–aminotransferases.[73] A schematic diagram shown below (**Figure 23)**, highlights key active site residues present in an ω-aminotransferase from *E. coli* (which catalyzes the transamination of gamma-aminobutanoic acid) and was used as a model for interpreting the results of the sequence alignment.[58] The results of the alignment were viewed in Jalview and conserved regions are shown (**Figures 24 – 26)** below.

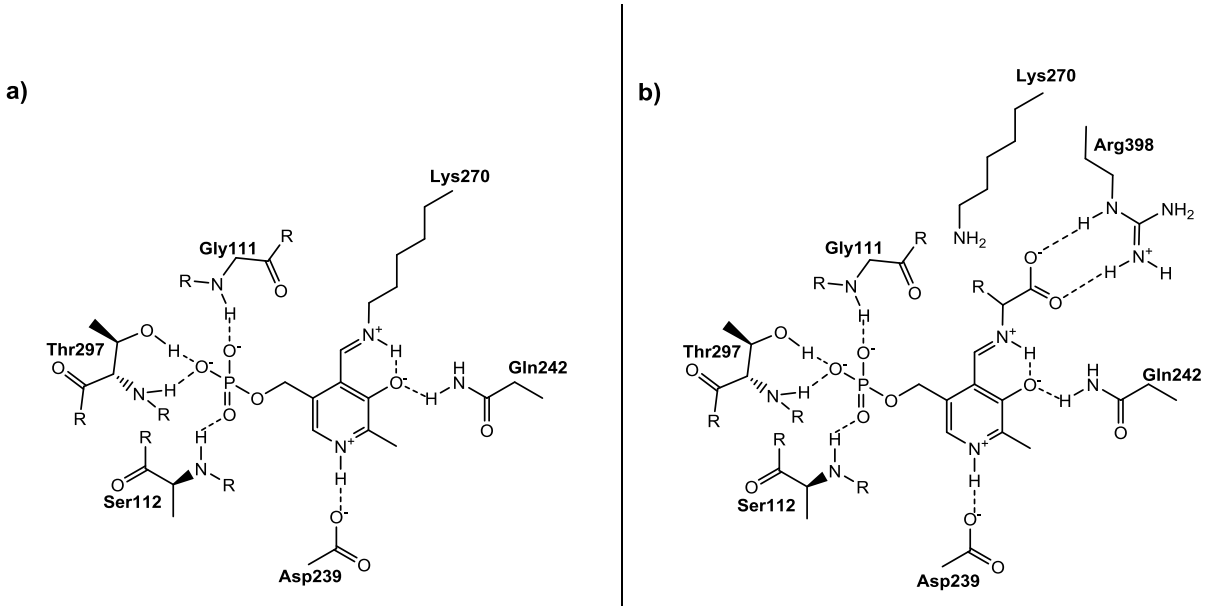


Figure 23: Active site model of a typical ω–aminotransferase (GABA-AT from *E. coli*) showing interactions of conserved residues with the **a**) PLP co-factor bound to lysine (internal aldimine) and **b**) PLP co-factor bound to a substrate (external aldimine).[58]

One common feature of all PLP-dependent enzymes is the presence of an aldimine bond between the PLP-cofactor and the Ɛ–amino group of a conserved lysine residue (**Figure 23**) in the holoenzyme active site.[36] Examination of the alignment (**Figure 24**) revealed the presence of this active site lysine residue (in all the aminotransferase protein sequences analysed) at position 338 in CpkG (the enzyme in this study). This lysine residue is required for acid-base catalysis and covalent attachment of the co-factor to the apo-enzyme.[36]

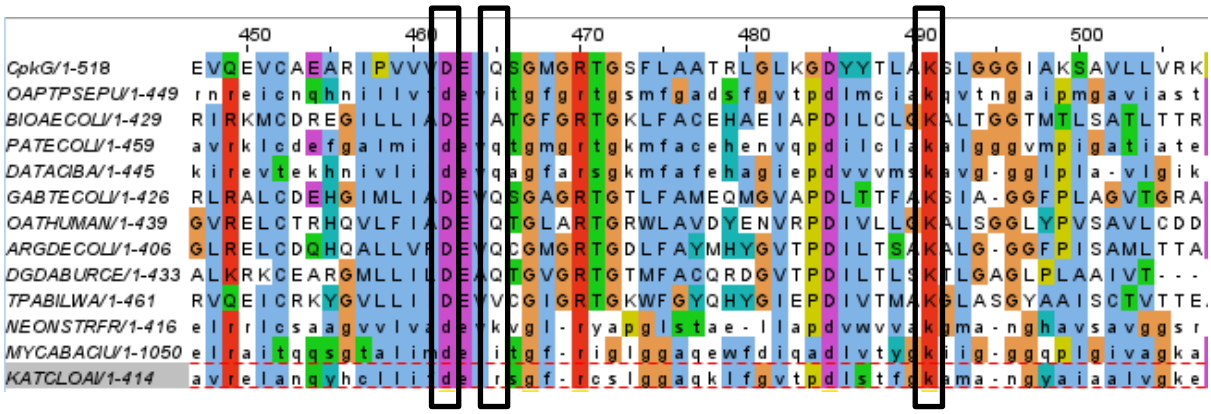


Figure 24: Sequence alignment of CpkG in with other ω–transferases of known (and putative) structure and function showing the strictly conserved active site lysine (K338) and aspartate (D309) residues. The glutamine (Q312) residue interacts with the phenolic oxygen of the co-factor.

Studies of the structure of the prototypical aminotransferase, aspartate aminotransferase showed that a second conserved residue *ca.* 26 amino acids upstream of the catalytic lysine, interacts with the pyridine nitrogen of the cofactor via its side chain carboxylate group.[82] This is an aspartate residue (D309 in CpkG) and its strict conservation in this group of aminotransferases can be seen clearly (**Figure 24**). The mechanistic importance of this residue has been discussed in section **1.5.1.2.**: it is needed to maintain the cofactor nitrogen atom in its protonated state to increase its electrophilicity and promote the formation of the quinoinoid intermediate thus driving the mechanism towards transamination and decreasing the likelihood of other possible side reactions. This residue is replaced in other PLP-dependent enzymes to change the outcome of the reaction. In the case of PLP-dependent racemases, an arginine residue replaces the aspartate found in their aminotransferase counterparts, resulting in the cofactor nitrogen being effectively un-protonated precluding transamination.[55] In the PLP-dependent ω–aminotransferases, γ–aminobutyric acid (GABA-AT) aminotransferase and ornithine aminotransferase (OAT), a glutamine residue (Q242 in GABA-AT and Q263 in OAT) *.ca* 3 residues downstream of the conserved aspartate residue, interacts with the co-factor phenolic oxygen via a hydrogen

bond. This residue is also conserved in CpkG (**Figure 24**) and is located at position 312 (Q312).

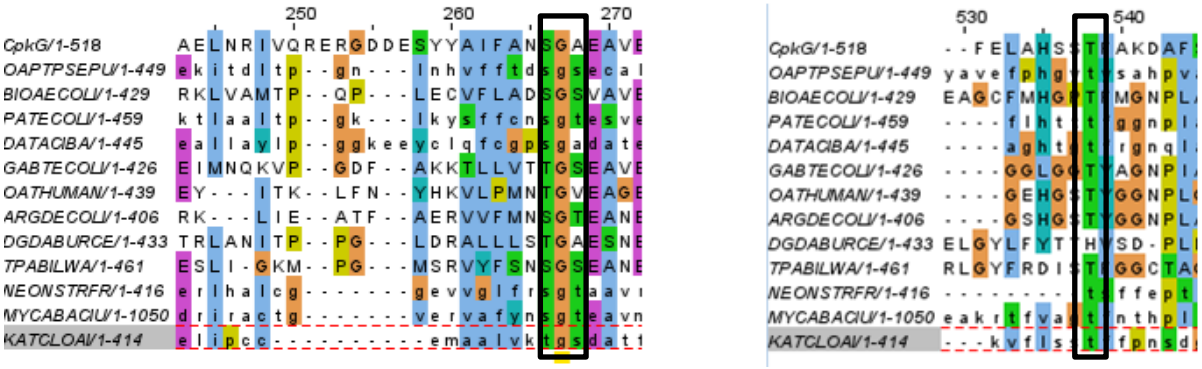


Figure 25: In CpkG a conserved glycine (G121) residue and an adjacent alanine (A122) residue are predicted to interact with the phosphate oxygens via the amide nitrogen atoms in the polypeptide chain. A threonine (T368) residue located further downstream is predicted to interact with a phosphate oxygen via hydrogen bonding.

Three other amino acid residues: a glycine, serine (or threonine or alanine) and a threonine are predicted to interact with the PLP phosphate group by hydrogen bonding via the amide group of the polypeptide backbone, and in the case of one of these residues, a conserved threonine, the side chain hydroxyl group (**Figure 25**). A combination of these interactions with other less conserved amino acids and in certain cases, a water molecule, makes up the "phosphate-binding cup" of the enzyme.[82,83] In CpkG and the other aminotransferases analysed (**Figure 25**), the first glycine and third threonine residues (of the phosphate binding group) are strictly conserved (G121 and T368 in CpkG; G111 and T297 in GABA-AT) the residue adjacent to glycine is variable (it can be either alanine, serine, threonine or valine); in CpkG, an alanine is located at this position (A122).

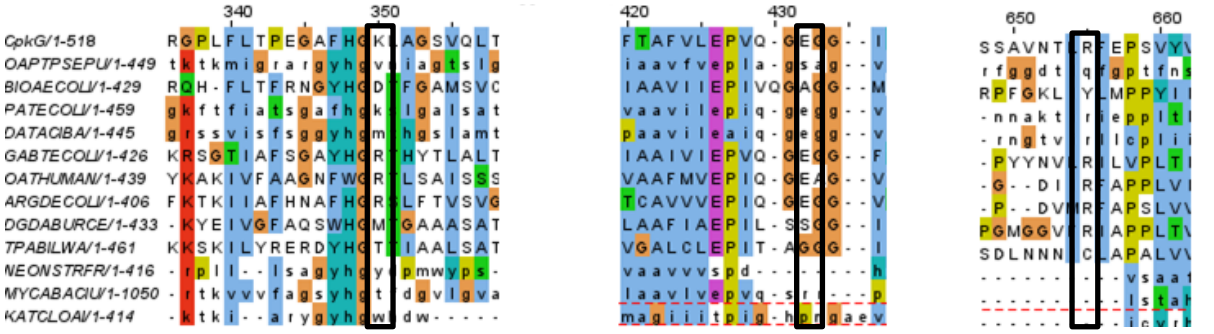


Figure 26: Sequence alignment showing residues predicted to be involved in substrate binding: in CpkG an arginine (R480) would be expected to interact with the α-carboxylate group of the amino acid substrate; a glutamate (E281) residue present only in α–ketoglutarate-specific ω–aminotransferases which forms a salt-bridge with an arginine residue as part of a gateway system; and a basic residue (K204 in CpkG) which is predicted to control substrate specificity towards a preference for linear amines (such as putrescine) that do not contain an α-carboxylate group.

Other potential interactions of interest that were observed by sequence alignment were the conservation of amino acid residues involved in substrate binding (**Figure 26**). It has been shown that varying some of these residues results in changes in the substrate preference of PLP-dependent enzymes.[58] An arginine residue (R398 in GABA-AT and R413 in OAT) binds the α-carboxylate group of the substrate amino acid positioning it for transamination of the amino group. This residue is present at position 480 in CpkG. A glutamate residue which is conserved only in α–ketoglutarate-specific aminotransferases (E211 in GABA-AT and E235 in OAT) is also conserved in CpkG (E281). This residue, together with the aforementioned arginine, make up the key elements of a gateway system used to ensure that the substrate is positioned and binds in a specific way that permits transamination of the ω–amino group while preventing competitive binding of α–amino acids.[58] The glutamate residue serves as a functional switch that is triggered by the formation of PMP. It forms a salt bridge with an arginine residue in the absence of PMP in

the first half of the reaction which prevents binding of α–amino acids (α-keto acids) and enables the ω–amino substrate to bind and react. Upon formation of PMP, the arginine residue is free to bind the α–ketoglutarate substrate which is then undergoes transamination.[58] An additional element identified in ω-aminotransferases that utilize α-KGA, is a second active site arginine residue (R141 in GABA-AT) which is used to form a salt-bridge with the γ-carboxylic acid group of α-KGA. This residue is replaced by a lysine (K204) in CpkG. A similar substitution was observed in putrescine aminotransferase from *E. coli*; which is specific *in vivo*, for amino donors (putrescine and cadaverine) that do not contain a carboxylate group, in common with the native substrate for CpkG. In at least one instance (a bacterial β-transaminase MesAT), certain ω-transaminases that have been predicted at primary sequence level to be specific for utilizing α-KGA as an amino acceptor, have been found to contain tertiary structural elements with pyruvate-specific ω-transaminases.[58] This can arise due to subtle structural re-arrangements resulting in an active site model that is a hybrid between α-KGA-specific and pyruvate-specific ω-transaminases.[58] These types of transaminases are highly versatile, utilizing a wide range of substrates including aromatic amino acids (such as phenyl alanine) and show potential for the industrial production of high value chemicals.

Overall, the sequence alignment shows that CpkG is likely to be a PLP-dependent ω–transaminase which would be capable of using α–ketoglutarate as an amino acceptor (glutamate as an aminodonor). This is in agreement with its predicted role in the coelimycin biosynthetic pathway.

3.2 Overproduction and Characterisation of His_6-CpkG

3.2.1 Preparation of an Expression Vector

PCR amplification of *cpk*G, was carried a product of the expected size: 1561 bp was obtained (**Figure 27**). The forward primer was designed to incorporate the vector recognition site CACC. Sequence integrity of the PCR amplicons was verified by DNA sequencing, prior to cloning in pET151 vector as discussed in section **2.1.2.**

In silico analysis of the expected vector map of the pET151 vector harboring the *cpk*G PCR product, was carried out using the Clone Manager software to select a suitable restriction enzyme for rapid clone analysis. *Pvu*II was predicted to cut the circular DNA sequence to produce four linear fragments of: 3637, 1502, 2085, 93 bp (**Figure 28**). Results of the incubation of the purified plasmid DNA with *Pvu*II were visualized by agarose gel electrophoresis (**Figure 29**). This showed the three largest predicted fragments. The smallest fragment is not visible likely because, it migrated at a significantly higher rate than the other fragments and probably ran off the gel into the electrophoresis buffer.

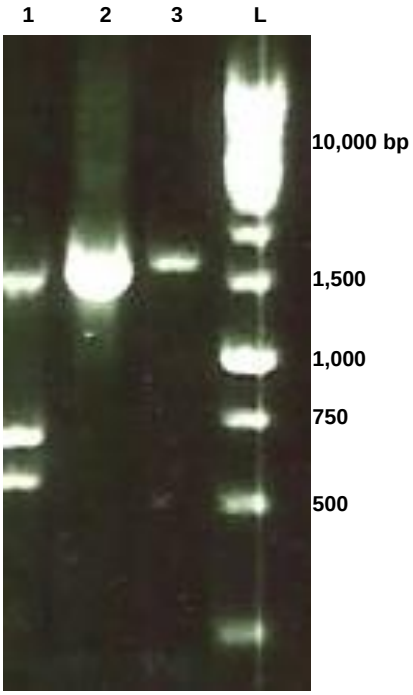


Figure 27: Gel electrophoresis analysis of the *cpk*G PCR product. **Lane 1**: failed PCR product (non-specific primer binding); **Lanes 2** and **3**: PCR products of targeted size. **Lane L:** DNA fragments of known chain lengths (bp).

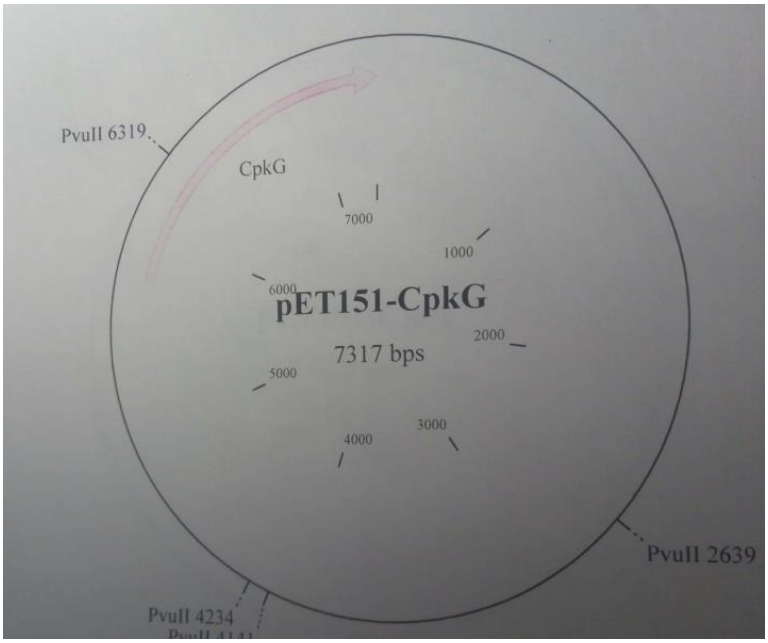


Figure 28: Restriction sites for *Pvu*II enzyme on the pET151-cpkG clone.

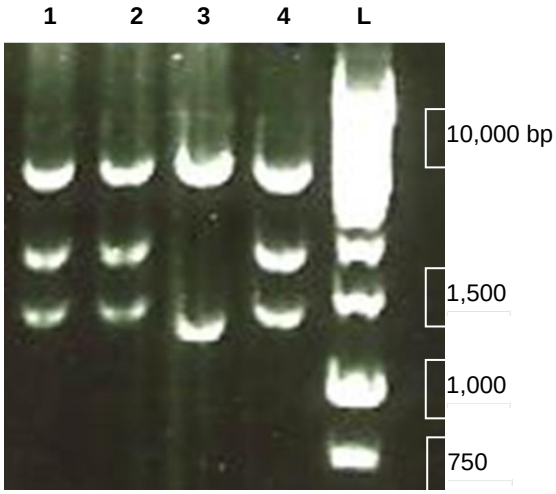


Figure 29: Gel electrophoresis analysis of *Pvu*II restriction digest of pET151-*cpk*G.
Lanes **1** to **4** contain plasmids of single *E. coli* colonies that were selected for further analysis. Positive clones **2** and **4** were confirmed by sequencing and used to transform *E. coli* BL21STR for protein expression. **Lane L:** DNA fragments of known chain lengths (bp).

3.2.2 Overproduction of His_6-CpkG

CpkG was overproduced in *E. coli* BL21STR as a soluble N-terminal hexa-histidine tagged fusion protein at yields of *ca.*14 mg/L. It was purified to homogeneity by FPLC. Its calculated weight is 60646 Da.[76] The purity of CpkG was ascertained by SDS-PAGE analysis and mass spectrometric analysis showed the molecular weight of CpkG to be 60645.2656 Da (**Figure 30**) which agrees with the calculated molecular weight.

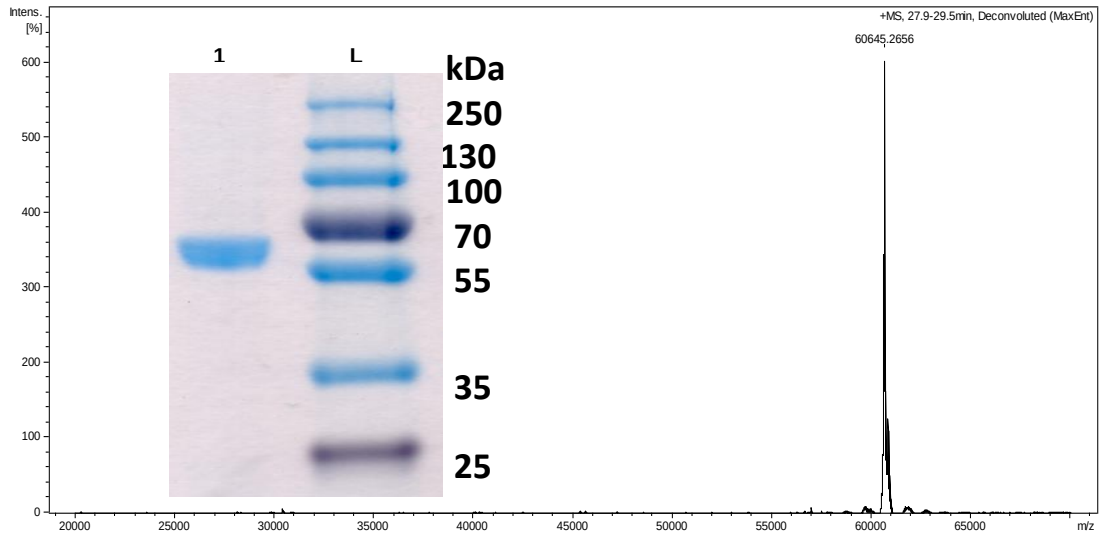


Figure 30: Analysis of N-terminal His_6-CpkG recombinant protein. **Insert:** SDS-PAGE analysis (**Lane 1:** His_6-CpkG, **Lane L:** proteins of known molecular weight) ; UHPLC-ESI-Q-TOF-MS analysis.

3.2.3 Oligomerization State of CpkG

The oligomerization state of CpkG was determined by size-exclusion chromatography (**Figure 31)**. The chromatography column was pre-calibrated with proteins of known molecular weight to determine their elution volumes. A calibration curve (**Figure 32**) was plotted and used to estimate the molecular weight of native CpkG. The equilibrium distribution coefficient K_{av}, was calculated for each protein of known molecular weight and plotted against the log of the protein molecular weight (in kDa). A trendline was used to estimate the molecular weight of native CpkG giving a value of *ca.* 137.56 kDa indicating that CpkG eluted as a dimer upon size-exclusion chromatography. This oligomerization state is consistent with class I aminotransferases which have been reported to function as homodimers (or higher order oligomers) where each monomer contributes essential residues to both active sites.[49]

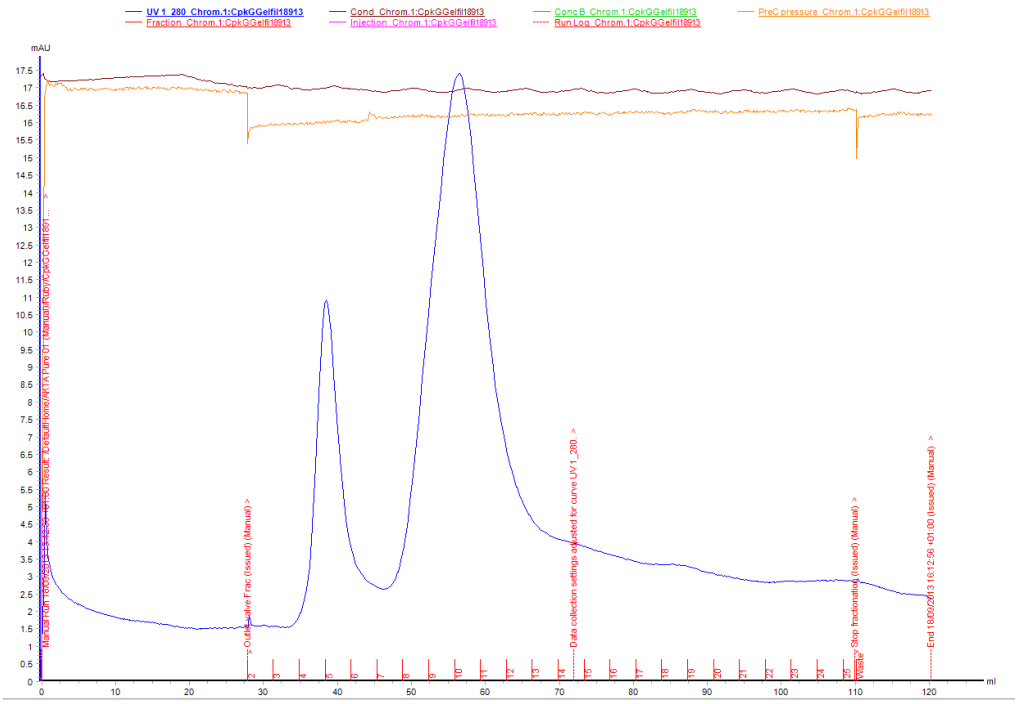


Figure 31: Size exclusion chromatogram of CpkG.

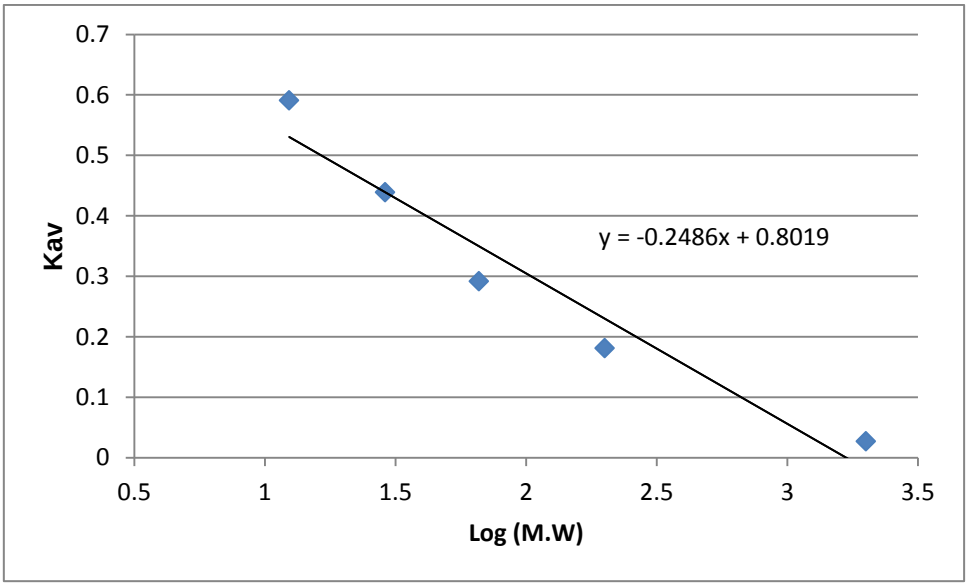


Figure 32: Size exclusion chromatography calibration curve. Standards used: 38 ml Blue dextran (2000 kDa), 49.4 ml β-amylase (200 kDa), 57.6 ml bovine serum albumin (66 kDa), 68.47 ml carbonic anhydrase (29 kDa), 79.74 ml cytochrome c (12.4 kDa).

3.2.4 Pyridoxal-5'-Phosphate Incorporation into the Active Site of CpkG

CpkG was analysed at pH 8 by UV-vis spectroscopy to detect if it could reversibly bind the PLP co-factor prior to enzymatic reaction assays (**Figure 33**). The UV-vis absorption of purified CpkG (**Figure 33**) showed two peaks at 330 nm and 420 nm corresponding to the absorption maxima of PMP and PLP-bound to the aminotransferase (the internal aldimine) respectively. The free co-factor has an absorption maximum of *ca.* 388 nm which is shifted upon binding with the active-site lysine to form the internal aldimine of the holo-enzyme.[84] Following incubation of CpkG with PLP in buffer solution, there was an increase in the absorption at 420 nm (**Figure 33**), indicative of incorporation of additional PLP in the active sites of more enzyme molecules causing a decrease in the absorption at 330 nm due to some displacement of PMP from the active sites, since PMP is not covalently attached to the enzyme. This showed that the purified His_6-CpkG could bind the co-factor necessary for catalysis.

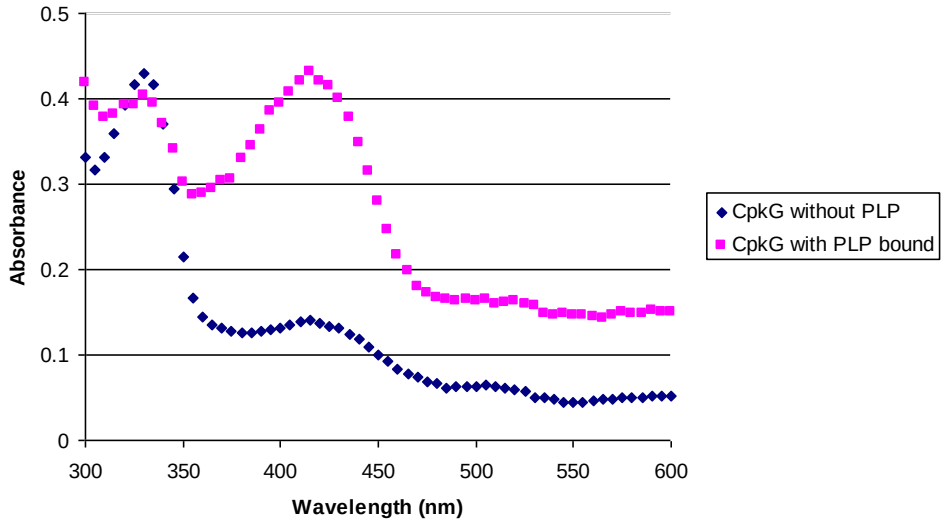
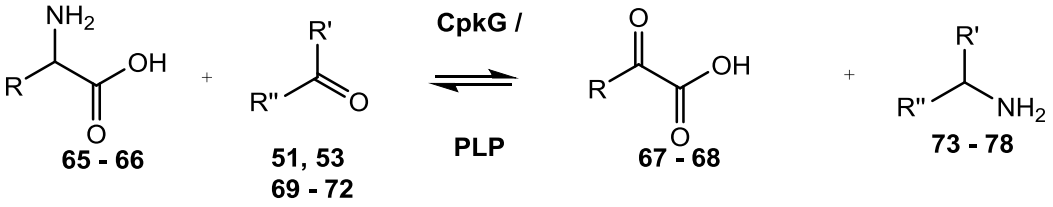


Figure 33: UV-vis spectrum of CpkG as purified (blue trace) after addition of 0.2mM PLP (pink trace).

3.3 Enzymatic Activity of His_6-CpkG

After establishing that CpkG is capable of binding PLP, additional experiments were designed to elucidate its enzymatic activity. The activity assays were carried out to detect the products of both the forward and reverse reactions (**Scheme 26**).



65, 67 R = CH_3
66, 68 R = CH_2CH_2COOH

69, 75	**R' = H, R'' = CH_3CH_2**
70, 76	**R' = H, R'' = $(CH_3)_2CH$**
71, 77	**R' = H, R'' = $NH_2(CH_2)_6$**
72, 78	**R' = H, R'' = C_6H_5**
73	**R' = H, R'' = $CH_3(CH_2)_7$**
74	**R' = CH_3, R'' = $CH_3(CH_2)_6$**

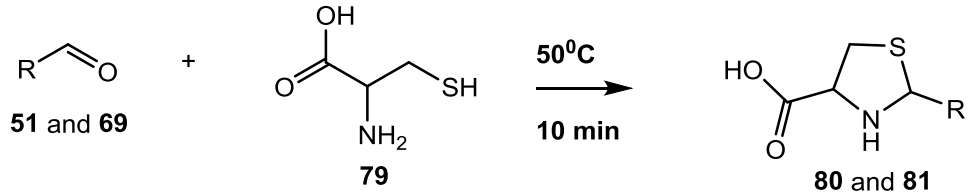
Scheme 26: Compounds examined as potential substrates for the reversible transamination reaction proposed to be catalyzed by CpkG.

3.3.1 Production of Aldehyde in the Reverse Reaction

Two aliphatic amines 1-propylamine and 1-octylamine were used separately (as amino group donors) in enzymatic reactions with CpkG to determine if they were converted to the corresponding aldehydes propanal and octanal.

Analysis of the enzymatic reaction containing 1-propylamine (**75**), pyruvate (**67**) and His_6-CpkG expected to produce propanal (**69**) was carried out by LC-MS in positive ion mode. Initial attempts to detect the product of the reaction by LC-MS analysis were unsuccessful, possibly due to the low ionization efficiency of aldehydes. It therefore, became necessary to carry out derivatization prior to LC-MS analysis. Derivatization with D-cysteine (**79**) converts aldehydes to their corresponding alkylthiazolidine-4-carboxylic acids (**Scheme**

27) which are more readily detected by ESI mass spectrometry due to their enhanced ionization efficiency.[85]



51, 80 : R = CH_3CH_2
69, 81 : R = $CH_3(CH_2)_6$

Scheme 27: Cysteine derivatization of aldehydes (propanal and octanal) for LC-MS detection.

After quenching with methanol to precipitate the protein, the propanal (**69**) was derivatized to ethylthiazolidine-4-carboxylic acid (**80**) which was detected by LC-MS (**Figure 34**).

The CpkG enzymatic reaction contained a peak (**Figure 34**) with a retention time of 2.5 mins giving rise to an ion with *m/z* = 162. This peak was absent in the control reaction lacking enzyme. The molecular formula of this compound $C_6H_{12}NO_2S^+$ was confirmed by HR-MS (calculated *m/z* of $[M+H]^+$ ion = 162.0589; observed 162.0583) corresponds to that of the $[M + H]^+$ ion of ethylthiazolidine-4-carboxylic acid (**80**).

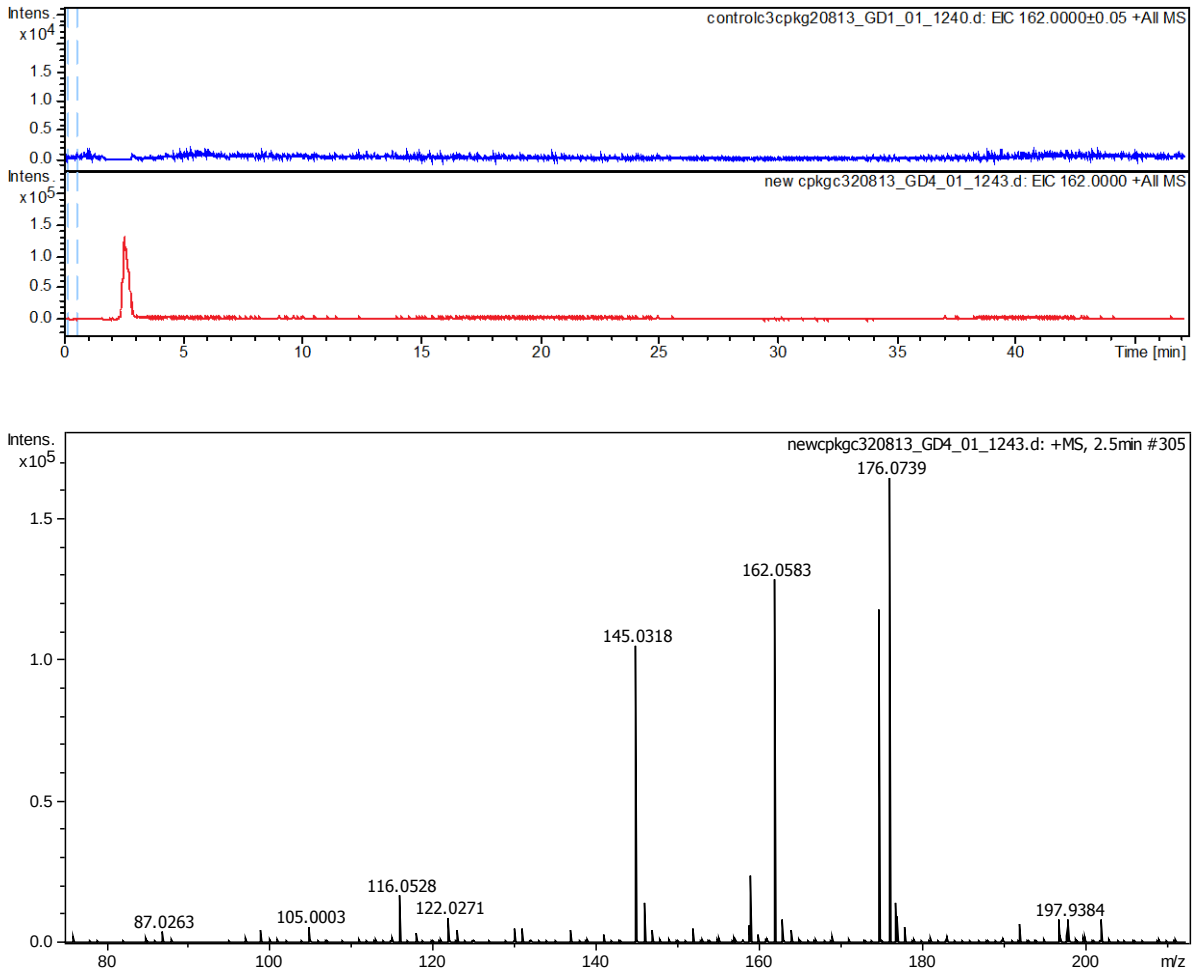


Figure 34: Extracted Ion Chromatogram at *m/z* = 162.1 from LC-MS analysis of CpkG corresponding to the product propanal (**69**) derivaised with D-cysteine (**79**). CpkG was incubated with 1-propylamine (**75**) as amino group donor and pyruvate (**67**) as amino group acceptor. Top: control reaction with no added enzyme and the bottom: enzyme assay. **b)** Mass spectrum of ethylthiazolidine-4-carboxylic acid (**80**).

The CpkG enzymatic assay was repeated using 1-octylamine (**73**). As a medium-chain alkylamine it is a closer mimic of the native substrate of CpkG and is easir to handle than 1-propylamine and propanal which are volatile. Both pyruvate (**67**) and α-KGA (**68**) were used as amino acceptors in separate reactions with 1-octylamine (**73**). The products

were derivatized with D-cysteine (**79**) which converted octanal (**51**) to hexylthiazolidine-4-carboxylic acid (**81**) as described above (**Scheme 7**). A standard solution of octanal (**51**) was also derivatized and analysed alongside to compare with the enzyme reaction sample (**Figure 35**).

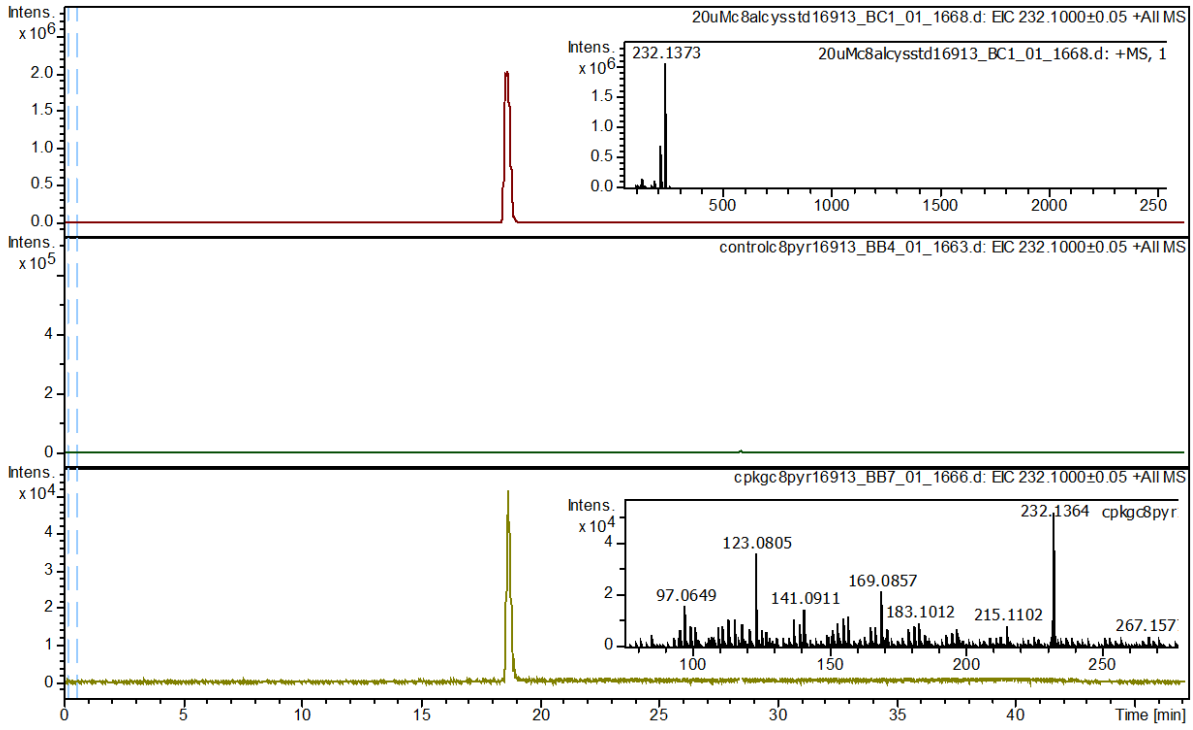


Figure 35: Extracted Ion Chromatogram at *m/z* = 232.1 corresponding to $C_{11}H_{22}NO_2S^+$ from LC-MS analysis of CpkG incubated with 1-octylamine (**73**) and pyruvate (**67**) showing D-cysteine derivatized octanal (**51**). Standard (top) control reaction with no enzyme (middle) and enzyme assay (bottom).

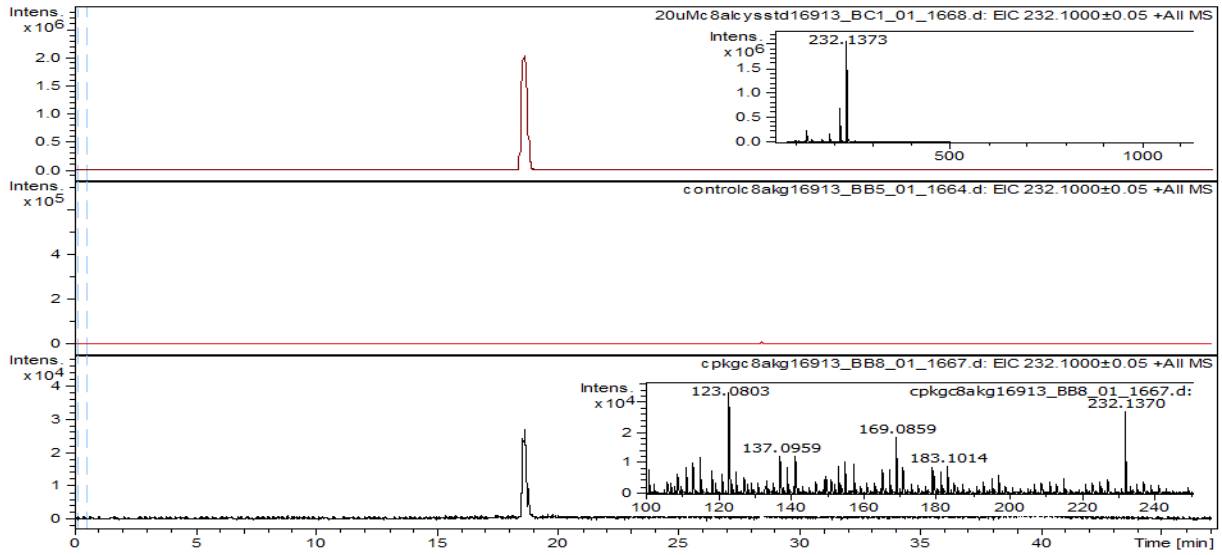


Figure 36: Extracted Ion Chromatogram of *m/z* = 232.1 corresponding to $C_{11}H_{22}NO_2S^+$ from LC-MS analysis of CpkG incubated with 1-octylamine (**73**) and α-KGA (**68**) showing D-cysteine derivatized octanal (**51**). Standard (top) control reaction with no enzyme (middle) and enzyme assay (bottom).

In the pyruvate and α-KGA containing reactions, respectively, a peak (**Figures 35** and **36**) of retention time 18.5 min giving rise to ions with *m/z* = 232.1364 and 232.1370 was observed. This was absent in the controls lacking enzyme. This peak had the same retention time and *m/z* value as an authentic standard of D-cysteine derivatized octanal (hexylthiazolidine-4-carboxylic acid). The molecular formula of the compound was confirmed by HR-MS (calculated *m/z* for $C_{11}H_{22}NO_2S^+$: 232.1371; observed: 232.1373). This confirmed that octanal was formed in the enzymatic reaction.

3.3.2 Production of Amino Acids in the Reverse Reaction

From the experiment in section **3.3.1**, it had been shown that CpkG accepts both pyruvate and α-KGA. Additional experiments were designed to assess the range of amino group donors which CpkG could accept. Pyruvate (**67**) and α–KGA (**68**) were incubated with

various amines (**69 – 72**) in the presence of CpkG and detection of the resulting amino acids, alanine (**65**) and glutamate (**66**), was carried out by employing a spectrophotometric assay for a blue amino acid-copper (II) complex with λ_{max} at 495 nm (**Figure 37**).[86]

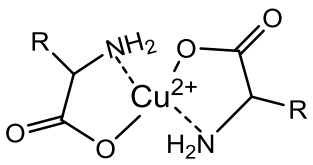
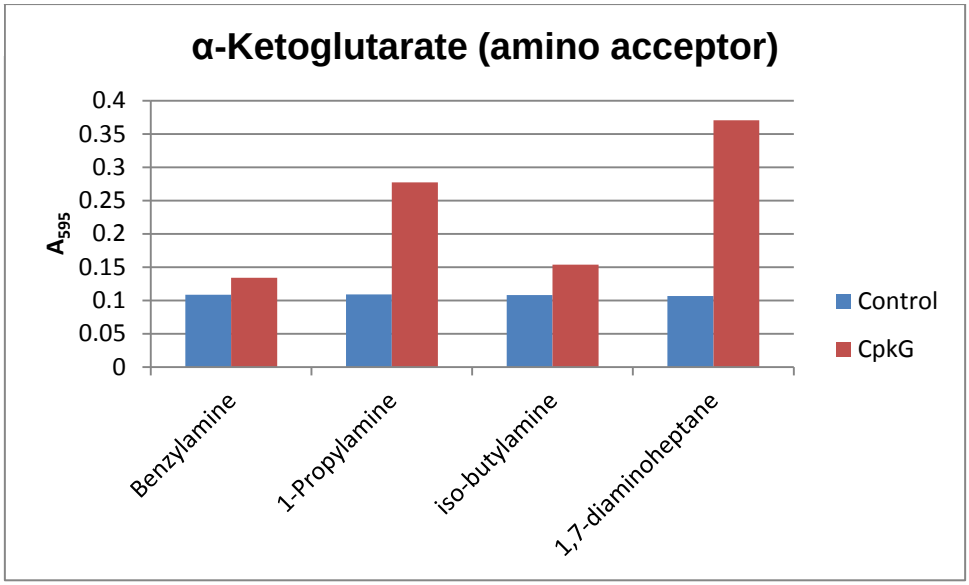


Figure 37: Structure of Cu (II) complex with amino acids with UV λ_{max} of 595 nm.[86,87]

A change in the absorbance was observed for each amino donor chosen to encompass aromatic, straight and branched chain amines showing that CpkG is an ω-aminotransferase, capable of utilizing amino group donors that do not contain a carboxylic acid functional group. CpkG was again shown to accept both α-KGA (**68**) and pyruvate (**67**) and not be strictly specific for α-KGA (**68**) contrary to the prediction made by the bioinformatics analysis in **section 3.1** above. This shows that *in silico* analysis for prediction of transaminase function is limited and there is a need for additional *in vitro* assays and structural studies to confirm enzymatic function predicted by bioinformatics analysis.

a



b)

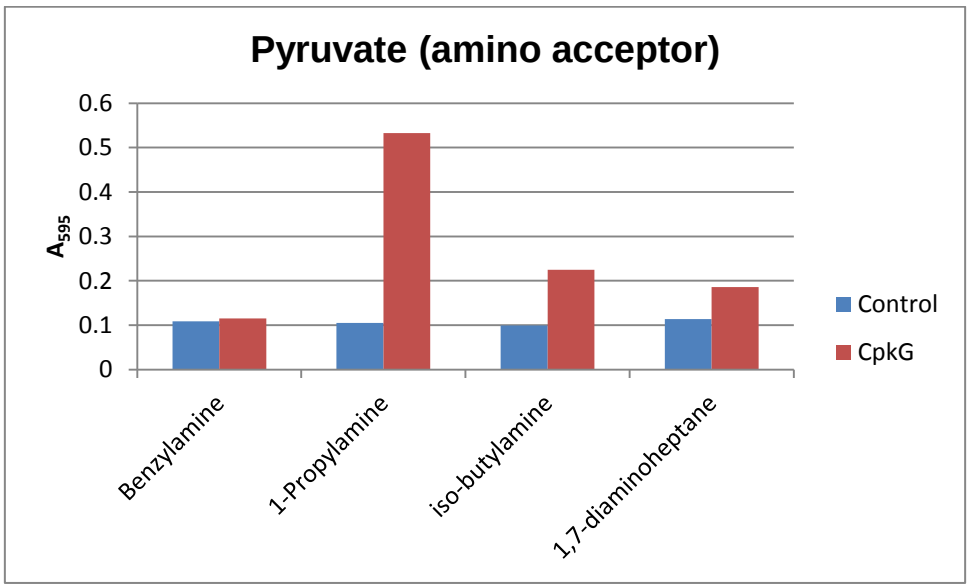


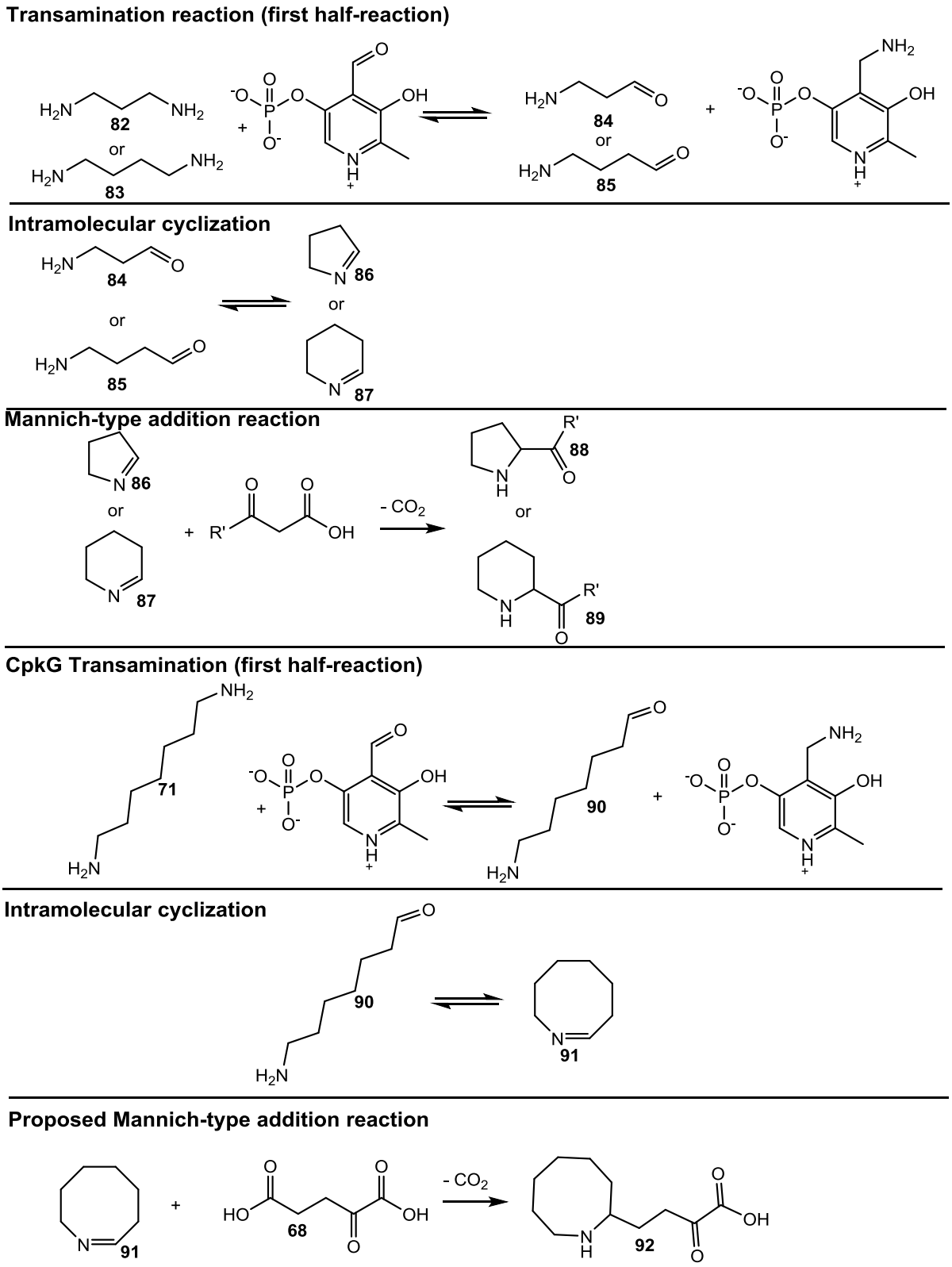
Figure 38: UV absorbance at 595 nm measuring **a)** glutamate and **b)** alanine Cu^{2+} complex formation.

For both amino acceptors (α-KGA and pyruvate) the change in absorbance at 495 nm for their reactions with benzylamine and isobutylamine were modest. This is very likely to be as a result of steric restrictions at the active site entrance.[58] Previous studies have shown that ω-aminotransferases which are specific for either α-KGA or pyruvate adopt distinct

mechanisms for substrate binding - one employs a gateway system (α-KGA-specific enzymes) and the other a bipartite substrate binding site that controls specificity by steric constraints and hydrophobic amino acid residues (pyruvate-specific enzymes).[58] It has however been observed, that exceptions to these mechanisms exist where a hybrid or intermediate system is utilized.[58] CpkG appears to be such a hybrid system, enabling it to utilize both α-KGA and pyruvate as amino group acceptors with ω-amines (amino group donors that do not bear an α-carboxylate functional group) as amino group donors. From the bioinformatics analysis, (**section 3.1**) the glutamate residue used as a functional switch in α-KGA-specific ω-aminotransferases is conserved. However, a second key arginine residue used that interacts with the γ-carboxylate of α-KGA is replaced by a lysine (K204 in CpkG and K183 in *E. coli* putrescine aminotransferase), as observed in another ω-aminotransferase: *E. coli* putrescine aminotransferase.[58] This amino acid substitution in CpkG may account for its activity towards aliphatic amines that do not contain a carboxylate functional group.[58] In pyruvate-specific ω-aminotransferases, the structural restrictions within the bipartite substrate binding site prevent the enzyme from selecting both substrates where one substrate bears an α-carboxylate and the second substrate bears a large side chain.[58] For the reaction to occur, a substrate with an α-carboxylate would only react with a second substrate bearing a minimal side chain.[58,88] On the other hand, a substrate with a bulky side chain would only react with a second substrate having no α-carboxylate (an aldehyde, ketone or amine).[58,88,89] This accounts for the modest enzymatic activity of CpkG towards the bulky amino group donors benzylamine and isobutylamine in the presence of α-KGA and pyruvate (since both amino group acceptors possess an α-carboxylate functional group).

The highest change in absorbance compared to a control was observed in the mixture containing pyruvate (**67**) and propylamine (**69**); while for the set of reactions utilizing α-KGA, the highest change in absorbance occurred when 1,7-diaminoheptane (**71**) was used as an amino group donor (**Figure 38 a** and **b**)**.** In both cases, the amino group donors

are straight chain amines without bulky side chains, the only difference being that one of them is an α,ω-diamine. The product of the first half of the transamination reaction of other naturally occurring α,ω-diamines (such as putrescine and cadaverine) is an ω-aminoaldehyde which spontaneously (and reversibly) cyclizes to the corresponding nitrogen-containing heterocycle (1-pyrroline or 1-piperideine) by the formation of an intramolecular imine bond.[90] It is therefore, conceivable that a similar transformation (resulting in the formation of the eight-membered hexahydro-1-azocine) is occurring in both reactions when CpkG is incubated with α-KGA and pyruvate. The higher activity observed in the presence of α-KGA may be as a result of a side reaction between the azocine intermediate and α-KGA. A decarboxylative Mannich-type addition reaction has been reported to occur between N-heterocyclic rings (1-pyrolline and 1-piperideine) and 3-keto acids yielding 2-substituted heterocyclic rings.[90,91] Since, the equilibrium for the first half-transamination reaction for related enzymes (putrescine aminotransferases) that utilize α,ω-amines lie on the side of aldehyde formation due to the spontaneous product cyclization,[58] it is expected that the azocine intermediate (**Scheme 28**) will accumulate in sufficient concentrations for the Mannich-type side reaction to be feasible. The likelihood of occurrence of this reaction further increased due to the concentration of the amino group acceptor (α-KGA) which was in four-fold excess relative to the amino group donor. This irreversible side reaction would drive the transamination half-reaction further to the right resulting in a higher concentration of the CpkG-PMP complex which would in turn readily convert a higher concentration of α-KGA to glutamate in the second transamination half-reaction. Additional experiments to detect the product of the addition reaction would therefore, be needed to provide supporting evidence for its formation. The absence of the additional driving force afforded by the Mannich-type reaction in the case of the pyruvate containing reaction may account for the observed difference in the concentration of the amino acid produced.

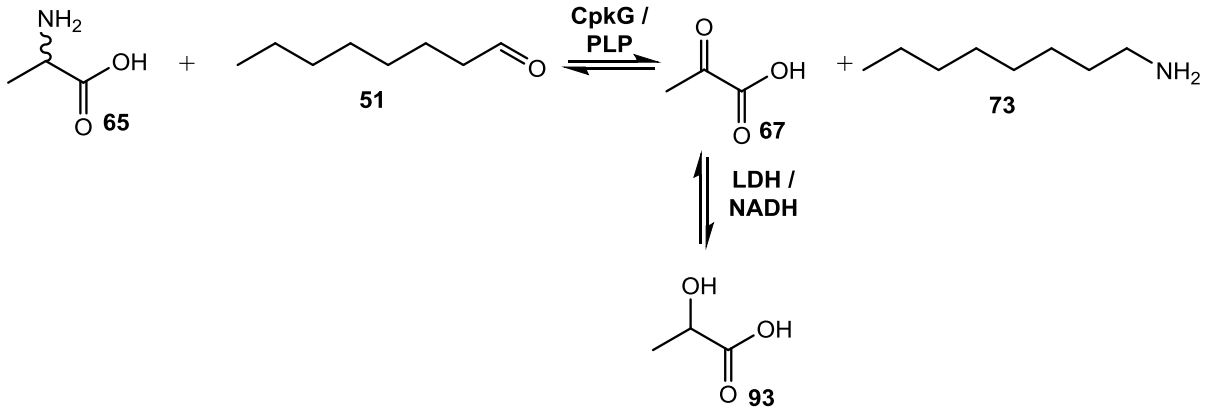


Scheme 28: Proposed intramolecular cyclization followed by Mannich-type addition of α-KGA.[90,91]

A structural steric constraint may account for the overall preferential utilization of the pyruvate / propylamine combination.[58] In addition, to the arginine to lysine mutation in the substrate binding site mentioned above, the results of a structural study of the closely related *E. coli* putrescine aminotransferase revealed that the active site entrance appears to be narrower and the amino acid residues are mostly hydrophobic in nature.[58] It would therefore, be easier for smaller and non-polar substrates to readily diffuse into the enzyme active site.

3.3.3 Production of Pyruvate in the Forward Reaction

An enzyme coupled assay was set up to detect the pyruvate (**67**) produced in the forward reaction (**Scheme 29**) by incubating octanal (**51**) with amino acid (L-and D-alanine) in the presence of CpkG, using NADH (**18**) to reduce the pyruvate (**67**) to lactic acid (**93**) in the presence of the enzyme lactate dehydrogenase (LDH).



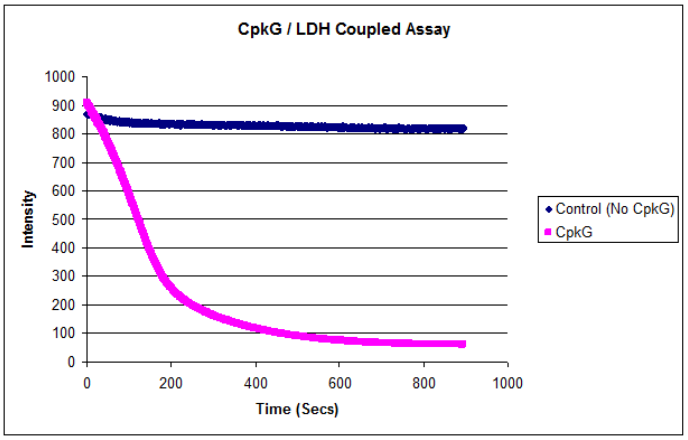
Scheme 29: Lactate dehydrogenase coupled assay for the detection of pyruvate formation during transamination of octanal (**51**) by CpkG using L- and D-alanine (**65**) as the amino group donor.

The reaction was monitored by measuring the decrease in fluorescent emission of NADH (**18**) as it is consumed (λ_{em} = 462 nm). A control sample with no CpkG added showed no change in the initial intensity of NADH fluorescence. Both L-alanine and D-alanine (**65**)

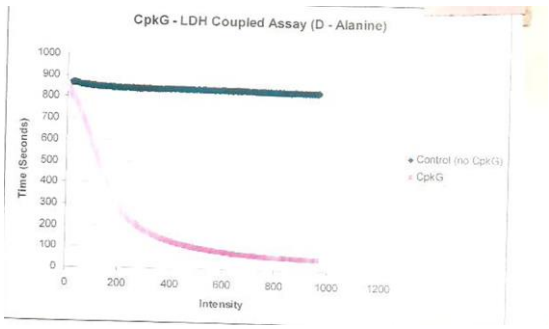
where accepted as substrates by CpkG. Substitution of octanal (**51**) with 2-octanone (**53**) yielded similar results (**Figures 39 a** to **c**).

These preliminary results indicate that CpkG accepts 2-octanone as a substrate in the presence of L-alanine, additional experiments need to be carried out to detect the expected product (2-octylamine) and determine the sterochemical outcome of the reaction (by preparing and analysing Mosher's amide derivatives).[78] This will provide useful data in support of on-going research focusing on characterizing ω-aminotransferases for bio-catalytic production of chiral building blocks for applications in the fine chemicals industry.[92] Useful applications already exist (for example in the production of sitagliptin an anti-diabetic drug)[93] and development of additional ω-aminotransferases with wide ranging substrate specificity and superior activity and conversion efficiency under various industrial conditions is desirable. Enzymatic cascade reactions of industrial relevance can also be further developed from studies involving the stereoselective ω-aminotransferases. For instance the conversion of α,ω-diamines to their corresponding piperidines and pyrolidines has also been applied to the synthesis of *N*-heterocyclic alkaloids.[91] *(R)*-selective ω-transaminases showing activity towards aromatic and bulky substrates are also needed as they are less abundant in nature. Insights gained from structural studies have shown that it is possible to introduce mutations in the enzyme active site in order to improve access of aromatic and bulky aliphatic substrates without compromising enzymatic activity significantly. It will, therefore, be necessary to continue to generate more data from research on this industrially important class of enzymes.

a)



b)



c)

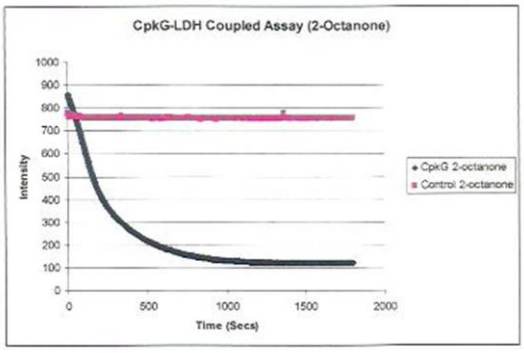


Figure 39: Fluorescent emission at 462 nm of NADH following incubation of CpkG with **a)** octanal and L-alanine **b)** octanal and D-alanine and **c)** 2-octanone and L-alanine. Wavelength for absorption: 340 nm; and emission 462nm. Enzyme concentration: 20 μM, NADH, octanal and alanine concentration: 200 μM.

3.3.4. Production of Amines

The enzymatic reaction used for the LDH coupled assay was also analysed for 1-octylamine (**73**) formation by ESI-LC-MS. The reaction was quenched with methanol to precipitate the protein and the supernatant was analysed by LC-MS in positive ion mode.

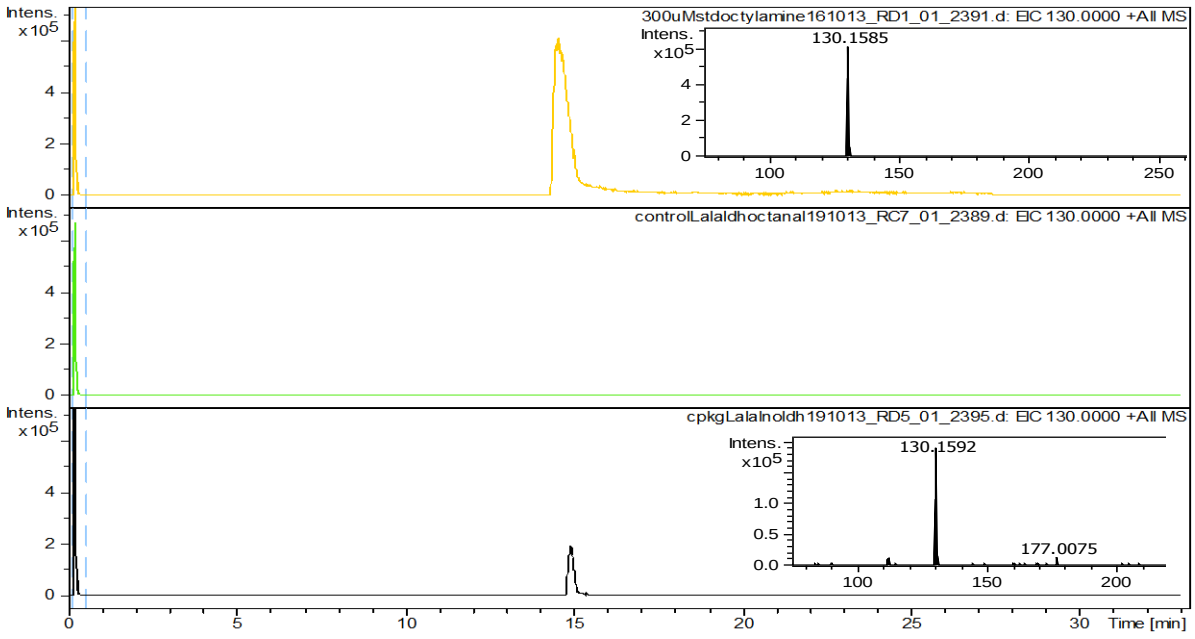


Figure 40: Extracted Ion Chromatograph at *m/z* = 130 from the LC-MS analysis of His_6-CpkG catalysed reactions. Top: octylamine (**73**) standard, Middle: control reaction with no added enzyme, bottom: CpkG catalysed reaction.

A peak with a retention time of 15 mins (**Figure 40**) was observed with *m/z* = 130 which is absent in the control reaction from which the enzyme was omitted. This new peak has the same retention time and *m/z* value as an authentic standard of 1-octylamine (**73**). The molecular formula of the product was confirmed by HR-MS (*m/z* calculated for $C_8H_{20}N^+$ 130.1596, observed 130.1592) consistent with the formation of octylamine (**73**) in the enzymatic reaction.

3.4 Conclusion

Pyridoxal-5'-phosphate dependent enzymes play important roles in the biosynthesis of various PKS and NRPS derived secondary metabolites. These enzymes are involved in the biosynthesis of the starter unit for the assembly of rifamycin (**14**).[94] They have also been found to operate *in cis* with the multi-modular enzyme machinery, catalysing chain termination in prodiginine (**41**) and fumonisin (**37**) biosynthesis.[31,34] In addition to the above functions, PLP-dependent enzymes are employed in post-PKS / NRPS tailoring reactions in oxytetracycline (**38**) and myxochelin (**27**) biosynthesis.[49,64] The diversity of reactions catalysed by PLP-dependent enzymes (which include C-C, C-S, C-O and C-N bond formation and bond cleavage)[95,96,97] and the possibility for their recruitment in polyketide biosynthetic pathways provides an additional layer of structural diversification to these biologically relevant chemical compounds.[98,99,100] This study shows that a *bona fide* PLP-dependent ω–aminotransferase CpkG is required for introduction of the nitrogen atom into coelimycin P1 (**45**). Its ability to use a wide range of substrates as amino group donors (α-amino acids and ω–amines) and amino group acceptors (pyruvate, α-ketoglutarate, aldehydes) has also been demonstrated.

CHAPTER FOUR

4.0 SUMMARY, CONCLUSIONS AND FUTURE WORK

4.1 Characterization of the mechanism for amine introduction in polyketide alkaloids

The *cpk* biosynthetic gene cluster encodes a type I PKS with a terminal domain predicted to catalyse chain release of the mature polyketide intermediate by an NAD(P)H – dependent reduction of the thioester bond between the polyketide chain and the ACP. This reduction was expected to yield an aldehyde product which would undergo subsequent modifications including a crucial PLP-dependent reductive amination to a primary amine; this amine functional group is critical to the intramolecular cyclization step that results in the formation of the six-membered piperidine ring in coelimycin P1.[7]

Prior to this study, similar NAD(P)H- dependent reductive release of thioester-bound intermediates had been reported in related biosynthetic systems (NRPSs and fungal PKSs). However, direct biochemical evidence for utilization of this mechanism in bacterial type I modular PKSs was lacking. In certain cases (the myxochelin NRPS and the zeamine type I iterative FAS/PKS),[40,49] a subsequent reductive amination of the aldehyde product occurs to yield the corresponding primary amino group which does not undergo further modification. During the course of this study, it has been shown that actinobacterial polyketide-alkaloid biosynthesis involves three conserved steps: TR-mediated chain release, reductive amination of the resulting aldehyde and cyclization to form an imine between the terminal amino group and a keto group at the Ɛ-position of the polyketide chain.[8]

Characterization of the Cpk-TR domain was achieved by *in vitro* biochemical experiments using a fluorimetric assay, which measured consumption of NAD(P)H. The Cpk-TR domain was shown to preferentially utilize NADH instead of NADPH; this differentiates the CpkC TR domain from other biochemically characterised TR domains.

Octanoyl-CoA (a commercially available analogue of the fully assembled ACP-bound polyketide chain) was tested as a substrate for the purified recombinant TR domain. Initial experiments to detect the aldehyde product were not reproducible; probably as a result of its transient nature. This aldehyde would be rapidly converted to the corresponding primary amine in the presence of the PLP-dependent transaminase (CpkG) *in vivo* (as predicted by bioinformatics and structural studies of coelimycin P1). The aldehyde was rapidly converted to the corresponding alcohol which reproducibly was detectable in the *in vitro* studies. These results indicate that the TR domain in Cpk was responsible for NADH-dependent reductive release of the fully-assembled polyketide chain as an aldehydewhich can be further reduced to the corresponding alcohol.

Further experiments to probe the kinetics of the reduction of octanoyl-CoA might provide insight into the rate of aldehyde formation relative to its further reduction to the alcohol and inform the design of experiments that could result in the detection of the aldehyde product either directly, or by trapping the aldehyde by derivatization. Since it has been shown in this study, that the TR domain also accepts ketones (2-octanone), it could potentially find applications in the enzymatic synthesis of chiral alcohols. The stereochemical course of the reaction could be probed, using the stereoselectively deuterated NADH synthesized in this study, to identify which diastereotopic hydrogen atom of NADH is transferred to the substrate. Conversion of the secondary alcohol product to a Mosher's ester and subsequent analysis could be used to determine the stereochemistry of the chiral alcohol resulting from the TR catalysed reaction.

An additional key step in polyketide-alkaloid formation that was identified during this study is the PLP-dependent, reductive amination of the aldehyde produced by the TR domain to yield the corresponding primary amine. Both the forward and the reverse transamination reactions were characterised using substrate analogues. 1-Octylamine was readily detected by LC-MS analysis, as the product of the reaction between octanal (as the

amino acceptor) and either L- or D-alanine (as the amino donor); formation of the pyruvate co-product of this reaction was followed using a coupled enzymatic reaction (with lactate dehydrogenase catalysing the NADH-dependent reduction of pyruvate to lactate). Conversion of an aldehyde to the corresponding primary amine is the reaction of most significance to this study because it provides experimental evidence to support the initial biosynthetic hypothesis made based on detailed structural and bioinformatics analysis of coelimycin P1 and its associated biosynthetic gene cluster.

The stereospecificity of the PLP-dependent transaminase, CpkG has subsequently been determined by Dr. Jade Ronan by analysing the reverse reaction (octylamine reacting with pyruvate). By derivatization of the amino acid formed (using Mosher's acid) and subsequent analysis by NMR spectroscopy it was shown that even though, CpkG accepted both D- and L-alanine in the forward reaction, only L-alanine is produced in the reverse reaction.[8]

4.2 Biosynthetic Mechanism for the Assembly of Polyketide Alkaolids

After the conclusion of this study, bioinformatics analysis of the genome sequences of actinobacterial species has showed that thioester reduction and aldehyde transamination are a universally conserved steps in the biosynthesis of polyketide-alkaloids.[8] In actinobacterial genomes alone, 22 biosynthetic gene clusters were found to encode a type I modular PKS containing a thioester reductase domain and an ω-aminotransferase.[8] The structures of the metabolic products of these gene clusters were predicted, showing that some of them are likely to direct the assembly of previously isolated polyketide-alkaloids (**Figure 41**) coelimycin P1 (**45**), pyridincin (**94**), latumcidin (**95**), streptazolin (**96**), cyclizidine (**98**), indolizidine (**99**) and nigrifactin (**101**).[8] The metabolic products of nine of the biosynthetic gene clusters analysed could not be predicted based on bioinformatics analysis alone and were hypothesized to direct the biosynthesis of novel polyketide-alkaloids.[8]

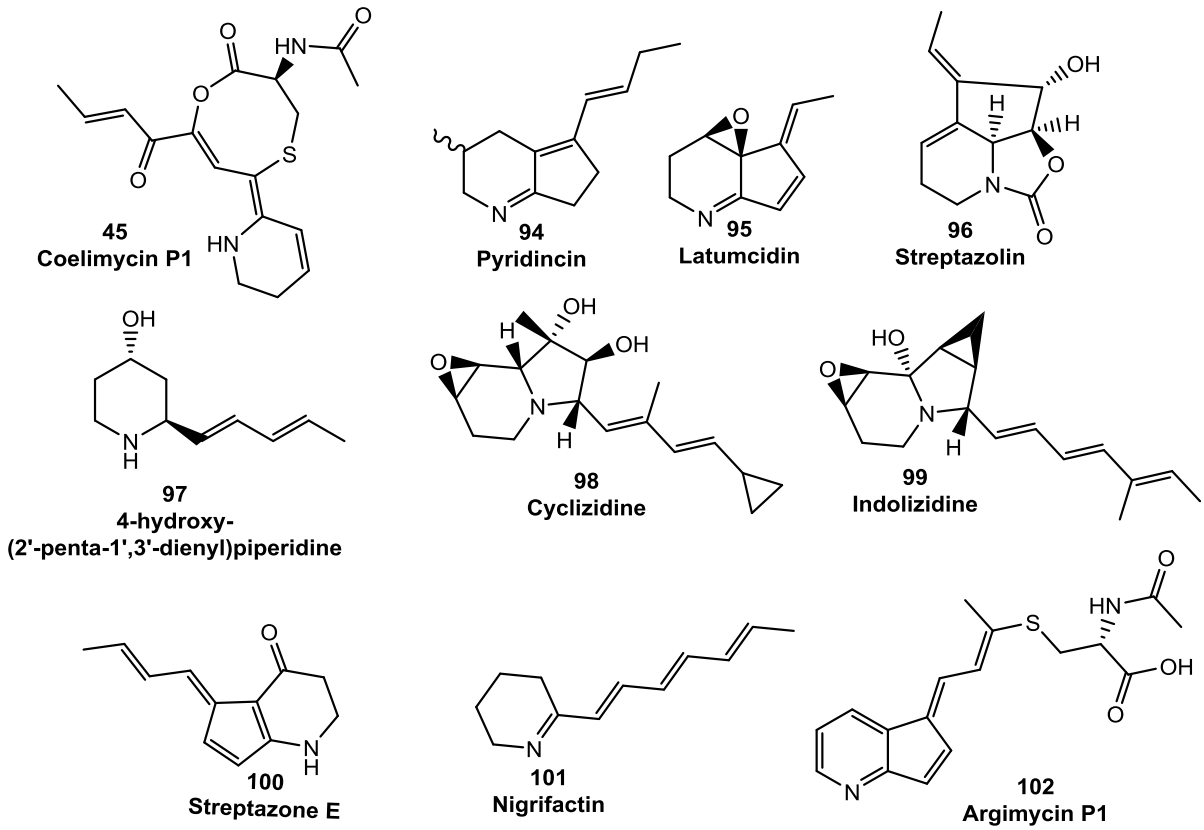
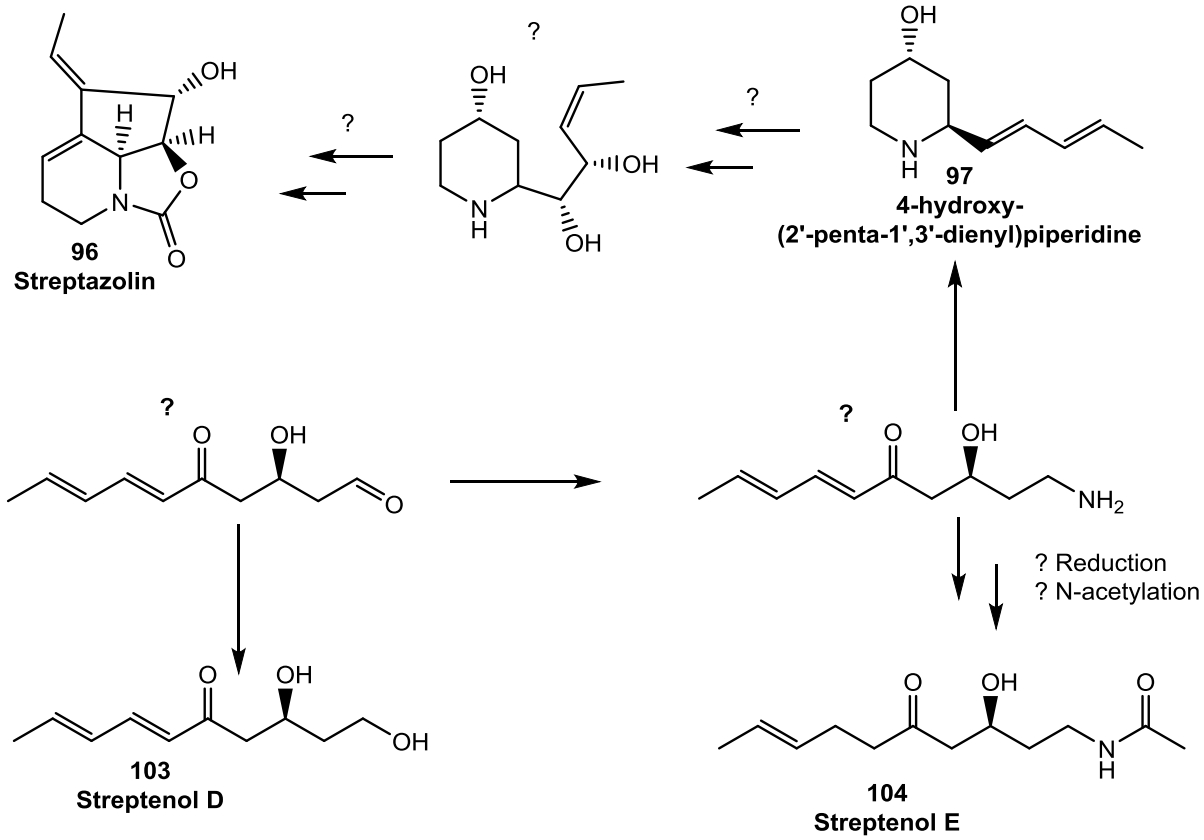


Figure 41: Representative structures of some polyketide alkaloids of actinobacterial origin.

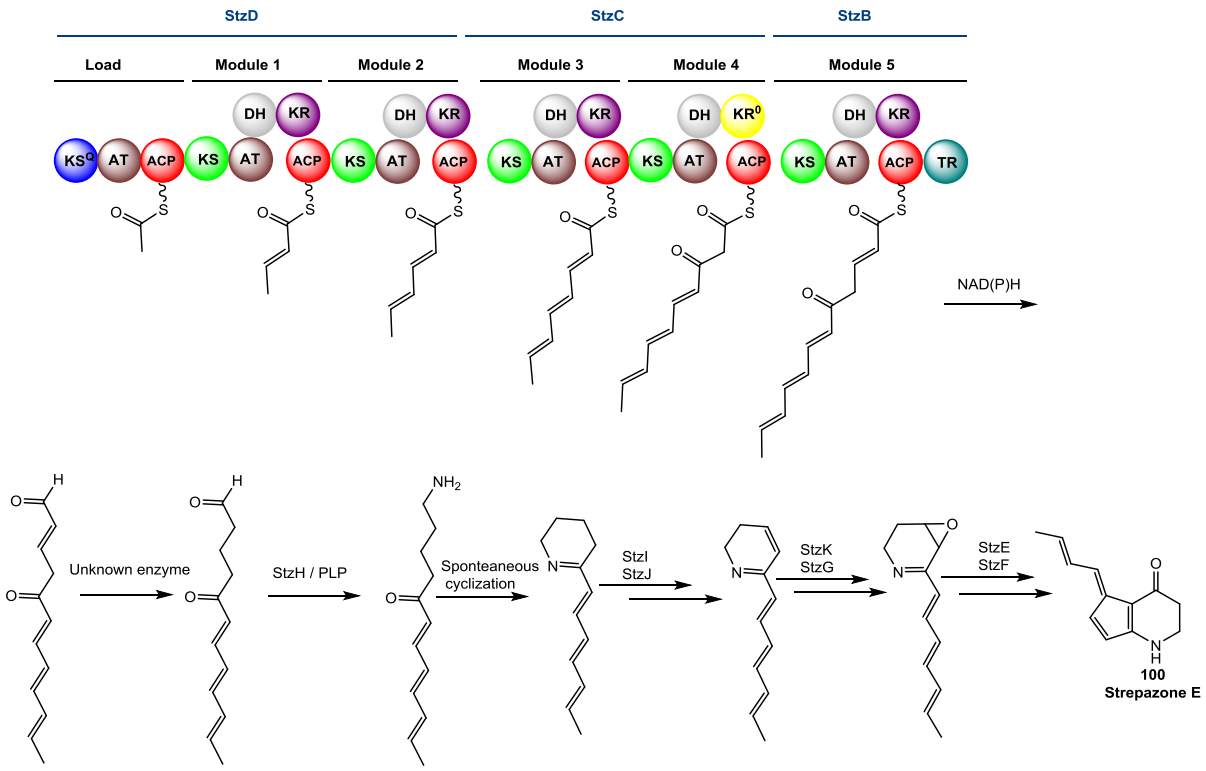
The structurally related piperidine alkaloids: pyridincin (**94**) and latumcidin (**95**) were isolated from *Streptomyces griseoflavus* var. *pyridicus* and *Streptomyces reticuli* var. *latumcidicus* respectively.[101,102] They have both been shown to be polyketide-derived by feeding experiments using ^{13}C-labelled precursors.[101,102] Isolation of biosynthetic intermediates and identification of the gene clusters encoding the PKSs involved in their biosynthesis was yet to be achieved at the time this study was carried out. Biosynthetic gene clusters for their assembly were subsequently identified by bioinformatics analysis of the genomes of: *Streptomyces venezuelae* (pyridincin), *Streptomyces fulvissimus* (latumcidin) based on the results from this study.[8]

In the 1990s, studies using isotope labelled precursors, revealed that the streptazolins are likely to be polyketide-derived, in common with the previously studied polyketide-alkaloids mentioned above.[103] Also, the nitrogen in the piperidine ring was shown to originate from glutamate.[103] Two, plausible biosynthetic precursors (**Scheme 10**) were isolated, alongside streptazolin (**96**), from the producing strain (*Streptomyces* sp. strain FH-S 2184): streptenol D (**103**) and 4-hydroxy-(2-penta-1',3'-dienyl) piperidine (**97**).[103] Additional studies on another streptazolin-producing *Streptomyces* strain (*Streptomyces* sp. Strain A1), resulted in the isolation of other biosynthetic precursors including: streptenol E (**104**).[103] Given that little or no information about the gene clusters (responsible for the assembly of polyketide-alkaloids of actinobacterial origin) was available at the time, prediction of the biosynthetic steps leading to the formation of these precursors or co-metabolites was not very straightforward.[7] Streptenol D (**103**) appears to be the product of a type I PKS which has been released by reductive cleavage to form the aldehyde and then subjected to an additional round of reduction to yield the alcohol. The aldehyde precursor to streptenol D (**103**), may undergo reductive amination in the presence of an aminotransferase to yield the corresponding amine *en route* to the final product. Streptenol E (**104**) appears to be an N-acetylated derivative of the amine intermediate.[103,104] The intermediate: 4-hydroxy-(2-penta-1',3'-dienyl) piperidine (**97**), would be the product of spontaneous cyclization of the amine intermediate, which would have undergone reduction of the imine bond, formed at the point of cyclization.[7,103] Additional steps which are yet to be characterized would ultimately lead to the formation of streptazolin (**96**). The streptazolin biosynthetic gene cluster has recently been identified in *S. fulvissimus*;[8] further research would likely reveal the exact nature and order of the biosynthetic steps responsible for the assembly of this metabolite.



Scheme 30: Shunt metabolites and putative intermediates in streptazolin (**96**) biosynthesis isolated from streptazolin-producing *Streptomyces* species.[103,104]

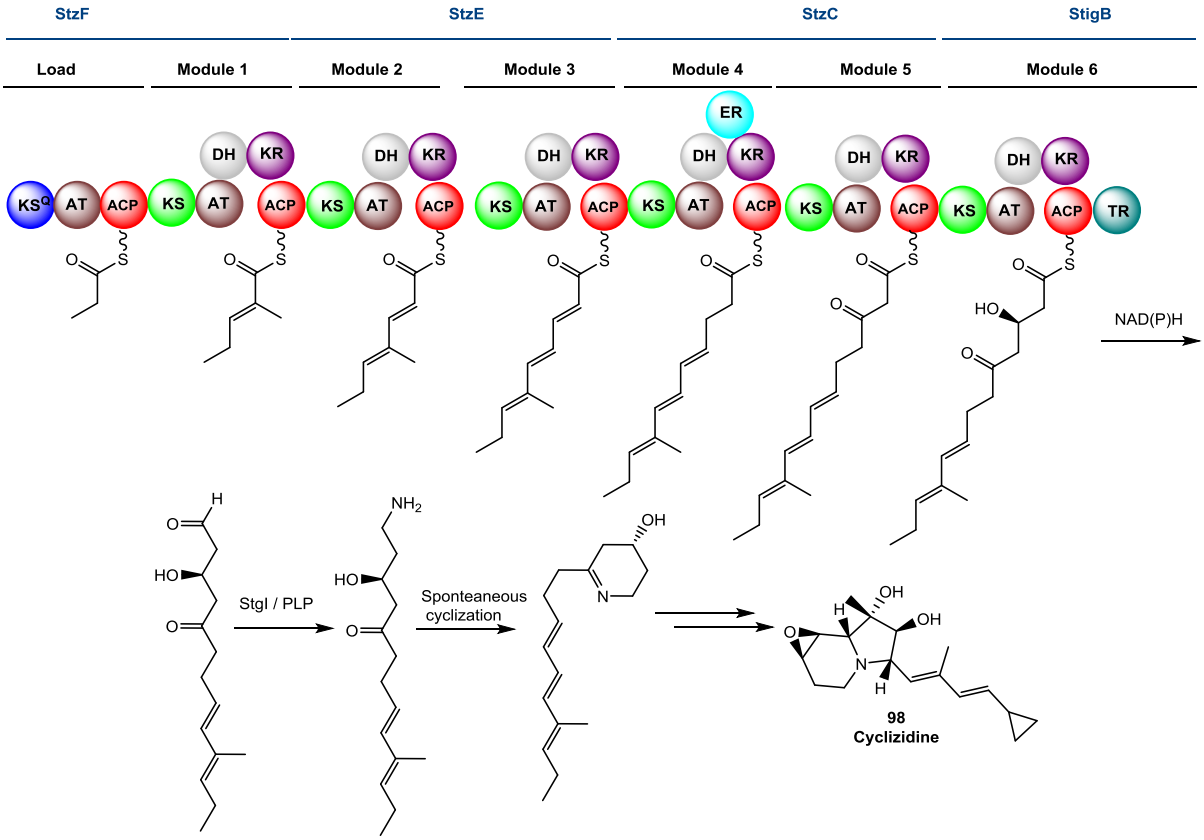
Subsequent characterization of the streptazone type biosynthetic gene cluster by Ohnishi and co-workers, revealed an almost identical biosynthetic pathway save for an additional PKS module resulting in a hexaketide intermediate as shown (**Scheme 31**).[105]



Scheme 31: Module and domain organization of the streptazone E type I modular PKS showing the conserved biosynthetic mechanism for nitrogen incorporation.[105]

Previous biosynthetic studies during the course of isolation and structural elucidation of the indolizidine alkaloids: cyclizidine (**98**) and indolizidine (**99**) from *Streptomyces* species, showed that their carbon skeleton including the piperidine moiety are of polyketide origin. This contrasts with indolizine alkaloids from fungal sources where the piperidine is derived from lysine.[106] Following on from our work the details of the biosynthetic steps for the assembly of indolizidine alkaloids from Actinobacteria were confirmed to involve the use of a similar biosynthetic logic (**Scheme 31**).[108] Qu and co-workers have now re-isolated the indolizidine-alkaloid cyclizidine (**98**) from *Streptomyces* sp. NCIB 11649 and biochemical studies using the purified enzymes encoded by its biosynthetic gene cluster, showed that a combination of thioester reduction, transamination and imine reduction is necessary for their

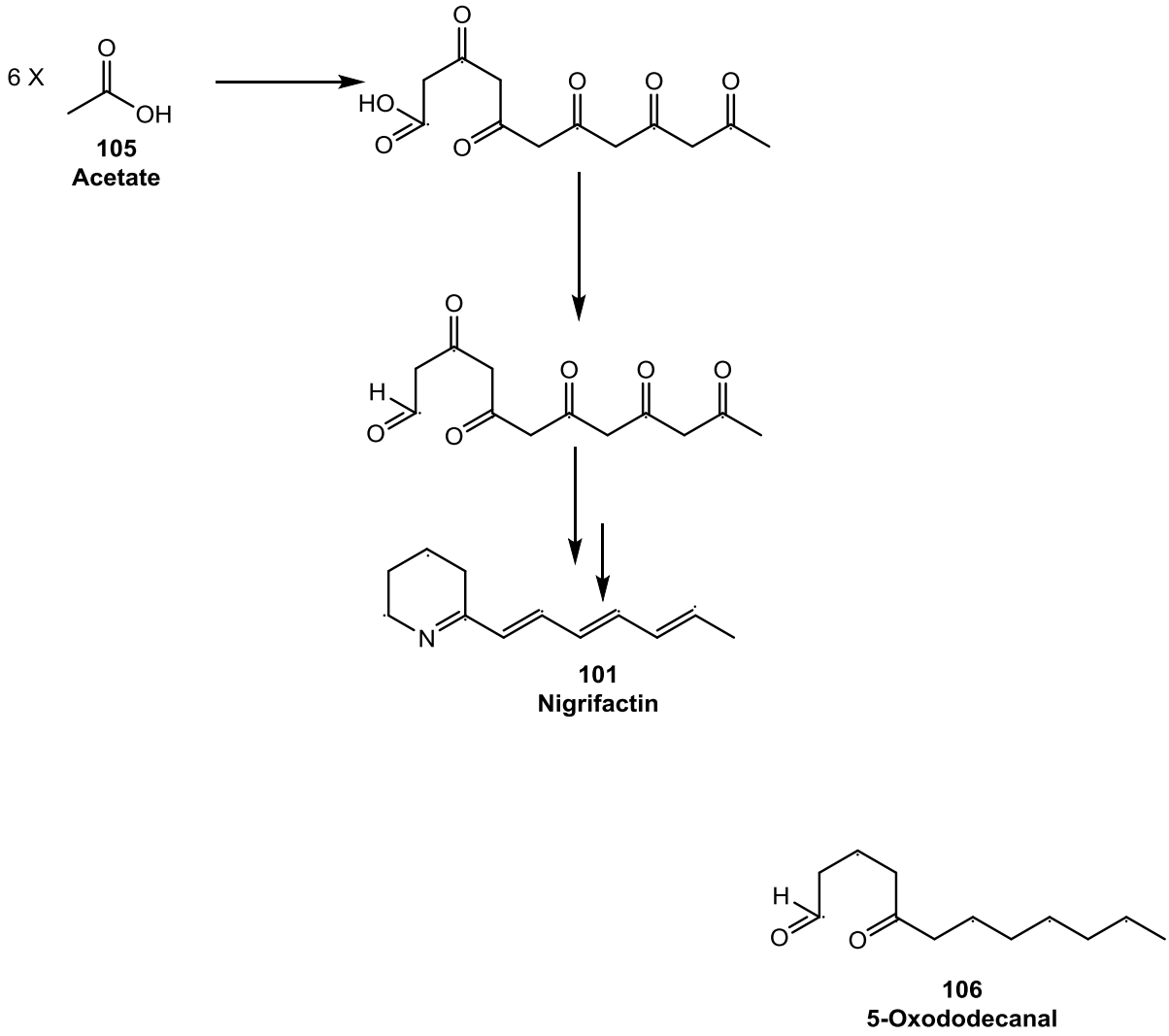
assembly.[97] The cyclizindine biosynthetic gene cluster was identified in *Streptomyces* sp. NCIB 11649 as well as in another actinobacterial species: *Actinosynnema mirium* by bioinformatics analysis.[8]



Scheme 32: Module and domain organization of the cyclizidine type I modular PKS showing the conserved biosynthetic mechanism for nitrogen incorporation.[108]

Nigrifactin (**101**) a simple piperidine alkaloid isolated from *S. nigrifaciens* var. FFD-101; possesses antihistamine and antihypertensive activity. [30,109] Studies carried out in the 1970s to probe its biosynthetic origin suggested that its carbon skeleton was derived from 6 acetate (**105**) units in a linear combination similar to the assembly of plant-derived alkaloids which had been shown to be of polyketide origin.[109] Labelled biosynthetic intermediates including the aldehyde 5-oxododecanal (**106**) predicted to be precursors of nigrifactin (**101**)

were also synthesized and used for further feeding experiments.[109] The aldehyde (**106**) synthetic analogue of the intermediate was incorporated into the re-isolated nigrifactin (**101**) probably via degradation to labelled acetate units prior to incorporation (**Scheme 33**).[109] Comparison of the structure of nigrifactin (**101**) with the biosynthetic precursor to coelimycin P1 (**45**), shows that both metabolites are likely to be derived from a closely related thioester intermediate (save for the absence of the last double bond in the predicted nigrifactin intermediate).[8] It would therefore, be logical to hypothesize, that reductive cleavage of the intermediate would yield a structural analogue of 5-oxododecanal (**106**). Hence, it would be expected to find a TR domain in the nigrifactin PKS. This was indeed the case when nigrifactin (**101**) was re-isolated from the Actinobacterium: *Streptomyces argillaceus* ATCC 12956).[110] A suite of closely related novel polyketide-alkaloids, the argimycins were discovered in the process.[110] Its type I PKS had previously been shown to be identical to the coelimycin PKS,[8] containing a chain-terminating TR domain, and was associated with an ω-aminotransferase utilized for assembly of the piperidine ring present in the indolizidine core of argimycin P1 (**102**).



Scheme 33: Proposed biosynthetic pathway to nigrifactin (**101**) showing the predicted intermediate: 5-oxododecanal (**106**) used for isotope labelling studies.[109]

4.3 Characterization of Bio-catalysts

In addition to understanding the biosynthetic pathway to a structurally intriguing class of secondary metabolites, this project initiated studies towards the development of two enzymes which could be developed further as biocatalysts for production of chiral alcohols and chiral amines. A combination of both enzymes (alongside other enzymes to form a biosynthetic cascade) could also be used for the synthesis of compounds containing a piperidine ring, which is often found in biologically active compounds of pharmaceutical and

agricultural interest. Both recombinant proteins are readily produced in stable, soluble form in *E. coli* in reasonable quantities and have exhibited broad substrate tolerance, including towards the pro-chiral substrate: 2-octanone. With the current heightened interest in green and sustainable methods for chemical synthesis on an industrial scale[111,113] these results may provide impetus for additional studies and scale-up. It has been shown that Actinobacteria not only represent a vast and largely untapped resource for the discovery of biologically active compounds of agricultural and pharmaceutical interest, but detailed studies of the steps employed in Nature for the assembly of these structurally diverse secondary metabolites will potentially provide an array of enzymes which will be invaluable components of the biocatalytic toolbox of the industrial synthetic chemist in the not so distant future.[111,112,113]

CHAPTER FIVE

5.0 MATERIALS AND METHODS

5.1 Chemicals and Reagents

All chemicals were purchased from Sigma-Aldrich (U.S.A), Fischer Scientific (U.K.), Fermentas (Lithuania) and BioLab (New Zealand) unless otherwise stated. Deionised water was used for all experiments unless otherwise stated.

5.2 Instruments and Equipment

An Eppendorf-Mastercycler-Personal was used to perform PCR. A NanoDrop spectrophotometer (Thermo Scientific) was used to measure DNA and protein concentrations. A BioRad tank and PowerPac 300 was used to carry out agarose gel electrophoresis. Gels were visualized using a UVP imaging system UV transilluminator. Centrifuge Eppendorf models 5810 R and 5424 were used. Protein purifications were carried out by FPLC using an Akta Purifier (GE Healthcare) fitted with a 1 mL HiTrap Ni-NTA sepharose affinity column (GE Healthcare). Fluorimetric analysis was carried out using a Perkin Elmer LS 50B luminescence spectrometer with a 3 mm path length quartz cuvette (Starna Scientific). Samples were submitted to the University of Warwick mass spectrometry service for high-resolution LC-MS analyses.

5.3 Bacterial Strains and Vectors

The following *E. coli* strains were used during this study:

- DH5α (lab stock)
- Chemically competent One Shot® Top10 (Invitrogen, U.S.A.)
- Chemically Competent BL21 Star™ (DE3) One Shot® (Invitrogen, U.S.A.)

The following vector was used during this study:

- pET151 – it was used for the directional cloning and overexpression of proteins with an N-terminal histidine tag. It was included in the Champion™ pET Directional TOPO® Expression kit (Invitrogen, U.S.A.).

5.4 Bacterial Growth Media and Culture conditions

All bacteria cells (*E. coli*) were grown in an incubator (as agar plates in petri dishes) or in a New Brunswick Scientific Innova shaker at 180 r.p.m. at 37 ^{o}C and 15 ^{o}C for protein overproduction.

Luria-Bertani Medium (LB)

Tryptone (10 g/L), yeast extract (5 g/L) and NaCl (10 g/L) were dissolved in water and sterilized by autoclaving prior to use. For LB agar, 15 g/L of Bacto agar was added to the mixture and the warm semi-solid mixture was poured into petri dishes in a fume cupboard and allowed to cool and solidify.

5.5 Isolation of Plasmid DNA (Miniprep)

A 5 mL liquid LB culture of the *E. coli* strain harbouring the plasmid of interest was prepared by inoculation with a loop full of glycerol stock solution and incubated at 37 ^{o}C overnight. The liquid culture was then centrifuged at maximum speed for 15 min at room temperature to precipitate the cell pellet. The supernatant was discarded and the pellet re-suspended in 100 µL of cold solution I (25 mM Tris-HCl pH 8, 10 mM EDTA pH 8, 50 mM glucose stored at 4^{o}C). After re-suspending the pellets, the 200 µL of cold solution II (0.2M NaOH, 1% SDS, stored at room temperature) was added to lyse the cells, the resulting suspension was mixed gently for 5 min. Genomic DNA and protein precipitation was achieved by adding 15 µL of cold solution III (60 mL 5M Potassium acetate, 11.5 mL glacial acetic acid, 28.5 mL H_2O, stored at 4 ^{o}C), the suspension was mixed by inverting the

Eppendorf tube 10 times. The suspension was centrifuged at maximum speed for 10 min at 4 ^{o}C to remove the cell debris and the supernatant was transferred to a fresh microfuge tube. A 200 µL aliquot of a mixture of phenol: chloroform: *i*-propyl alcohol (25:24:1) was added to the supernatant and the tube was vortexed for 30 seconds for thorough mixing. The mixture was centrifuged again and the aqueous layer was transferred to a fresh Eppendorf tube to which 500 µL of ice-cold *iso*-propanol was added to precipitate the plasmid DNA. The tube was mixed again and centrifuged for 20 min at 4 ^{o}C. The supernatant was discarded and the plasmid DNA was washed with 100 µL ice-cold 70% ethanol. The ethanol was removed carefully and the plasmid DNA was allowed to dry before being re-suspended in 35 µL deionized water containing RNase the mixture was incubated at 37 ^{o}C for 30 minutes prior to storage for downstream applications.

5.6 Polymerase Chain Reaction

Primers for the polymerase chain reaction (see **Table 1**) where designed by obtaining sequence information from the gene sequences of interest. For the ACP-TR di-domain and TR domains, the cpkC gene was analysed using PKS-NRPS analysis software to determine the putative domain regions. The CACC sequence was incorporated into all the forward primers for recognition by the cloning site of pET151/D-TOPO® vector. Primers were ordered from Sigma-Aldrich.

A 50µL PCR reaction mixture was prepared using Readymix Taq PCR mastermix; DMSO was added to a final concentration of 5 % to reduce the annealing interaction of GC rich DNA (from Streptomyces sp), 100 (pmoles / µL) each of the forward and reverse primers as well as 1µL of cosmid DNA (from *S. coelicolor* M145 cosmid library st2C4 and st1G7). The PCR reaction conditions comprised of an initial denaturation step of 94 ^{o}C for 2 min, and 30 cycles of denaturation at 94 ^{o}C for 45 s, annealing at 60 ^{o}C for 45 s and extension at 72 ^{o}C for 90 s with a final extension round at 72 ^{o}C for 5 mins.

Table 1: PCR Primers

Primer Name	Sequence 5' to 3'
*cpkC*_ACP-TR_fw	CACCGGCATCGTGCTGGGGCAG
*cpkC*_ACP-TR_rv	TCACCCGGCCGCCGGGAA
*cpkC*_TR_fw	CACCGGCATCGTGCTGGGGCAG
*cpkG*_fw *cpkG*_rv	CACCATGACCGACCAGCC TCAGCCGATCGGAGCCA

5.7 Agarose Gel Electrophoresis

The PCR products were analysed by agarose gel electrophoresis to determine if PCR resulted in the formation of non-specific amplification products, whether PCR amplicons of expected size were obtained and to separate and purify the target DNA segment for downstream cloning reactions. A 1% w/v agarose gel was prepared by dissolving 1 g of agarose in 1X TBE buffer by heating in a microwave till a clear solution was obtained. The mixture was cooled to a semi-solid consistency and then 5 µL of GelRed DNA staining solution was added to the mixture and poured into a setting tray and allowed to cool at room temperature until completely set (*ca.* 30 min) a comb of appropriate size was inserted into the warm mixture at the point of casting in order to create wells in the gel for each sample. Each PCR sample was mixed with 6X DNA loading dye to a convenient loading volume for the sample wells (final concentration of the loading dye would be 1X). For visual estimation of the DNA fragment sizes, a ladder marker with standard DNA fragments of known sizes was run alongside the samples on the agarose gel. The solid agarose gel was placed in a buffer bath containing 1X TBE and electrophoresis was run at 50 -100 V for 20 – 30 min at room temperature.

The GelRed stained gel was visualised upon completion of electrophoresis via UV light using a UVP UV TransIlluminator and a photo of the gel was taken for further analysis.

5.8 PCR Product Purification from Agarose Gel

Upon confirmation by agarose gel electrophoresis, that the PCR product of expected size was obtained, the portion of the gel containing the DNA fragments of interest was excised and the DNA was purified using a GeneJET^{TM} gel extraction kit adhering to the manufacturer's protocol. A 1:1 weight : volume ratio of the excised gel to the binding buffer (provided in the kit) was added to an Eppendorf tube and incubated at 50-60 ^{o}C for 10 min. to melt the gel, the tube was inverted periodically to ensure complete melting. The solution was then transferred to the GeneJET^{TM} purification column and centrifuged in a microcentrifuge for 30 – 60 s at maximum speed (14 000 rpm) to bind the DNA to the column, after centrifugation, the flow through was discarded. The bound DNA was washed with 700 µL of wash buffer to remove any impurities the column was centrifuged for 1 min at maximum speed after addition of the buffer and after discarding the flow through; it was centrifuged again for an additional minute. Elution was carried out using the elution buffer, 50 µL of the elution buffer was added to the column and allowed to stand for 1 min then the column was centrifuged for 1 min at maximum speed and the flow through was collected.

Quantification of DNA concentration was carried out at 260 nm using a NanoDrop spectrophotometer (Thermo Scientific). Aliquots of the purified PCR product were sequenced to ensure there were no amplification errors.

5.9 Cloning and Ligation

Ligation to the linearized expression vector from the pET151 Directional TOPO® Expression kit (Invitrogen) was carried out according to the manufacturer's protocol. The PCR product was diluted to a convenient concentration or used neat, in order to obtain a 1:1

molar ratio of PCR product to TOPO vector (1 ng of PCR product). Typically, a 5 µL reaction was set up by adding 1 µL of PCR product to 1 µL TOPO vector, 1 µL salt solution and 2 µL sterile distilled water. The reaction mix was mixed gently and incubated for 30 min at room temperature.

5.10 Transformation of Chemically Competent *E. coli* Cells

Chemically competent *E. coli* cells (One Shot® TOP10 Competent cells) from the pET151 Directional TOPO® Expression kit (Invitrogen) were transformed with the plasmids obtained by the ligation reaction obtained above. A 3 µL aliquot of the cloning reaction was added to a vial of One Shot® TOP10 chemically competent cells and mixed gently. The mixture was incubated on ice for 30 min and then the cells were heat shocked for 30 seconds at 42 ^{o}C and immediately transferred to the ice, 200 µL of LB medium was added to the transformation mix which was then incubated at 37 ^{o}C and 180 r.p.m. for 1 hour. After this initial incubation, the mixture was spread on LB agar plates (100 µg/mL) and incubated overnight.

The antibiotic resistant clones were analysed in order to ensure proper insertion into the linearized vector and re-circularization, 4 to 5 colonies were picked and incubated in 200 µL LB medium (supplemented with ampicillin to a final concentration of 100 µg/mL) overnight at 37 ^{o}C and 180 r.p.m. Plasmids were purified using GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific). The bacterial culture was harvested by centrifugation at 8000 rpm at room temperature in a micro-centrifuge and then the supernatant was decanted leaving the cell pellet at the bottom of the tube. The pellet was re-suspended in 250 µL resuspension solution and vortexed to mix, 250 µL of lysis solution was then added and the tube was inverted 4-6 times. The mixture was neutralized using 350 µL neutralization solution and centrifuged at maximum speed (14 000 rpm) in a micro-centrifuge. The supernatant was transferred to the GeneJET spin column and centrifuged at maximum speed for 1 minute.

The column was washed to remove other impurities by adding 500 µL of wash solution and then centrifuged again for 1 min at maximum speed. Purified plasmid DNA was eluted from the column by adding 50 µL elution buffer before centrifugation for 2 minutes at maximum speed to collect the flow through. The plasmid DNA was analysed first by restriction enzyme digests followed by sequence confirmation by Sanger sequencing (GATC Biotech).

Chemically competent *E. coli* BL21 Star (DE3) cells were transformed (as described above for One Shot® TOP10 Competent cells) with the plasmid DNA from one positive clone for expression of each of the three genes of interest.

5.11 Protein Purification and Characterization

5.11.1 Overproduction of Proteins

Protein overproduction in *E. coli* BL21 Star(DE3) was carried out by first of all, inoculating a 10 mL LB culture (supplemented with ampicillin to a final concentration of 100 µg/mL) with a single positive clone. The 10 mL pre-culture was incubated overnight (at 37 ^{o}C with shaking at 180 rpm) and then a 1 mL aliquot was used to inoculate 1 L of LB medium (100 µg/mL) which was incubated for approximately 3 hours (until an OD_{600} of 0.5-0.6 was reached) at 37 ^{o}C with shaking at 180 rpm. The 1 L LB culture was cooled in the cold room at 4 ^{o}C for 30 min prior to induction of recombinant protein expression by the addition of isopropyl-β-*D*-thiogalactopyranoside to a final concentration of 0.5 mM. The induced culture was incubated further at (15 ^{o}C with shaking at 180 rpm) overnight.

5.11.2 Protein Purification by Nickel Affinity Chromatography

Cells from the induced culture were then harvested by centrifugation (for 15 min at 5000 rpm, maintained at a temperature of 4 ^{o}C) and re-suspended in 10 mL washing buffer (50 mM Tris-HCl pH 8.0, 200mM NaCl, 10 mM imidazole, 10% glycerol). The cells were lysed using a high pressure cell disruptor (Constant Systems Ltd. TS-Series Cabinet) at 20,000 psi and the suspension was centrifuged at (10000 rpm for 25 minutes and 4 ^{o}C) to separate the cell debris.

Clear cell-free lysate was applied to a pre-equilibrated (with washing buffer) 1 mL HiTrap Ni-NTA sepharose affinity column (GE Healthcare) and protein purification was carried out by FPLC using an AKTA Purifier (GE Healthcare) at a constant flow rate of 0.45 mL/min. After sample injection, non-specifically bound proteins and impurities were removed from the column using washing buffer (washing buffer volume was twice the volume of injected sample); the absorbance of the eluate was observed at 280 nm to determine the point of complete removal of the impurities (when a constant minimum absorbance was detected for 30 min). His_6-tagged recombinant protein was eluted from the column using a buffer of higher imidazole concentration in order to displace specifically-bound target protein (elution buffer: 50 mM TRIS-HCl pH 8.0, 200 mM NaCl, 200 mM imidazole, 10% glycerol). The eluate absorbance was monitored at 280 nm and fractions containing recombinant protein were pooled.

Protein-containing fractions were filtered through an Amicon Ultra centrifugal filter (Millipore) with a 30 kDa molecular weight cut-off membrane and concentrated to a volume of 250 µL. Concentrated protein was exchanged into gel filtration buffer (for His_6-CpkC-TR and His_6-CpkC-ACP-TR: 50 mM TRIS-HCl pH 8.0, 200 mM NaCl, 10% glycerol; for His_6-CpkG: 50 mM HEPES pH 7.8, 100 mM NaCl, 20 µM PLP, 10% glycerol) using the same

centrifugal filters (Amicon) and 20 µL aliquots were flash frozen in liquid nitrogen and stored at -80 ^{o}C.

5.11.3 Gel Filtration Chromatography

Gel filtration chromatography was carried out (in order to determine the native oligomeric state of His_{6}-CpkG) by FPLC using an AKTA Purifier (GE Healthcare). Purified recombinant protein that had been buffer-exchanged into gel filtration buffer and concentrated was diluted to 1 mL and injected unto a 110 mL HiLoad^{TM} 75 Superdex^{TM} 12 prep grade gel filtration column (Amersham Biosciences) at a flow rate of 1 mL/min. Eluate absorbance at 280 nm was observed and the apparent molecular weight was estimated by comparison of the elution volume with the elution volume of proteins of known molecular weights (12,000 – 20,000 Da) used for calibration. The proteins used were: blue dextran (2000 kDa), β-amylase (200 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa) and cytochrome C (12.4 kDa).

5.12 Polyacrylamide Gel Electrophoresis Analysis

All protein purifications were carried out using a 12% polyacrylamide gel. The stacking gel and running gel were prepared using similar recipes by mixing aliquots of stock solutions:1.5 M Tris-HCl pH 8.8 (2.5 mL for running gel only), 0.5 M Tris-HCl pH 6.8 (1.25 mL for stacking gel only), 30% acrylamide / bisacrylamide (2 mL for stacking and running gels); , N,N,N',N'-tetramethylenediamine (**TEMED**: 15 µL for stacking and running gels), 10% sodium dodecyl sulphate (**SDS:** 50 µL for stacking and running gel), 10% w/v ammonium persulphate (25 µL for stacking and running gel), made up to 5 mL with dH_2O. Care was taken to add TEMED last to the gel preparation mixture because it catalyzes the rapid polymerization process. The running gel mixture was added to a BIORAD casting plate and allowed to set for 30 min. The semisolid running gel mixture was overlaid with 70% ethanol (1 mL) during the process of setting. After the gel had been set, the ethanol solution

was removed prior to the addition of the stacking gel mixture (and the sample well comb was inserted) which was left for another 30 min to set.

The running buffer was 1X Tris-glycine running buffer and the sample buffer was made up of 100 mM Tris, pH 6.8, 2% SDS, 5% β-mercaptoethanol, and 15% glycerol. After solidification, the cast gel was immersed into a buffer bath containing the running buffer and each sample was mixed separately with the sample buffer and loaded onto the polyacrylamide gel. Each experiment included a standard ladder marker containing a mixture of proteins of known molecular weight for comparison. The gel in the buffer bath was connected to standard electrodes and run for 1 hour.

5.13 Determining Protein Concentration by UV-Spectrophotometry

Protein concentrations were estimated by measuring the UV absorbance at 280 nm on a NanoDrop spectrophotometer (Thermo Scientific). Extinction coefficients for CpkC-TR: 53525; CpkC-ACP-TR: 55015 and CpkG: 19495 L M^{-1} cm^{-1}.

5.14 Thioester Reductase Enzyme Activity Assays

5.14.1 Fluorimetric Analysis of NAD(P)H Consumption by Cpk-TR and Cpk-ACP-TR

Initial experiments to detect enzymatic activity of the reductase domain and didomain involved measurement of the change in intensity of NAD(P)H fluorescence using a Perkin Elmer LS 55 fluorescence spectrometer. Enzymatic reactions were initiated by the addition of either His_6-CpkC-TR or His_6-CpkC-ACP-TR (20 µM) to a mixture of octanoyl-CoA (200 µM), NADH or NADPH (200 µM) in 10 mM Tris-HCl buffer at a pH of 7.2 to a final reaction volume of 125 µL (control reactions using denatured enzyme were analysed alongside the samples for comparison). Both NADH and NADPH consumption was observed in separate assays to determine enzyme specificity. The progress of the reaction was monitored by

incubation for 30 min at 25^{o}C while measuring the change in fluorescence emission at 462 nm after excitation at 340 nm.

5.14.2 LC-MS Detection of Acyl-CoA Consumption and Coenzyme A formation

Additional evidence in support of enzymatic activity was sought by observing the consumption of the acyl-CoA thioester co-substrate (commercially available analogues of the His_{6}-TR substrate) and attendant formation of the reduction product coenzyme A thiol; this was carried out by LC-MS analysis with mass spectrum detection in negative ion mode. The reaction mixture described above was scaled up prior to LC-MS analysis as follows: His_{6}-TR (80 µM), acyl-CoA (2 mM), NADH (2 mM) in 10 mM Tris-HCl buffer pH 7.2 to a final reaction volume of 125 µL. The mixture was incubated at 25 ^{o}C overnight and protein was precipitated by centrifugation after addition of 250 µl of methanol. A 50µl aliquot of the clear supernatant was injected into an Eclipse XDB-C18 column (150 x 4.6 mm, 5 µm, column temperature 25 ^{o}C, Agilent) which was connected to an Agilent 1100 instrument equipped with a binary pump and diode array detector. The total run time was 45 minutes at a flow rate of 1 mL/min The HPLC outflow was connected via a splitter (10% flow to MS, 90% flow to waste) to a Brucker HCT mass spectrometer equipped with an electrospray source operated in negative ion mode with parameters as follows: nebulizer flow 40 psi, dry gas flow 10.0 L/min, dry temperature 300 ^{o}C, capillary - 4kV, skimmer 40V, capillary exit 106 V, ion charge control target 100,00. **Table 2** below shows the gradient elution profile used to analyse the His_{6}-TR reaction mixture.

Table 2: Gradient elution profile for LC-MS analysis of enzymatic reaction mixture.

Time (min)	% H_20 + 0.1% NH_4	% MeOH + 0.1% NH_4
0	100	0
5	100	0
25	0	100
30	0	100
38	100	0
45	100	0

5.14.3 Detection of Product (Alcohol) by GC-MS Analysis

Reduction of the cognate thioester intermediate to its corresponding aldehyde by the TR domain in the proposed coelimycin P1 biosynthetic assembly line, is one of the reactions of interest for this study. Attempts at direct determination of the aldehyde intermediate were however, irreproducible, probably due to the transient existence of the aldehyde which could potentially undergo further reduction to its corresponding alcohol. Experiments to detect the TR-mediated conversion of acyl-CoA substrates and aldehydes to alcohol by GC-MS were optimized by Drs. Joleen Masschelein and Jade L. Ronan. For conversion of octanal to octanol: the reaction mixture contained His_6-CpkC-TR (100 µM), NADH (20 mM), octanal (20 mM) in ammonium bicarbonate buffer (25 mM); while conversion of octanoyl-CoA to octanol: the reaction mixture contained His_6-CpkC-TR (125 µM), NADH (40 mM), octanoyl-CoA (20 mM) in ammonium bicarbonate buffer (25 mM). The 125 µL mixture was incubated for 24 h

at room temperature then an equal volume of $CHCl_3$ was added to extract the analyte and stop the reaction by enzyme precipitation. The organic phase was dried with $MgSO_4$ and filtered through a dropper fitted with cotton wool. Derivatization of the analyte was carried out in order to enhance its detection by GC-MS, one drop of *N,O*-Bis-(trimethylsilyl)acetamide (BSA) was added to the samples to convert the alcohol to its corresponding trimethylsilylether. A negative control sample with denatured enzyme was prepared and handled identically. Each sample (5 µL) was injected into a VF-5ms column (30 m length, 0.25 mm internal diameter) fitted to a Varian 400 GC-MS using a CP8400 sample injector in splitless mode. The temperature profile was: 50 ^{o}C for 1 min, ramped up to 300 ^{o}C, held constant at 300 ^{o}C for 9 min and then ramped back to 50 ^{o}C at a rate of 25 ^{o}C/min. Helium gas was used as the carrier gas.

5.14.4 Enzymatic synthesis of [4*R*-^{2}H] NADH

Deuterated diastereomers of NADH were synthesized using established enzymatic methods. Yeast (*Saccharomyces cerevisiae*) alcohol dehydrogenase (8 U) was added to a 2 ml reaction mixture containing 7 mM NAD^{+} and 3.2 M *d4*-methanol in 50 mM sodium phosphate buffer (pH 8.0). The mixture was incubated for 4 hr at 25 ^{o}C and the formation of NADH was monitored at intervals by diluting 10 µl aliquots of the reaction mixture to 1 ml and measuring the increase in absorbance at 340 nm. The reaction mixture was loaded onto a DEAE cellulose column (1 X 15cm) which had been pre-equilibrated with 20 ml 25 mM ammonium bicarbonate buffer (pH 8.0). The column was washed with 50 mM ammonium bicarbonate (pH 8.0) and product was eluted with 250 mM ammonium bicarbonate (pH 8.0). Eluate fractions with an absorbance greater than 0.5 at 340 nm were pooled, freeze-dried and re-dissolved in D_2O prior to NMR analysis.

5.14.5 Enzymatic synthesis of [4*S*-^{2}H]NADH

Glucose dehydrogenase from *Pseudomonas* sp was used to prepare [4*S*-^{2}H] NADH using the same method described above for the synthesis of [4*R*-^{2}H]NADH. The reaction was incubated for 1 hr and the product was purified and analysed as described previously.

5.15 Aminotransferase Enzyme Activity Assays

5.15.1 Spectrophotometric Measurement of PLP-bound CpkG

Purified N-terminal His_6-CpkG was diluted to a concentration of 10 mg/ml and incubated with 0.2 mM PLP at 4 ^oC for 1 hr then for 30 min at room temperature (a control sample was also prepared identically without the addition of PLP). The protein was exchanged into assay buffer (50 mM Hepes pH 7.8) using an Amicon Ultra centrifugal filter unit (Millipore) prior to analysis by UV-vis spectroscopy. UV spectra were recorded between 300 and 600 nm on a Cary 50 UV/ Vis spectrometer (Agilent Technologies).

5.15.2 Spectrophotometric Analysis of Amino Acid Formation

Aminotransferase activity was investigated by utilizing a spectrophotometric assay designed to detect amino acid formation. The concentration of a blue α-amino acid-copper (II) complex can be measured at 595 nm using UV-vis spectrometry. Purified protein (100 µM, His_6-CpkG) was incubated with 5 mM amino donor (either benzylamine, 1-propylamine, isobutylamine or 1,7-diaminoheptane) and 20 mM amino acceptor (α-ketoglutaric acid or pyruvate) in 100 mM potassium phosphate buffer pH 7.0 to a final reaction volume of 160 µl at room temperature overnight in separate reactions in a 96-well plate. A saturated copper sulphate / methanol staining solution was prepared by dissolving 300 mg $CuSO_4.5H_2O$ in 500 µL water and then adding 30 mL of methanol. The solution was centrifuged at 9,000 g for 1 min to precipitate undissolved salt and the supernatant was collected and used for further analysis. The copper sulphate / methanol solution (40 µl) was added to each reaction

mixture in the 96-well plate; the solution served to both precipitate out the enzyme from the reaction and to form a complex with the α-amino acids formed. The resulting suspension was then filtered by centrifugation using PVDF micro spin filters (Grace Davison Discovery Sciences) to remove the precipitated protein. Negative control samples were prepared alongside each sample (by eliminating the enzyme solution from the reaction mixture) and a buffer blank (160 µl of 100 mM potassium phosphate buffer with 40 µl staining solution) was also prepared and measured. A Microplate reader (Tecan GENios Microplate Reader) was used to measure the UV absorbance at 595 nm for each sample.

5.15.3 Detection of Octanal Formation

A 125 µL enzymatic reaction mixture was set up (90 µM His6-CpkG, 200 µM octylamine with 200 µM pyruvate or α-ketoglutarate in 25 mM phosphate buffer at pH 7.0) and incubated at 25^{o}C overnight. An excess of *D*-cysteine (400 µM) was added to the reaction mixture which was then incubated for 10 min at 50^{o}C. The *D*-cysteine was used to derivatize the aldehyde product (converting it to its corresponding thiazolidine-4-carboxylic acid derivative) in order to improve its detection by LC-MS analysis. Protein was precipitated out of the reaction mixture by the addition of 125 µL methanol followed by centrifugation and filtration using PVDF micro spin filters (Grace Davidson Discovery Sciences). Each derivatized sample was injected onto a Zorbax Eclipse Plus column (C18, 100 X 2.1 mm, 1.8 µm) which was connected to a Dionex UltiMate 3000 RS UHPLC coupled to a Bruker MaXis IMPACT mass spectrometer.

5.15.4 Detection of Pyruvate Formation

A lactate dehydrogenase (LDH)-coupled enzymatic assay was used to observe the formation of pyruvate by transamination of *L*- and *D*-alanine by His_6-CpkG. The reduction of the pyruvate formed by the transamination reaction is catalysed by LDH using NADH as a co-factor; the change in NADH flourimetric emission corresponds to the consumption of the

reducing co-factor which is directly as a result of pyruvate formation. The enzymatic reaction was made up of 90 µM His6-CpkG, 200 µM octanal or 2-octanone, 200 µM L- or D-alanine, 1 u/ml lactate dehydrogenase (from Lactobacillus leichmanii), 200 µM NADH in 25 mM phosphate buffer pH 7.0 to a final reaction volume of 125 µL. The reaction mixture was incubated for 30 mins and analysed by fluorimetry as described earlier for the TR activity assay. A control reaction with no added enzyme was analysed in an identical manner for comparison.

5.15.5 Detection of Octylamine Formation

The same reaction used for the LDH coupled assay was carried forward for further analysis by LC-MS in order to detect the formation of octylamine from the transamination of octanal. Octylamine is the co-product of pyruvate and it would be expected to be present in the reaction mixture alongside the pre-detected pyruvate. Proteins were precipitated by the addition of 250 µL methanol and the filtered through centrifugal filters (PVDF micro spin filters, Grace Davison Discovery Sciences). After filtration, the supernatants were analysed by UHPLC-ESI-Q-TOF-MS using a Dionex UltiMate 3000 RS UHPLC fitted with a Zorbax Eclipse Plus column (C18, 100 x 2.1 mm, 1.8 µm) coupled to a Bruker MaXis IMPACT mass spectrometer.

5.15.6 Mass Spectrometer Settings

The mass spectrometer was operated in positive ion mode with a scan range of 50-3000 *m/z*. Source conditions were: end plate offset at -500 V; capillary at -4500 V; nebulizer gas (N_2) at 1.6 bar; dry gas (N_2) at 8 L min^{-1}; dry temperature at 180 ^{O}C. Ion transfer conditions were: ion funnel RF at 200 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 55 *m/z*; collision energy at 5.0 eV; collision RF at 600 Vpp; ion cooler RF at 50-350 Vpp; transfer time at 121 µs; pre-pulse storage time at 1 µs. Calibration was performed with 10 mM sodium formate through a loop injection of 20 µL at the beginning of each run.

REFERENCES

1. K. J. Weissman. *Nat. Prod. Rep.*, 2015, **32**, 436.
2. J. Mann. *Chemical Aspects of Biosynthesis;* Oxford Chemistry Primers, Oxford University Press, 1999.
3. J. E. McMurry. *Organic Chemistry with Biological Applications*. Cengage Learning, Stamford, U. S. A., 2014.
4. C. J. Booker, R. Bedmutha, T. Vogel, A. Gloor, R. Xu, L. Ferrante, K. K.-C. Yeung, I. M. Scott, K. L. Conn, F. Berruti, and C. Briens. *Ind. Eng. Chem. Res.* 2010, **49**, 10074.
5. A. J. Wright. *Mayo. Clin. Proc.*, 1999, **74**, 290.
6. E. A. Crane and K. Gademann. *Angew. Chem. Int. Ed.*, 2016, **55**, 3882.
7. J. P. Gomez-Escribano, L. Song, D. J. Fox, V. Yeo, M. J. Bibb and G. L. Challis. *Chem. Sci.,* 2012, **3**, 2716.
8. U. R. Awodi, J. L. Ronan, J. Masschelein, E. L. C. de los Santos and G. L. Challis. *Chem. Sci*., 2017, 8, 411.
9. B. Shen. *Curr. Opin. Chem. Biol.* 2003, **7**, 285.
10. L. Katz and R. H. Baltz. *J. Ind. Microbiol. Biotechnol.,* 2016, **43**, 155.
11. L. Song, F. Barona-Gomez, C. Corre, L. Xiang, D. W. Udwary, M. B. Austin, J. P. Noel, B. S. Moore, and G. L. Challis, *J. Am. Chem. Soc.,* 2006, **128**, 14754.
12. S. G. Van Lanen and B. Shen. *Curr. Top. Med. Chem.*, 2008, **8**, 448.
13. J. Staunton and K. J. Weissman. *Nat. Prod. Rep.*, 2001,**18**, 380.
14. A. Das and C. Khosla. *Acc. Chem. Res.*, 2009, **42**, 631.
15. J. F. Barajas, J. M. Blake-Hedges, C. B. Bailey, S. Curran, J. D. Keasling. *Synth. Syst. Biotechnol.* 2017, **2**, 147.
16. A. T. Keatinge-Clay. *Nat. Prod. Rep.,* 2012, **29**, 1050.
17. C. Bisang, P. F. Long, J. Cortes, J. Westcott, J. Crosby, A. Matharu, R. J. Cox, T. J. Simpson, J. Staunton and P. F. Leadlay. *Nature.* 1999, **401**, 502.

18. T. A. Gulder, M. F. Freeman and J. Piel. *Top. Curr. Chem.* 2011, 1-55, Epub DOI: 10.1007/128_2010_113.

19. C. Waldron, P. Matsushima, P. R. Rosteck Jr., M. C. Broughton, J. Turner, K. Madduri, K. P. Crawford, D. J. Merlo and R. H. Baltz. *Chem. Biol.,* 2001, **8**, 487.

20. L. Laureti, L. Song, S. Huang, C. Corre, P. Leblond, G. L. Challis, and B. Aigle. *Proc. Natl. Acad. Sci. U. S. A.,* 2011, **108**, 6258.

21. O. BilykI, M. Samborskyy, P. F. Leadlay. PLoS ONE, 2019, **14**, e0215958.

22. H. Hong, Y. Sun, Y. Zhou, E. Stephens, M. Samborskyy and P. F. Leadlay. *Beilstein J. Org. Chem.* 2016, **12**, 2164.

23. T. Yao, Z. Liu, T. Li, H. Zhang, J. Liu, H. Li, Q. Che, T. Zhu, D. Li, and W. Li. *Microb. Cell. Fact.,* 2018, **17**, 98.

24. S. J. Kakavas, L. Katz, and D. Stassi. *J. Bacteriol*. 1997, **179**, 7515.

25. S. Hwang, N. Lee, S. Cho, B. Palsson and B-K. Cho. *Front. Mol. Biosci.*, 2020, **7**, 87.

26. M. A. Skiba, C. L. Tran, Q. Dan, A. P. Sikkema, Z. Klaver, W. H. Gerwick, D. H. Sherman, and J. L. Smith. *Structure*, 2019, **28**, 1.

27. J. Young, D. C. Stevens, R. Carmichael, J. Tan, S. Rachid, C. N. Boddy, R. Muller, R. E. Taylor. *J. Nat. Prod.*, 2013, **76**, 2269.

28. M. M. Alhamadsheh, N. Palaniappan, S. DasChouduri, and K. A. Reynolds. *J. Am. Chem. Soc.*, 2007, **129**, 1910.

29. B. J. Beck, Y. J. Yoon, K. A. Reynolds, and D. H. Sherman. *Chem. Biol.*, 2002, **9**, 575.

30. H. Hong, Y. Sun, Y. Zhou, E. Stephens, M. Samborskyy and P. F. Leadlay. *Beilstein J. Org. Chem*. 2016, **12**, 2164.

31. L. Du and L. Lou. *Nat. Prod. Rep.,* 2010, **27**, 255.

32. I. T. Nakou, M. Jenner, Y. Dashti, I. Romero-Canelon, J. Masschelein, E. Mahenthiralingam and G. L. Challis. *Angew. Chem. Int. Ed.,* 2020, **59**, 23145.

33. L. Vieweg, S. Reichau, R. Schobert, P. F. Leadlay, R. D. Sussmuth. *Nat. Prod. Rep.,* 2014, **31**, 1554.

34. D. X. Hu, D. M. Withal, G. L. Challis and R. J. Thomson. *Chem. Rev.,* 2016, **116**, 7818.

35. M. W. Mullowney, R. A. McClure, M. T. Robey, N. L. Kelleher and R. J. Thomson. *Nat. Prod. Rep.,* 2018, **35**, 847.

36. T. D. H. Bugg, Introduction to Enzyme and Coenzyme Chemistry, 2nd Edition, 2008, Blackwell Publishing, 211-218.

37. Y. Kalberg, U. Oppermann, H. Jorvall, B. Persson. *Euro. J. Biochem.* 2002, **269**,4409.

38. J. F. Barajas, R. M. Phelan, A. J. Schaub, J. T. Kliewer, P. J. Kelly, D. R. Jackson, R. Luo, J. D. Keasling, and S-C. Tsai. *Chem. Biol.,* 2015, **22**, 1.

39. A. Chhabra, A. S. Haque, R.K. Pal, A. Goyal, R. Rai, S . Joshi, S. Panjikar, S. Pasha, R. Sankaranarayanan, and R. S. Gokhale. *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 5681.

40. J. Masschelein, C. Clauwers, U. R. Awodi, K. Stalmans, W. Vermaelen, E. Lescrinier, A. Aertsen, C. Michiels, G. L. Challis and R. Lavigne, *Chem. Sci.,* 2015, **6**, 923.

41. A. Miyanaga, J. E. Janso, L. McDonald, M. He, H. Liu, L. Barbieri, A. S. Eustaquio, E. N. Fielding, G. T. Carter, P. R. Jensen, X. Feng, M. Leighton, F. E. Koehn and B. S. Moore. *J. Am. Chem. Soc,* 2011, **34**, 13311.

42. D. E. Ehmann, A. M. Gehring and C. T. Walsh, *Biochemistry*, 1999, **38**, 6171.

43. K-D. An, H. Nishida, Y. Miura and A. Yokota. *BMC Evol. Biol.*, 2002, **6**, 2.

44. J. A. Read and C.T. Walsh. *J. Am. Chem. Soc.*, 2007, **51**, 15762.

45. N. Schracke, Uwe Linne, Christoph Mahlert, and Mohamed A. Marahiel. *Biochemistry*, 2005, **44**, 8507.

46. F. Kopp, C. Mahlert, J. Grunewald, M. A. Mahariel. *J. Am. Chem. Soc.*, 2006, **126**, 16478.

47. B. Silakowski, G. Nordsiek, B. Kunze, H. Blocker, R, Muller. *Chem. Biol.*, 2001, **8**, 59.

48. K Gerth, R Jansen, G Reifenstahl, G Höfle, H Irschik, B Kunze, H Reichenbach, G Thierbach. *J. Antibiot.,* 1983, **36**, 1150.

49. Y. Li, K. J. Weissman and R. Muller. *J. Am. Chem. Soc*. 2008, **130**, 7554.

50. J. S. Schorey and L. Sweet. *Glycobiol.*, 2008, **11**, 832.

51. L. M. Halo, J. W. Marshall, A. A. Yakasai, Z. Son, C. P. Butts, M. P. Crump, M. Heneghan, A. M. Bailey, T. J. Simpson, C. M. Lazarus and R. J. Cox. *ChemBioChem.*, 2008, **9**, 585.

52. J. W. Sims and E. W. Schmidt. *J. Am. Chem. Soc.*, 2008, **130**, 11149.

53. C. Olano, C. Mendez and J. A. Salas. *Nat. Prod. Rep.,* 2010, **27**, 571.

54. A. C. Elliot and J.F. Kirsch, *Annu. Rev. Biochem.*, 2004, **73**, 383.

55. T. Mukherjee, J. Hanes, I. Tews, S. E. Ealick and T. P. Begley, *Biochim. Biophys. Acta*, 2011, **1814**, 1585.

56. D.E. Metzer, M. Ikawa and E.E. Snell. *Physiol. Rev.* 1953, **33**, 509.

57. A. I. Denesyuk, K. A. Denessiouk, T. Korpela and M. S. Johnson. *J. Mol. Biol.,* 2002, **316**, 155.

58. D. Schiroli and A. Peracchi, *Biochim. Biophys. Acta,* 2015, **1854**, 1200.

59. W. R. Griswold and M. D. Toney. *J. Am. Chem. Soc.*, 2011, **133**, 14823.

60. J. P. Richard, T. L. Amyes, J. Crugeiras, A. Rios. *Curr. Opin. Chem. Biol.,* 2009, **4**, 475.

61. Y. Morino and E.E. Snell. *Proc. Natl. Acad. Sci. U. S. A.,* 1967, **57**, 1692.

62. Y-L. Du and K. S Ryan. *Nat. Prod. Rep.,* 2019, **36**, 430.

63. M. Takaishi, F. Kudo, T. Eguchi. *J. Antibiot.,* 2013, **12**, 691.

64. W. Zhang, K. Watanabe, X-Cai, M. E. Jung, Y. Tang and J. Zhan. *J. Am. Chem. Soc.*, 2008, **19**, 6068.

65. C. Olano, S. J. Moses, A. F. Brana, R.M. Sheridan, V. Math, A. J. Weston, C. Mendez, P. F. Leadlay, B. Wilkinson, J. A. Salas. *Mol. Microbiol.,* 2004, **52**, 1745.

66. T. P. Tim Cushnie, B. Cushnie, A. J. Lamb. *Int. J. Antimicrob. Agents.*, 2014, **44**, 377.

67. S. D. Bentley, K. F. Chater, A.-M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth,

C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, 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.

68. G. L. Challis. *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 219.
69. F. Malpartida and D. A. Hopwood. *Nature.,* 1984, **309**, 462.
70. K. Pawlik, M. Kotowska, P. Kolesinski. *J. Mol. Microbiol. Biotechnol.*, 2010, **3**, 147.
71. J. L. Ronan, Elucidating molecular mechanisms of actinobacterial polyketide alkaloid biosynthesis, PhD thesis, University of Warwick, 2017.
72. M. Z. Ansar, G. Yadav, R. S. Gokhale and D. Mohanty. *Nucleic Acids Res.*, 2004, **32**, W405.
73. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplis, W. Li, R. Lopez, H. McWilliam, M. Remmart, J. Soding, J. D. Thompson and D. G. Higgins. *Mol. Syst. Biol.,* 2011, 539.

74. J. C. Carrington and W. G. Dougherty. *Proc. Natl. Acad. Sci.*, 1988, **85**, 3391.
75. C. Cheng and S. Shuman. *Mol. and Cell. Biol.*, 2000, **20**, 8059.
76. M. R. Wilkins, E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, D. F. Hochstrasser. Methods Mol. Biol., 1999, **112**, 531.
77. N. J. Turner and E. O'Reilly. Nat. Chem. Biol., 2013, **9**, 285.
78. T. R. Hoye, C. S. Jeffrey and F. Shao. Nat. Protoc. 2007, **10**, 2451.
79. S. B. Mostad and A. Glasfeld. *J. Chem. Educ.*, 1993, **6**, 504.
80. G. Ottolina, S. Riva, G. Carrea, B. Danieli and A. F. Buckmann. *Biochim. Biophys. Acta.*, 1989, **998**, 173.
81. S. F .Altschul, W. Gish, W. Miller, E.W. Myers and D. J. Lipman. *J. Mol. Biol.,* 1990, **215**, 403.
82. K. A. Denessiouk, A. I. Denesyuk, J.V. Lehtonen, T. Korpela and M. S. Johnson. *Proteins*, 1999, **35**, 250.

83. A. E. Beattie, D. J. Clarke, J. M. Wadsworth, J. Lowther, H. L. Sin, D. J. Campopiano. *Chem. Commun.*, 2013, **49**, 7058.

84. R. A. Vacca, S. Giannattasio, G. Capitani, E. Marra and P. Christen. *BMC Biochem.,* 2008, **9**,17.

85. H. J. Kim and H-S. Shin. *Anal. Chim. Acta.*, 2011, **702**, 225.

86. B-Y. Hwang and B-G. Kim. *Enzyme Microb. Technol.*, 2004, **34**, 429.

87. A. Stanila, A. Marcu , D. Rusu, M. Rusu, L. David. *J. Mol Struct.*, 2007, **834**, 364.

88. G. G. Wybenga, C. G. Crismaru, D. B. Janssen, and B. W. Dijkstra. *J. Biol. Chem.* 2012, **287**, 28498.

89. H. J. Cha, J-H. Jeong, C. Rojviriya, Y-G. Kim. *PLoS ONE.*, 2014, **9**, 11: e113212.

90. J. L. Galman, D. Gahloth, F. Parmeggiani, I. Slabu, D. Leys and N. J. Turner. *Front. Bioeng Biotechnol*, 2018, **6**, 205.

91. J. L. Galman, I. Slabu, F. Parmeggiani and N. J. Turner. *Chem. Commun.* 2018, **80**, 11316.

92. M. D. Truppo, J. D. Rozzel, N. J. Turner. *Org. Process Res. Dev.*, 2010, **14**, 234.

93. F. Parmeggiani, A. R. Casamajo, D. Colombo, M. C. Ghezzi, J. L. Galman, R. A. Chica, E. Brenna, N. J Turner. 2019, *Green Chem*., **21**, 4368.

94. B. S. Moore and C. Hertweck. *Nat. Prod. Reports.*, 2002, **19**, 70.

95. W. Zhang, B. Ostash and C. T. Walsh. *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 16828.

96. Z. D. Aron, P. C. Dorrestein, J. R. Blackhall, N. L. Kelleher, C. T. Walsh. *J. Am. Chem. Soc.*, 2005, **127**, 14986.

97. A. M. Cerdeno, M. J. Bibb, G. L. Challis. *Chem. Biol.*, 2001, **8**, 817.

98. R. Gerber, L. Lou, L. Du. *J. Am. Chem. Soc*. 2009, **9**, 3148.

99. M. Ma, J. R. Lohman, T. Liu, B. Shen. *Proc. Natl. Acad. Sci. U. S. A.,* 2015, **112**, 10359.

100. Y. Huang, X. Liu, Z. Cui, D. Wiegmann, G. Niro, C. Ducho, Y. Song, Z. Yang, and S. G. Van Lanen. *Proc. Natl. Acad. Sci. U. S. A.,* 2018, **115**, 974.

101. H. Seto, T. Sato and H. Yonehara. *J. Am. Chem. Soc.*, 1973, **95**, 846.

102. Y. Iwai, K. Kumano and S. Omura. *Chem. Pharm. Bull.*, 1978, **26**, 736.

103. M. Mayer and R. Thericke. *J. Org. Chem.*, 1993, **58**, 3486.

104. S. Puder, S. Loya, A. Hizi and A. Zeeck. *J. Nat. Prod.*, 2001, **64**, 42.

105. S. Ohno, Y. Katsuyama, Y. Tajima, M. Izumikawa, M. Takagi, M. Fujie, N. Satoh, K. Shin-ya, and Y. Ohnishi. *ChemBioChem.,* 2015, **16**, 2385.

106. F. J. Leeper, P. Padmanabhan, G. W. Kirby, G. N. Sheldrake. *J. Chem. Soc., Chem. Commun.*, 1987, **7**, 505.

107. W. Huang, S. J. Kim, J. Liu, W. Zhang. *Org. Lett.,* 2015, **21**, 5344.

108. H. Peng, E. Wei, J. Wang, Y. Zhang, L. Cheng, H. Ma, Z. Deng, X. Qu. *ACS Chem. Biol.*, 2016, **11**, 3278.

109. T. Terashima, E. Idaka, Y. Kishi and T. Goto. *J. Chem. Soc., Chem. Commun.*, 1973, 75.

110. S. Ye, B. Molloy, A. F. Brana, D. Zabala, C. Olano, J. Cortes, F. Moris, J. A. Salas and C. Mendez. *Front. Microbiol.,* 2017, **8**, 194.

111. Y. Huang, N. Liu, X. Wu, Y. Chen. *Curr. Org. Chem.* 2010, **14**, 1447.

112. N. Ran, L. Zhao, Z. Cheng and J. Tao. *Green Chem.* 2008, **10**, 361.

113. R. A. Sheldon and J. M. Woodley. *Chem. Rev.,* 2018, **118**, 801.