STUDIES ON THE BIOSYNTHESIS OF NON-RIBOSOMAL PEPTIDES

Andrew N. Holding
St. John's College

September 2009

A dissertation submitted to the
UNIVERSITY OF CAMBRIDGE
for the degree of
DOCTOR OF PHILOSOPHY

J. B. SPENCER GROUP
DEPARTMENT OF CHEMISTRY

UNIVERSITY OF CAMBRIDGE

For

Kim

Preface

This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated in the text.

The work was carried out under the supervision of Dr. Jonathan B. Spencer and Dr. Finian J. Leeper in the Department of Chemistry at the University of Cambridge between October 2005 and September 2009 and has not previously been submitted to any other university for any degree.

The dissertation does not exceed the word limit prescribed by the Degree Committee for Physics & Chemistry.

Andrew N. Holding
Cambridge, September 2009

Acknowledgements

Much of this work is inspired and directed by the ideas and thoughts of the late Dr. Jonathan B. Spencer, for without his guidance much of the work contained within these pages could never have been achieved. It is only tragic that such a life was cut short as both his academic and personal life was in its prime. He will be sorely missed by all of us within the group, many of whom saw him not just as a supervisor, but also a friend.

I wish to thank Dr. Finian J. Leeper and Prof. Peter F. Leadlay who did a remarkable job of standing in and supporting the group in the many months after Dr. Spencer's death. Many thanks are also due to the rest of the Spencer group for their time and effort: to Dr. Fanglu Huang for her endless energy and constant eagerness to help with molecular biology; to Dr. Manuela Tosin for her guidance in synthetic chemistry; to Dr. Helen O'Hare for the considerable amount of work she has put into the directed evolution work on HmaS and HppD before I joined the group, making the contribution of work I gave to the project run so smoothly; to Miss Helen Potter for her company during many months on end within the synthetic lab; to Dr. Andrew Truman for his endless knowledge about chemical biology, and frankness and honesty as well as his consistent dedication to fixing the LCMS after the endless abuse

it suffered. Thanks also to those within the department, without whom this project could not have been achieved: Mr. Paul Skelton and Dr. Elaine Stephens within the mass spectrometry services for their efforts; and finally to Rev. Tim Layt and Mr. Kevin Judd, our floor technicians, who have always been there to fix things when they inevitably break.

I also owe a great deal of thanks to the BBSRC for funding my time at Cambridge, without whom I would never have had the privilege to have studied here; and to St. John's College for providing my home and grants for conferences as well as some of the most splendid evenings of excess, which I doubt I will ever experience again elsewhere.

Finally, I must thank the endless support of my wife Kim, who has stood by me the whole way through my course and has been a source of endless encouragement.

Summary

Studies on the Biosynthesis of Non-Ribosomal Peptides

The problem of bacterial resistance is of growing concern within the medical community. Eventually, even with responsible use of antibiotics, new compounds will be required to bypass the resistance that bacteria have acquired. Thus the expansion of knowledge and understanding of antibiotics is key in the development of new compounds in the fight against infection. One attractive starting point for the development of new compounds are those natural products generated by non-ribosomal peptide synthetases (NRPSs), which include a range of clinically relevant glycopeptide antibiotics.

Several aspects of the biosynthesis of glycopepetide antibiotics were examined. Firstly, is the investigation to identify by the use of directed evolution if 4-hydroxymandelic acid synthase (HmaS) from the gene cluster of the antibiotic chloroeremomycin may have evolved from its homologue 4-hydroxyphenylpyruvate dioxygenase (HppD). The summation of this work is published in FEBS Lett. 2006; 580:3445. Following on from this work was an investigation into the hypothesis that HmaS catalyses the turnover of the non-natural phenylpyruvic acid to produce a

product with an inverted chiral centre compared to that of the natural substrate due to differences in substrate binding. Results showed that while not the major product, the inverted product was detected *via* chiral GCMS. Secondly, it is shown that all three cytochrome P450 enzymes (OxyA–C) that catalyse the sequential formation of three essential oxidative cross-links within the chloroeremomycin molecule do so with the retention of the oxygen atom on the peptide backbone and without the incorporation of oxygen in the air. This portion of the work is published in ChemBioChem 2008; 9:2209. The final part of the study was the development of a high-throughput screening method for NRPS A-domains, with the aim of both rapid characterisation and directed evolution of novel substrate specificity. This lead to the identification that the amino acid loaded by the first A-domain of the teicoplanin NRPS was shown to load the D-amino acid in preference to the L-amino acid. This is in contrast to the equivalent domain in the chloroeremomycin gene cluster that loads the L-amino acid.

Abbreviations

4HPP	4-hydroxyphenyl pyruvate
A	adenylation domain
ADP	adenosine 5′-diphosphate
Ahb	3-amino-5-hydroxybenzoate
α-KAO	*alpha*-keto acid dependant oxygenase
Ala	alanine
Arg	arginine
AT	acyltransferase domain
ATP	adenosine 5′-triphosphate
αKA	α-keto acid
αKAO	α-keto acid dependent oxygenase
Boc	*tert*-butyloxycarbonyl
Bz	benzyl
β-HT	β-hydroxytyrosine
C	condensation domain
Cam	chloramphenicol
CDA	calcium-dependent antibiotic
CoA	coenzyme A
CP	carrier protein
CSM	common streptomyces medium
DCC	*N,N*′-dicyclohexylcarbodiimide
DCM	dichloromethane
DH	dehydratase domain
Dhb	dihydroxybenzoic acid
DMAP	4-(dimethylamino)pyridine

DMSO	dimethylsulphoxide
DNA	2-deoxyribonucleic acid
dNTP	deoxynucleoside 5′-triphosphate
DOPA	dihydroxyphenylalanine
Dpg	3,5-dihydroxyphenylglcyine
DTNB	5,5-dithiobis(2-nitrobenzoic acid)
DTT	dithiothreitol
E	epimerase domain
EDC	*N*-ethyl-*N*′-(dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
Ent	enterobactin
Ery	erythromycin
ESI	electrospray ionization
FAD	flavin adenine dinucleotide
FAS	fatty acid synthase
FMN	flavin mononucleotide
$FMNH_2$	reduced flavin mononucleotide
FTMS	fourier transform mass spectroscopy
gDNA	genomic 2-deoxyribonucleic acid
Glu	glutamate
GC	gas chromatography
Gtf	glycosyltransferases
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulphonic acid
His	histidine
Hma	hydroxymandelic acid
HmaS	hydroxymandelic acid synthase
HOBt	1-hydroxybenzotriazole
Hpg	hydroxyphenylglycine
HPLC	high-performance liquid chromatography
HppD	hydroxyphenylpyruvate dioxygenase
HPQ	hexahydroxypterylenequinone
HRMS	high resolution mass spectroscopy
HRP	horseradish peroxidase
IPTG	isopropyl β-D-thiogalactopyranoside
ITC	isothermal titration calorimetry
Kan	kanamycin
KR	ketoreductase domain
KS	ketosynthase domain
LB	Luria-Bertani medium
LC	liquid chromatography

LDH	lactate dehydrogenase
Lys	lysine
MIC	mean inhibitory concentration
MRSA	methicillin-resistant *Staphylococcus aureus*
MS	mass spectrometry
NAD^+	oxidized nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NDP	nucleotide diphosphate
NMR	nuclear magnetic resonance
NRPS	non-ribosomal peptide synthetase
NTBC	2-[2-nitro-4-(trifluoromethyl)benzoyl]-1,3-cyclohexanedione
ORF	open reading frame
Ox	oxidised
PBP	penicillin binding protein
PCR	polymerase chain reaction
PCP	peptidyl carrier protein
PdX	putidaredoxin
PKS	polyketide synthase
PMT	photomultiplier tube
PP_i	pyrophosphate
4′PP	4′-phosphopantetheine
PPTase	4′-phosphopantetheinyl transferase
Red	reduced
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SNAC	*N*-acetyl-cysteamine
T1Y	type I tyrosinaemia
TE	thioesterase domain
TEII	type II thioesterase
TFA	trifluoroacetic acid
THN	1,3,6,8-tetrahydroxynaphthalene
TLC	thin layer chromatography
Thb	trihydroxybenzoic acid
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet
UMP	uracil monophosphate
VRE	vancomycin-resistant *enterococcus*
VISA	intermediate vancomycin-resistant *Staphylococcus aureus*
VRSA	vancomycin-resistant *Staphylococcus aureus*

Contents

List of Figures

List of Tables

1

Introduction

1.1 Antibiotic discovery

Antibiotics have been hailed as the greatest medicinal triumph of the 20th century.[1] Before their discovery, there was little available to help a patient once they had become infected. The earlier development of vaccination had introduced immunity to some diseases and sterilisation had helped to reduce the chance of infection from surgery.[2] While these were both great achievements in preventative medicine, neither was able to provide a cure. With the subsequent formulation of germ theory[3] and work identifying the role of specific bacteria in the diseases anthrax and tuberculosis,[4] the search for a cure began.

Pasteur and Joubert were the first to observe that *Anthrax bacillus* could be killed by the presence of other bacteria, suggesting that a method to destroy these pathogens was indeed possible. As time went on, many micro-organisms were shown to be capable of producing compounds that were effective against bacteria. In 1929, nearly 250 years after van Leeuwenhoek first discovered bacteria, Fleming noted that the growths of bacteria could be inhibited by the presence of a mould (*Penicillium notatum*).[5] He concluded that this effect was caused by a metabolic product from the

Table 1.1: A list of the major antibiotic classes and their modes of action.

Antibiotic	Target	Resistance Mechanism
Cell Wall		
β-Lactams	Transpeptidases / transglycosylases (PBPs)[a]	β-Lactamases, PBP mutants
Vancomycin, Teicoplanin	D-ala-D-ala termini of peptidoglycan and of lipid II	Reprogramming of D-Ala-D-Ala to D-Ala-D-Lac or D-Ala-D-Ser
Protein Synthesis		
Erythromycins	Peptidyltransferase/ribosome	rRNA methylation/efflux
Tetracyclines	Peptidyltransferase	Drug efflux
Aminoglycosides	Peptidyltransferase	Drug modification
Oxazolidinones	Peptidyltransferase	Unknown
DNA Replication / Repair		
Fluoroquinolones	DNA Gyrase	Gyrase mutations

[a]PBP, penicillin-bind protein

mould that was interacting with the staphyloccocal culture. This insight eventually led to the creation of the prototype drug penicillin G. Penicillin was the first of the family of β-lactam antibacterials that now forms the largest share of the antibacterial market.

In 1944, streptomycin was discovered to be effective in treating tuberculosis. This was the first of the aminoglycoside antibiotics and provided the first example of an antibiotic that was effective against Gram-negative bacteria (this is in comparison with penicillin, which has low activity against such bacteria).[6] At the same time, Brotzu's work introduced cephalosporins.[7] Since then, more than ten classes of antibiotics have been discovered (see table 1.1, which shows the current major antibiotic classes) and with this rapid success and the continuation of work on vaccination, it was believed that humanity might finally overcome infectious disease.[8]

1.2 Appearance of antibiotic resistance

This belief that the discovery of antibiotics made them widely available, which only accelerated the appearance of antibiotic resistance, devaluing them. This is aptly named the "antibiotic paradox".[9] The appearance of resistance should have been foreseen. The biochemical warfare between micro-organisms dates back much further than any human intervention, and these organisms had developed defences during that time against the very weapons we have tried to turn on them. The truth is shown no more clearly than in the history of penicillin. Within a decade of antibiotic use, resistance was already being observed, first in hospitals and later in the community.[10]As time has progressed, more antibiotics have succumbed to the rising levels of resistance with the effectiveness of most antibiotics dropping. By 1993, the dosages for these antibiotics had increased from 10- to 100-fold the dosage when they were first introduced.[11]

With the appearance of methicillin-resistant *Staphylococcus aureus* (MRSA), the problem has only grown further. While at present MRSA is mostly confined within hospitals and the care community, the risk of community-acquired infection is very real and something leading to much worry within both the medical profession and general public.

1.3 Origins of bacterial resistance

With the ongoing struggle for dominance between organisms, it is inevitable that over time resistance to various compounds will appear. It is hard to imagine that antibiotics could exist at all without some degree of resistance by the host organism, for without resistance the antibiotic kills the cell that produced it. There are three main reasons for antibiotic resistance: the first is that of producer organisms; the second is the resistance found in bacteria that do not produce an antibiotic but live in

environments where it is present; and the third is that the organism is naturally resistant to an antibiotic since its cell make-up differs as a result of different evolutionary factors. An example of this is the poor activity of penicillin towards Gram-negative bacteria, due to its inability to penetrate the cell wall.

In hospitals, there is constant exposure to antibiotics leading to a continual selective pressure on the bacteria in the environment. This enables bacteria that have acquired resistance to antibiotics to survive in competition with other bacteria and even dominate. This effect can be seen on a large scale by comparing the emergence of MRSA and the use of cephalosporin antibiotics. A study carried out between 1982 and 1994 showed that there is a clear link between the increasing use of cephalosporins and increasing prevalence of MRSA.[12]

An alternative method for the acquisition of antibiotic resistance is the transfer of resistance genes from one bacterium to another. This is promoted by the presence of resistance genes on extrachromosomal elements such as plasmids, that are mobile and able to transfer between cells, and on transposons, which are smaller transposable DNA elements. It is this latter process that spreads the vancomycin resistance gene (VanA) and it is found embedded in a plasmid within the vancomycin-resistant *enterococcus* (VRE) cell.[13]

The method by which the organism actually protects itself from the antibiotics, however, is much more varied than the methods by which these resistances are acquired.

In many cases of naturally antibiotic-producing bacteria, the enzymes involved in the final steps of the creation of the antibiotic also act to pump it out of the cell. One example of this is the protein complex involved in the making of lantibiotic proteins where the efflux and synthesis of the antibiotic are directly coupled.[14] These pumps keep the concentration of the antibiotic low at a safe level inside the cell. The mechanism employed by vancomycin-producing bacteria works on a different prin-

ciple, that of modifying the target site for the antibiotic. This results in the toxicity of the antibiotic being reduced so the cell can continue to function in its presence. The target of vancomycin is the D-Ala-D-Ala building block involved in the cell wall synthesis. Once the bacterium reaches stationary phase, the cell starts producing both vancomycin and the enzymes VanH, VanA, and VanX, which togther act to replace D-Ala-D-Ala in cell wall biosynthesis to D-Ala-D-lactate.[15] This modification reduces the affinity of vancomycin to the cell wall by over a 1000-fold[16] and allowing the cell to continue to function while producing the antibiotic.

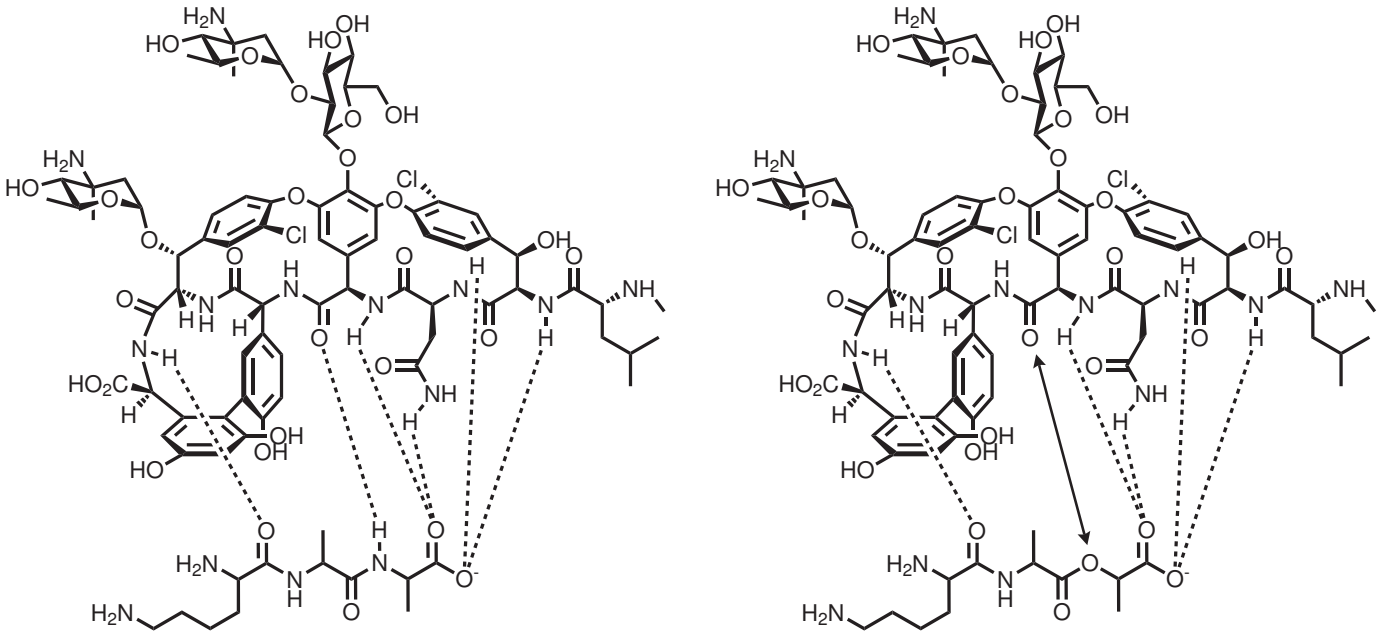


Figure 1.1: The vancomycin-like antibiotic chloroeremomcyin bound to D-Ala-D-Ala (left) and D-Ala-D-Lactate (right). The latter complex lacks a critical H-bond interaction and is significantly destabilised.

The third form of antibiotic self-resistance employed in nature is the inactivation

of antibiotics. When a bacteria requires self-protection from its own antibiotic production, the antibiotic is reversibly inactivated so that, once the antibiotic has left the cell, it reverts to the active version. In contrast, bacteria which show resistance to an exogenus antibiotic may irreversibly transform it into a form that is harmless to the cell. For example, a bacteria may generate penicillinase which inactivates penicillin by the opening of the β-lactam ring.[17]

1.4 Drugs of last resort

Due to the nature of MRSA, when a patient is unfortunate enough to acquire an infection, the resulting treatment often involves the use of vancomycin, perhaps with gentamicin or another drug. As such, vancomycin is thought of as a drug of last resort. While vancomycin is currently an effective treatment in the battle against MRSA, vancomycin resistance is already well documented in vancomycin-resistant enterococci (VRE).[18, 19] This, and the emergence of MRSA with resistance to vancomycin (VRSA) in a small number of isolated cases,[20] is definitely a cause for concern, and lends real urgency to the search for new derivatives of known antibiotics, as well as wholly novel structure class.

1.5 Relevance of non-ribosomal peptide synthetases

Non-ribosomal peptide synthetases (NRPSs) are large assembly-line enzymes that represent an alternative mechanism for the production of peptide chains. While the ribosome is limited to just 20 amino acids (not including selenocysteine and pyrrolysine), NRPSs have been shown to use hundreds.[21] NRPSs often possess catalytic domains capable of modification such as epimerisation, methylation or cyclisation. However, they appear to lack the multiple proof-reading mechanisms of the ribosome, offering the benefit of more diverse products but lacking the precision required

for primary metabolites. The differences between these processes results in a large and much more diverse pool of NRPS products compared with those generated by the ribosome, and some examples of already known and medically useful NRPS products are shown in figure 1.2. Cyclosporin is used as an immunosuppressant while teicoplanin, tyrocidine and daptomycin are clinically-used antibiotics. Other known biological uses for NRPS products are siderophores (iron chelating compounds), and cytostatics. A further feature of NRPS products are the varied modifications, such as found on antibiotics like fengicin, surfactin or teicoplanin, which all contain varying lengths of fatty acid chain.[22, 23, 24]

1.6 The NRPS machinery

The NPRS consists of multiple modules, which in themselves are made up from several enzymatic domains (figure 1.3). Each module catalyses the incorporation of a single building block (for example an amino acid) into the peptide chain. In bacterial NRPSs, the individual modules are often distributed over multiple proteins. For example, the bacterium *Actinoplanes teicomyceticus* requires seven modules distributed over four separate protein sub-units in the correct order to act like an assembly line to synthesise the peptide backbone of teicoplanin.[25] This is in contrast to fungal systems where it is far more common for them to use NRPSs consisting of a single protein leading to much larger enzymes.[26, 27]

The catalytic domains within the NRPS modules can be classified into several different types. The four main types of domains are the adenylation domain (A-domain), the peptidyl carrier protein (PCP), the condensation domain (C-domain), and the thioesterase domain (TE-domain), which together comprise the minimal set of domains required for a fully working NRPS.

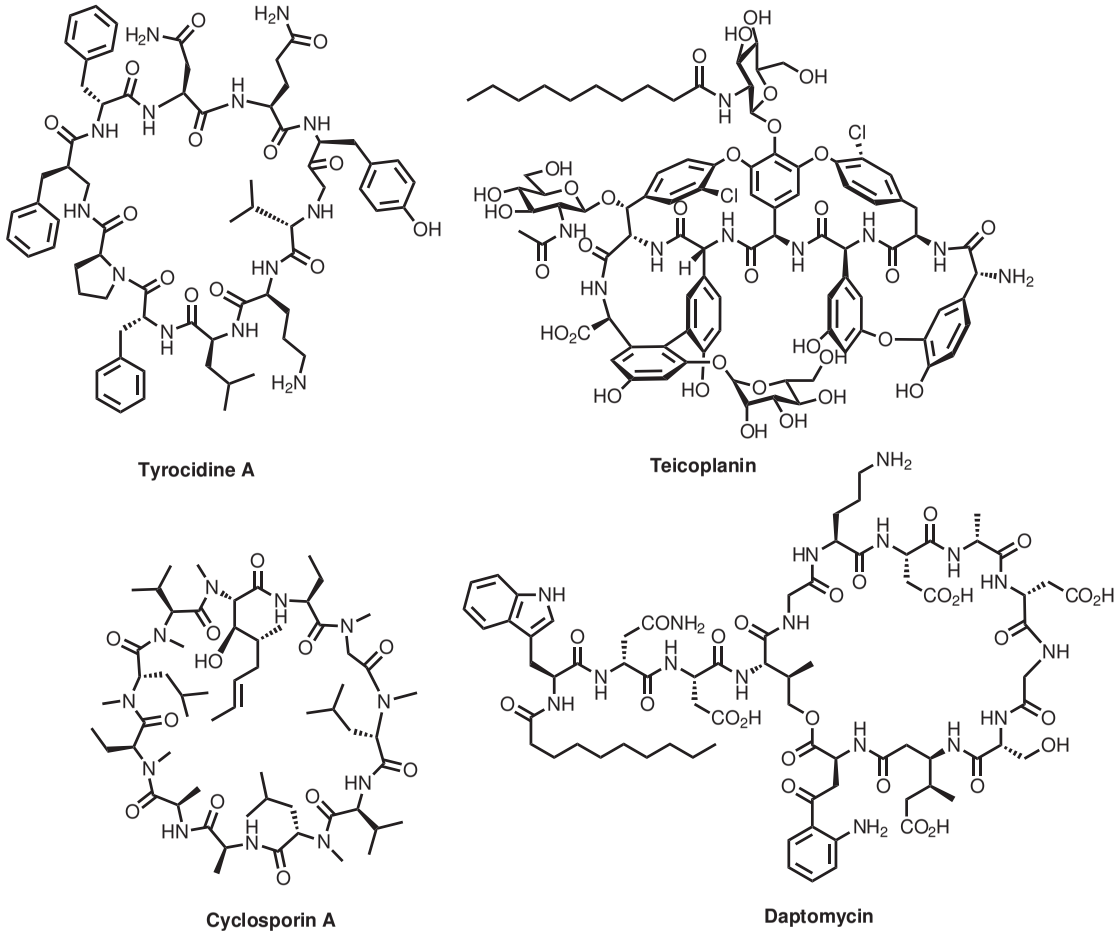


Figure 1.2: Examples of products from nonribosomal peptide synthetases, demonstrating the variation in structures and the variation in post-NRPS modification.

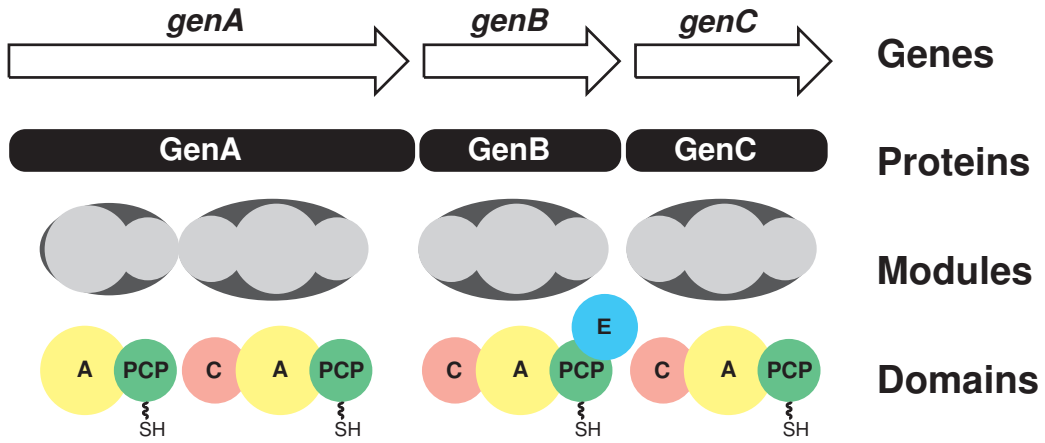


Figure 1.3: Overview of how the genes, modules and domains of a NRPS are related. A = adenylation domain, C = condensation domain, pcp = peptidyl carrier protein, E = epimerisation domain. Each gene relates to a single protein, which can then be made up of one of more modules. These modules are then generally made up of three domains, shown here in varying colours.

1.6.1 The adenylation domain (A-domain)

The role of the A-domain is to activate a particular building block – usually an amino acid – by catalysing the reaction with ATP to form the adenylate, with the release of pyrophosphate (PP_i). This adenylated amino acid is then loaded onto the PCP by the nucleophilic attack of the free thiol on the 4′-phosphopantetheine arm, a covalently attached cofactor of the PCP. The A-domain is the main site for amino acid recognition and recruitment onto the PCP as a covalently attached thioester, which in turn helps control which amino acid is incorporated into the final peptide at each position. Analysis of multiple A-domains and their respective active sites, guided by X-ray crystal structures of the A-domain of GrsA from gramicidin NRPS,[28] led to the recognition of the key role of eight amino acid side chains in the A-domain binding pocket in determining the substrate specificity of the domain.[29, 30] The deductions from this ‘non-ribosomal code’, has also led to a few examples of the altering of substrate specificity by manipulation of these specific residues. However, many other

attempts to exploit the non-ribosomal code to confer a change in substrate specificity have resulted in no change or no activity. This suggests that while the code provides a useful predictive tool in genome mining other factors are important in determining substrate specificity during peptide chain growth.

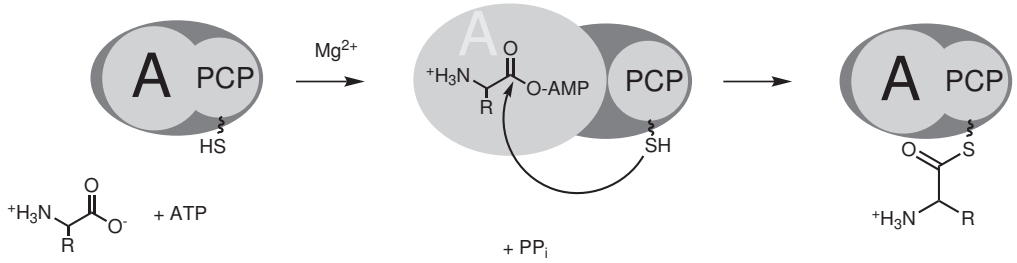


Figure 1.4: Mechanism of amino acid activation and loading by the adenylation domain. The first step shown is the docking of the amino acid into the A-domain and it's activation as an amino acid adenylate. The second step is the loading onto the neighbouring PCP by the formation a thioester and the loss of AMP.

1.6.2 The petidyl carrier protein (PCP)

The PCP is the domain that accepts the activated amino acid from the A-domain. This domain is small in size compared to the other domains within the NRPS at about 100 amino acids in length. The PCP is responsible for the movement of the amino acid or growing chain between the active sites within the NRPS. This role is aided by the amino acid or growing chain being attached to the thiol group at the end of a flexible 4′-phosphopantetheine arm (4′PP), which itself is covalently attached to a conserved serine in the PCP. The 4′PP linker is attached to the PCP by a 4′-phosphopantetheinyl transferase, which catalyses the transfer of 4′PP from coenzyme A (CoA) onto the *apo*-PCP to give the active *holo* from. This method can also be used to load acyl-CoAs onto the PCP to provide a carrier protein already loaded with a chosen acyl group.[31, 32] This is generally achieved by the use of recombinant Sfp, the 4′-phosphopantetheinyl transferase from the surfactin gene cluster in

Bacillus subtilis due to its broad substrate specificity.[33]

1.6.3 The condensation domain (C-domain)

The next phase of the process is catalysed by the C-domain. The domain simultaneously binds the acyl-PCP from both the previous module and its own module (figure 1.5) and catalyses the nucleophilic attack on the upstream aminoacyl or peptidyl-PCP by the amine of the amino acid attached to the downstream PCP, hence extending the peptide chain by one unit and transferring it onto the next PCP in the assembly line. This cycle repeats, until the peptide reaches the C-terminus of the NRPS. An important feature of C-domains is that they are reported to have some substrate selectivity for the downstream nucleophile[34, 35, 36] but little selectivity for the upstream amino acid or peptide chain beyond what is required to prevent the premature condensation. Nonetheless a C-domain that is meant to accept an upstream peptide must not accept the upstream thioester when it is of a simple amino acid. This result is important, as it strengthens the role of the A-domain as the gatekeeper of the NRPS systems.[34, 37, 38, 36]

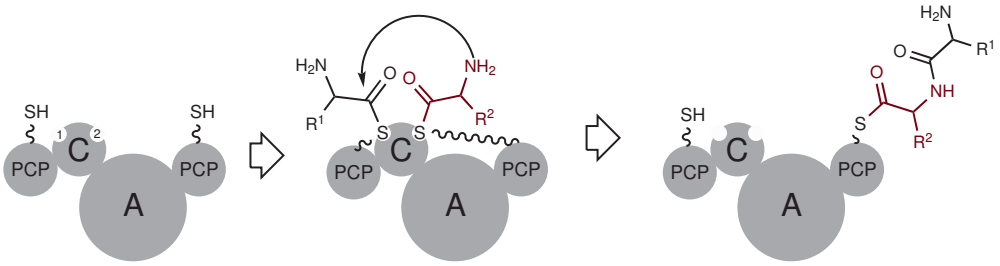


Figure 1.5: Mechanism of the condensation domain. R^1 are R^2 and the amino acid side chains. 1 and 2 are the binding sites for the acyl-PCP from the previous module and its own module, receptively.

1.6.4 The thioesterase domain (TE-domain)

The final domain in a modular NRPS is normally the thioesterase domain (TE-domain), which has the role of terminating the growing chain by unloading it from the final PCP. The TE-domain carries out nucleophilic attack on the PCP to form an acyl enzyme intermediate, which is then released either by hydrolysis or by intramolecular nucleophilic attack. This difference is reflected in the low sequence similarity between TE-domains from different chain clusters compared to the high level of similarity in the other domains.

1.6.5 Mispriming of carrier proteins

A second discrete thioesterase is often encoded within the NRPS gene cluster. These so-called type II thioesterases only show around 10% sequence identity with the type I thioesterases found integrated at the end of an NRPS. As type I thioesterase domains contain all that is needed for the termination of the growing chain chain other roles have been sought for these discrete thioesterases. Targeted deletion of the gene for typical NRPS type II thioesterases reduces, but does not abolish, peptide production. This is consistent with the current-accepted proposal that Type II thioesterases (TEII) are involved in the regeneration of carrier proteins that have been mis-acylated. While A-domains are often specific to a substrate, there is a chance, depending on the selectivity of the A-domain, that it will load the wrong amino acid onto the carrier protein. If this occurs it has been shown within the gene clusters of some NRPS such as surfactin ($TEII_{srf}$) and bacitracin ($TEII_{bac}$) that TEIIs are responsible for the hydrolysis of the misacylated thiol groups of the $4'$-phosphopantetheine arm.[39] A second role for these enzymes is when the $4'$-phosphopantetheinyl transferase utilises acyl-CoA rather than CoASH as a $4'$PP donor.[33] This would lead to a misprimed, blocked NRPS which can be reactivated by the action of the TEII.

1.6.6 The NRPS assembly line

The NRPS extension modules are combined together across several multienzymes to create a linear assembly line. Three different types of NRPS have been characterised, as shown in figure 1.6, linear (the most common), iterative (where two copies of a peptide are made and then dimerised later) and non-linear. The starter module is generally made up of simply an A-domain and a PCP, and it is this A-domain that activates the first amino acid and initiates the chain growing. This activated-amino acid is then loaded onto the PCP. The condensation domain then catalyses the nucleophilic attack from the downstream amino acid that has already been loaded on to the second PCP by the second A-domain. This forms a dipeptide on the second PCP, which is subsequently attacked by the next downstream amino acid. This cycle can then repeat till the end of the NRPS is reached, building up the peptide chain stepwise. In the case of the third module of the surfactin NRPS, there is an epimerisation domain (E-domain) present. This is an example of a tailoring domain, commonly the amino acids loaded onto the PCP by the A-domain are L-amino acids (though sometimes the D-amino acid is loaded directly)[40] this aminoacyl thioester is then racemised by the adjacent E-domain. Other common modifications include N-methylation and cyclisations. The correct interactions between successive multienzymes are promoted by so-called com domains at the N- and C- termini of appropriate sub units.[41] In some cases, the direct correspondence between the sequence of modules and the peptide product (the 'co-linearity rule') is not maintained, as for example in the biosynthesis in the hydroxamate sideophore coeliochelin from *Streptomyces coelicolor*.[42]

1.7 Introduction to glycopeptides

Glycopeptide antibiotics, such as vancomycin and teicoplanin, are biosynthesised by actinomycetes[43] and are in clinical use as drugs of last resort. Chloroeremomycin

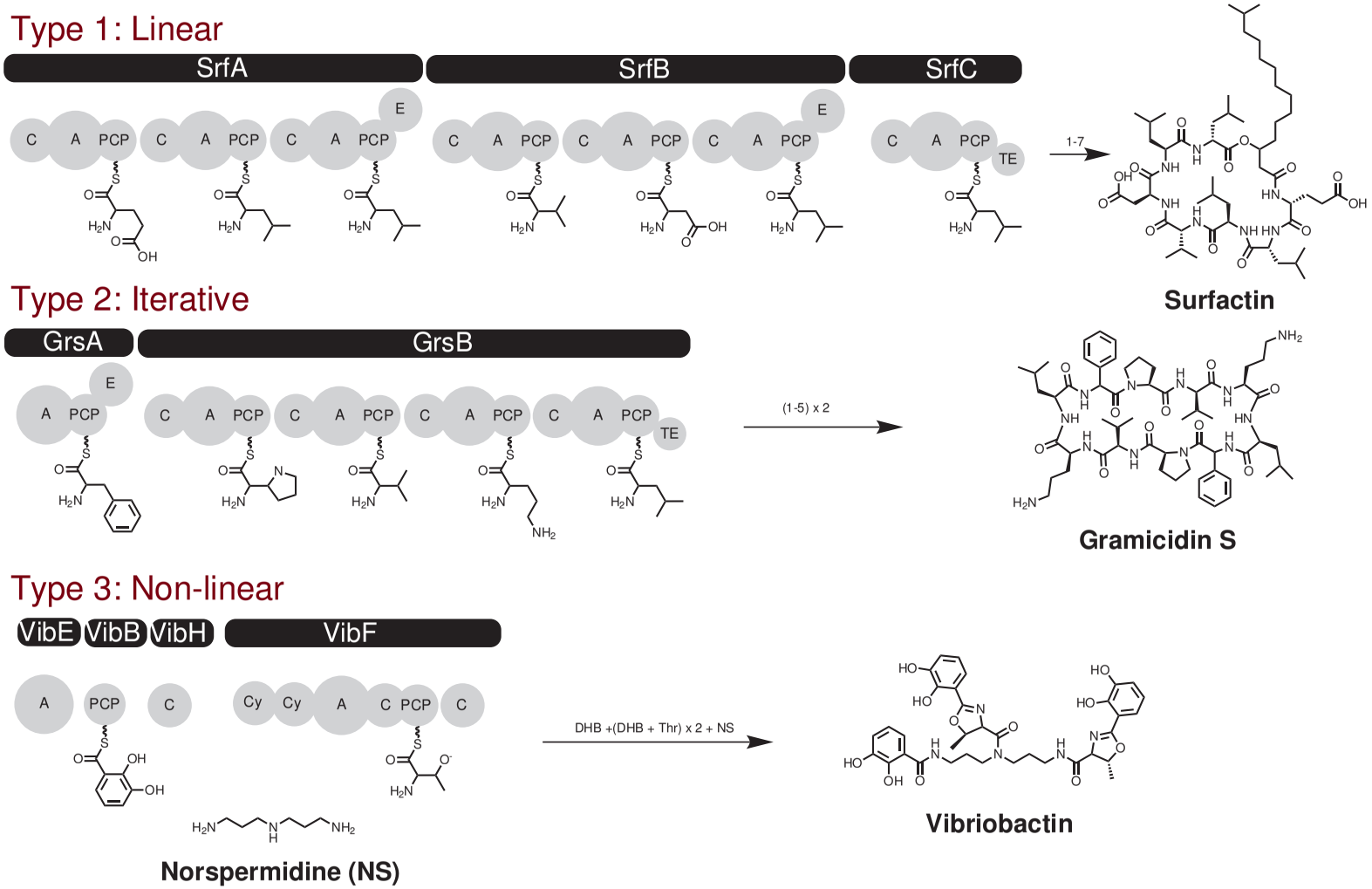


Figure 1.6: Examples of the three types of NRPS multienzyme assembly lines so far uncovered. It should be noted that while they vary in the how the genes reflect the final product, they all makes use of the same set of catalytic domains.

(figure 1.7) is a glycopeptide produced by *Amycolatopsis orientalis* which shares the same aglycone structure as vancomycin and balhimycin. While chloroeremomycin is not in clinical use, the semi-synthetic chloroeremomycin derivative oritavancin is currently approaching the end of phase III clinical trials.[44] This family of glycopeptide antibiotics shares a similar biosynthetic pathway starting with the formation of a linear heptapeptide chain, which is produced by three non-ribosomal peptide synthetase (NRPS) enzymes utilising a range of non-proteinogenic amino acids.[45, 46]

There are five classes of glycopeptides: types I-IV have antibiotic activity while type V displays anti-viral properties. Type I glycopeptides, such as balhimycin or chloroeremomycin have aliphatic side chains at amino acid positions 1 and 3. Type III and type IV contain an extra oxidative cross link. In addition type IV have a fatty acyl chain attached to a amino sugar as is found in teicoplanin, and type V have a characteristic tryptophan at residue 2.[47]

1.8 Structure of chloroeremomycin

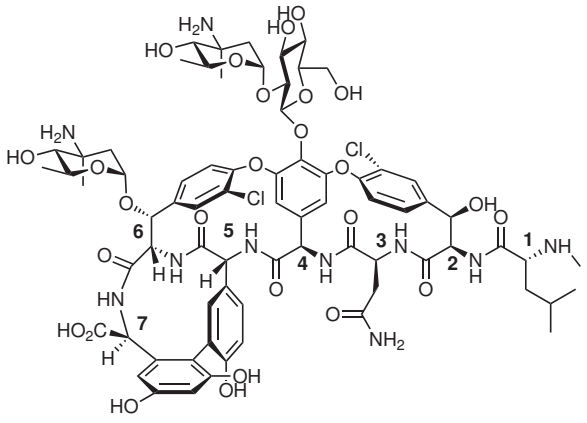


Figure 1.7: Chloroeremomycin structure (amino acid numbering shown in bold).

The heptapeptide backbone of chloroeremomycin incorporates both D- and L-amino acids as well as the non-proteinogenic amino acids 4-D-hydroxyphenylglycine

(Hpg) at positions 4 and 5; β-hydroxy-m-chlorotyrosine at positions 2 and 6 and (*S*)-3,5-dihydroxyphenylglycine (Dpg) at position 7.

The backbone is then crosslinked by a series of phenolic coupling reactions catalysed by cytochrome P450 enzymes between the aromatic residues at positions 2, 4 and 6 to form C-O-C bonds, and at positions 5 and 7 to form C-C bonds (see figure 1.7. The aglycone is further modified by the successive addition of three sugars, one glucose moiety on the *para*-hydroxyl of the Hpg residue at position 4, and two L-epivancosamine moieties, the first of which is attached to the β position of the sixth amino acid residue and the second to the glucose.

1.9 Biosynthesis of chloroeremomycin

The vancomycin gene cluster from the producer strain *A. orientalis* (figure 1.8) has been previously identified by researchers at Eli Lilly, who purified a glucosyl transferase from the cell free extract.[48] This enzyme was shown to be responsible for attaching glucose to the vancomycin aglycone. A cosmid library of the chloroeremomycin *A. orientalis* was then screened using DNA primers based on the enzyme's protein sequence. This led to the isolation and sequencing of two of the cosmids that contained the gene cluster.[45] The whole cluster is 72 kb in size and contains 39 genes. The function of the majority of the enzymes has since been deduced, either by homology searches or *via in vitro* and *in vivo* assays.

The most striking feature of the chloroeremomycin gene cluster is the large size of the NRPS genes CepA to CepC. It is these genes that are expressed to give the massive assembly-line style system that produces the peptide backbone (see figure 1.12). Also striking, and as is normal for prokaryotes, all the other genes for the biosynthetic pathway are co-located (clustered) together.

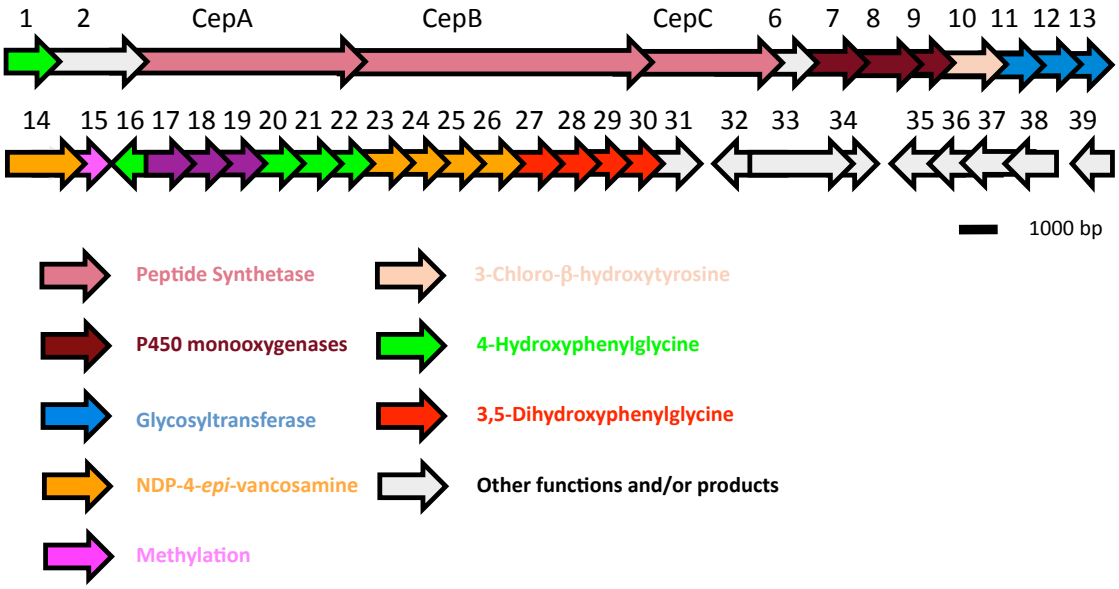


Figure 1.8: The chloroeremomycin gene cluster.

1.10 Biosynthesis of non-proteinogenic amino acids.

Chloroeremomycin, like most glycopeptides incorporates several amino acids that are non-proteinogenic. These amino acids are 3,5-dihydroxyphenylglycine (Dpg), hydroxyphenylglycine (Hpg) and β-hydroxytyrosine (β-HT) and each have their own biosynthetic pathway encoded in the gene cluster.

1.10.1 4-Hydroxyphenylglycine (Hpg)[49]

Hydroxyphenylglycine is found at two positions within the chloroeremomycin backbone and is derived from tyrosine *via* four enzyme-catalysed steps.[50] Chorismate is produced by the shikimate primary metabolite pathway and is then converted into prephenate. This then undergoes decarboxylation to form 4-hydroxyphenylpyruvate, which acts as the substrate for hydroxymandelic acid synthase (HmaS) to produce 4-hydroxymandelic acid which is then in turn oxidised enzymatically to form 4-hydroxybenzoylformate (Hbf) (see figure 1.9). Hbf then undergoes a transamination

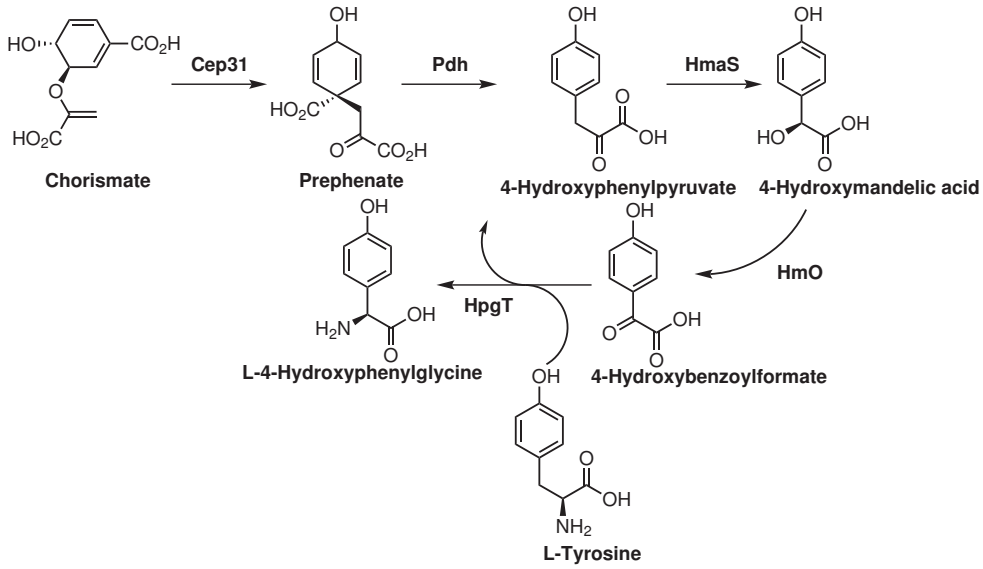


Figure 1.9: L-4-Hydroxyphenylglycine biosynthetic pathway.

reaction to form Hpg using tyrosine as the amino group donor. The deaminated product of tyrosine, 4-hydroxyphenylpyruvate, can then feed back into the pathway as a substrate for HmaS.

1.10.2 3,5-Dihydroxyphenylglycine (Dpg)

While Dpg shares a very similar carbon skeleton to Hpg, it has a radically different biosynthetic origin. Feeding experiments with isotopically labelled acetate demonstrated that the residue was made up from several acetate units, and hence it was most likely produced by a polyketide synthase (PKS).[50] This supposition was subsequently substantiated by the genetic investigations which showed five genes responsible for the biosynthesis of this secondary metabolite.[51, 52] It has been shown that the glycopeptide producers *A. orientalis*, *A.mediterranei* and *A. teichomyceticus*, which all produce Dpg, have homologues of these five enzymes, named DpgA-D and HpgT, within the gene cluster for the respective products and these genes are organised into a putative single transcription unit.[25]

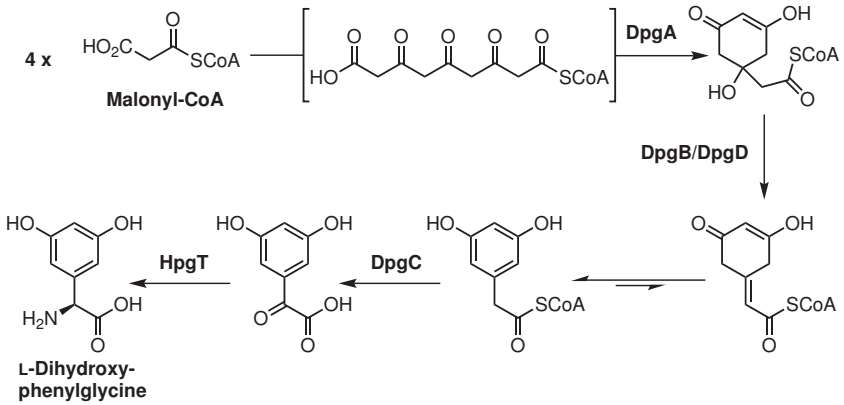


Figure 1.10: L-Dihydroxyphenylglycine biosynthetic pathway.

DpgA, which shows homology to the chalcone synthases, is a type III PKS that forms a linear tetraketide from four malonyl-CoA units. The linear product is then decarboxylated and cyclised.[53] It is proposed that the intermediate is then dehydrated

and aromatised by DpgB and DpgD. DpgC is responsible for a 4-electron oxidation to produce 3,5-dihydroxyphenylglyoxalate, and is particularly unusual because it catalyses a spin-forbidden oxidation utilising molecular oxygen without the use of a co-factor. Very few such enzymes have been characterised.[54] The final transamination is catalysed by HpgT, the same enzyme as found in the pathway to Hpg, with L-tyrosine as the amino donor (see figure 1.10).

1.10.3 β-Hydroxytyrosine (β-HT)

In studies of the vancomycin-type glycopeptides, specifically balhimycin, feeding studies have shown that the β-hydroxylation occurs before the assembly of the heptapeptide. Genetic analysis has identified a three-gene operon in both the balhimycin and chloroeremomcyin gene clusters which governs the β-hydroxylation of tyrosine.[55] There are no homologues of these genes within the teicoplanin cluster, one of the few differences between the vancomycin and teicoplanin-like biosyntheses. The β-HT operon in the balhimycin cluster encodes the loading module of a NPRS, a thioesterase and a cyctochrome P450 monooxygenase (figure 1.11).

In teicoplanin a different system appears to operate. In the organism that produces A40926, a close analogue of teicoplanin, inactivation studies of Orf28 (a homologue of Orf12* in the teicoplanin cluster) suggested that a single enzyme is responsible for this process. A homologue of this enzyme is found within the chloramphenicol cluster, an antibiotic that contains a β-hydroxylated phenylalanine residue in the peptide backbone, which reinforces this hypothesis. [56]

1.11 Growth of the peptide chain

The peptide backbone of chloroeremomcyin is synthesised by a NRPS made up of three proteins designated CepA-C. As described in section 1.6 these are modular proteins where each module contains the necessary domains to carry out one round

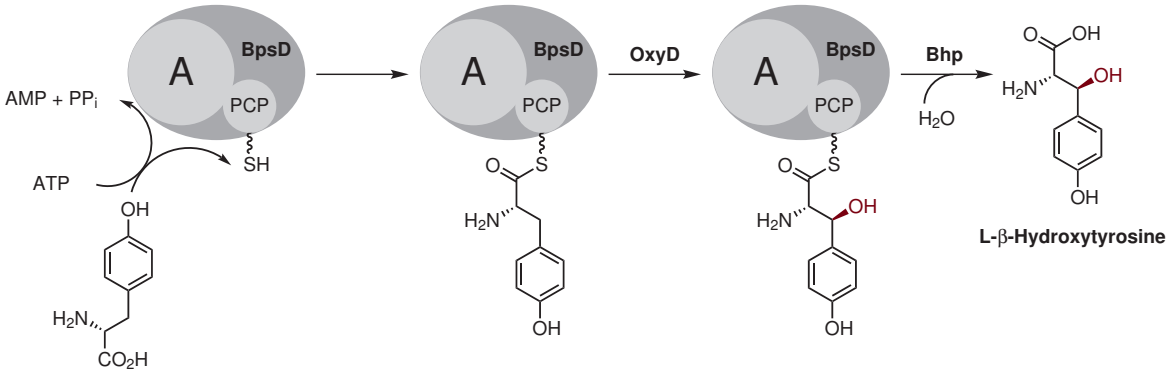


Figure 1.11: Biosynthesis of β-hydroxytyrosine in balhimycin – a vancomycin-like glycopeptide

of peptide growth and, as is normal for NRPS, the ordering of domains is consistent with the enzymatic product.[57, 58, 59] The amino acid extension modules are distributed 3-3-1 across the three NRPS proteins in contrast to teicoplanin and complestatin-like glycopeptides, which are distributed 2-1-3-1 across four NRPS proteins[25] (see figure 1.12).

The assignment of the NRPS module functions and of the amino acid that is incorporated by each A-domain are mainly in agreement with those predicted.[60, 30] As with all glycopeptide NRPSs so far characterised the first module incorporates a D- rather than an L-amino acid even though no E-domain is present. Even more curious is that in chloroeremomycin the module shows a 6:1 preference for the L-amino acid.[61] The C-domain might have dual activity and carry out the epimerisation on the amino acid as shown for the recently discovered C/E hybrid domain found in the arthrofactin pathway.[62] All the glycopeptide C-domains have been compared in a phylogenetic study, which showed them to be members of the ${}^{D}C_{L}$ family.[63] This is a family of C-domains that link an upstream D-amino acid to a downstream L-amino acid and they are usually found next to E-domains. The epimerisation activity may

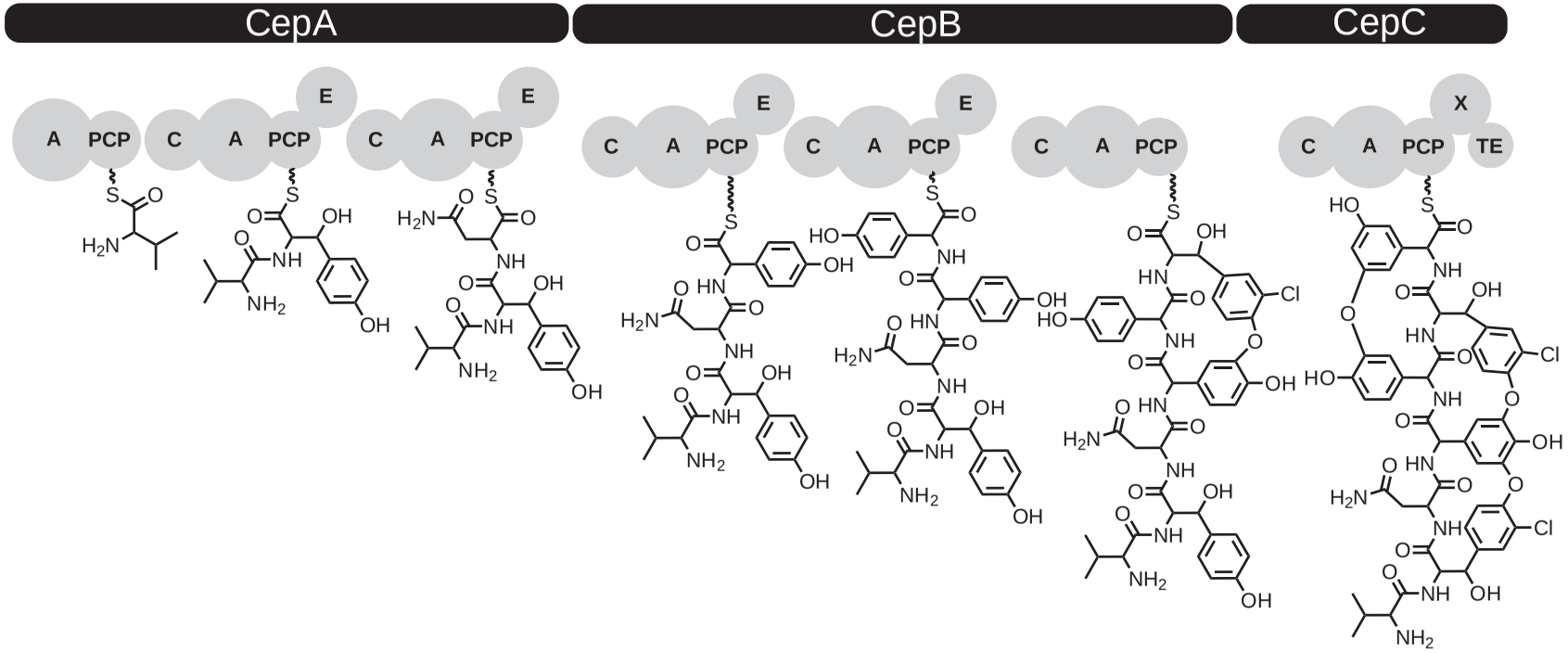


Figure 1.12: Overview of the chloreremomycin NRPS. The chloroeremomycin NRPS is spread across three proteins and seven modules reflecting the seven amino acid residues in the final structure. The module X is of unknown function, though shows homology to a C-domain. Actual timing of the cross linking steps is unknown with the exception of the first cross-linking being shown to occur at hexapeptide stage.

also be due to a gene from outside the cluser.

Within all glycopeptides, there is a degenerate module (shown as X in 1.12) adjacent to the final TE-domain. While sequence analysis would suggest it was once a C-domain,[63] it would no longer appear to serve any direct catalytic function. However, many polyketide synthases (PKSs) are known to contain inactive domains to aid stability of the synthases, while within the vibriobactin pathway a catalytically inactive domain is responsible for VibF dimerisation.[64]

1.11.1 Halogenation

Halogen-carbon bonds are abundant in nature with over 3000 different compounds isolated thus far.[65] In glycopeptides, it has been shown that the presence of chlorine atoms has a striking effect on the antibiotic activity of the molecule.[66] The mono- or non-chlorinated form of the antibiotic has been shown to have a much lower bioactivity because of a reduced binding to D-Ala-D-Ala.[67] This may also be caused by an alteration in the dimerisation of the antibiotic.[68] In balhimycin biosynthesis, it was demonstrated by the use of gene deactivation that BhaA (Orf10 in chloroeremomycin), a $FADH_2$-dependent halogenase, was responsible for the chlorination of the β-HT and that the chlorination process takes place before the oxidative cross-linking reactions occur. Thus both chlorination reactions occur before the peptide is released from the NRPS (see next section).[55, 69]

1.11.2 Cyclisation

The rigid concave shape of glycopeptides (figure 1.13) is formed by multiple oxidative cross-links between aromatic residues on the peptide backbone.[70] It has been shown in several of the glycopeptides that each cross-link is catalysed by a separate cytochrome P450 oxygenase. For the vancomycin-like antibiotics which have three cross-links, there are three enzymes (OxyA, OxyB and OxyC) that catalyse these

reactions, while the teicoplanin family of antibiotics have four cross-links and four discrete enzymes, with Orfs 5*, 7* and 9* replicating the function of OxyA-C and Orf6* providing the coupling between residues 1 and 3. By the use of gene inactivation it has been shown that in balhimycin biosynthesis the first cross-link is carried out by OxyB, coupling residues 4 and 6. OxyA then couples between residues 2 and 4, and finally OxyC couples between residues 5 and 7.[69, 71, 72]

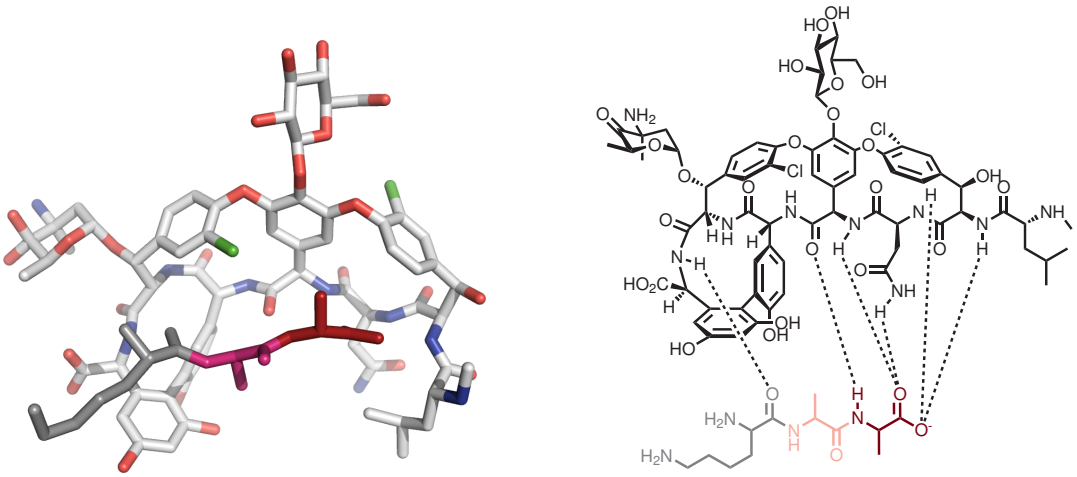


Figure 1.13: Left – Crystal structure[73] showing balhimycin bound to D-Ala-D-Ala with a resolution of 0.98 Å PDB Code:1GO6; Right – Schematic diagram of Balhimycin bound to D-Ala-D-Ala

In vivo and *in vitro* studies have been used to investigate the timing of the cross-links with respect to the NRPS. Analysis of the products from the gene inactivation studies suggested that all the cross-linking occurred before the release from the NRPS. The results of the *in vitro* work also showed that OxyB only catalyses the cross-linking reaction of the PCP-bound hexapeptide and not the full length heptapeptide or the equivalent unbound chains.[74] As OxyB is known to catalyse the first of the cross-linking reactions, it can therefore be deduced that no cross-linking can happen before the hexapeptide is formed. The cross-linking of glycopeptides is essential for antibiotic activity with the possible exception of the fourth coupling

reaction only found in the teicoplanin-like antibiotics.

1.11.3 Glycosylation

The modification of the aglycone core by the addition of sugars and for teicoplanin-like antibiotics lipids, is a feature that increases chemical diversity and is essential for optimal activity. The sugars found on glycopeptides have been shown to be responsible for aiding the dimerisation and hence binding of the antibiotic to the target sites.[68] In teicoplanin it has been shown that without the fatty acid chain, the resultant product shows a loss of activity against several *Enterococcal* and *Staphylococcal* strains.[75] This type of diversity is not just limited to glycopeptides and in some cases it is essential. Erythromycin A, for example, is inactive before the addition of two deoxysugars to the polyketide backbone as without these residues it cannot bind to its target: the 23S rRNA component of the bacterial ribosome.[76] Another example is the (*S*)-4-amino-2-hydroxybutyrate side chain in the amino glycoside butirosin.[77] The side chain, unlike the erythromycin A sugar is not essential for anti-microbial activity, rather it has been shown that it protects the antibiotic from several resistance mechanisms.

In the glycosylation of chloroeremomycin three glycosyltransferases (Gtfs) are used, one for each position of which a sugar is attached (figure 1.14). The Gtfs are all members of the superfamily GT-B, as shown by the comparison of the crystal structure of GtfA,[78] with those of MurG, a UDP-*N*-acetylglucosaminyltransferase from *E. coli*, which is responsible for the transfer of *N*-acetyl glucosamine onto the peptidoglycan precusor involved in cell wall biosynthesis;[79, 80] and of β-glucosyltransferase from T4 phage, which glucosylates hydroxymethyl cystosine of DNA.[81, 82] Many more Gtfs have since been characterised, and thousands of putative enzymes from within the GT-B superfamily have been identified.

The structure of these Gtfs in the GT-B superfamilies is highly conserved, despite

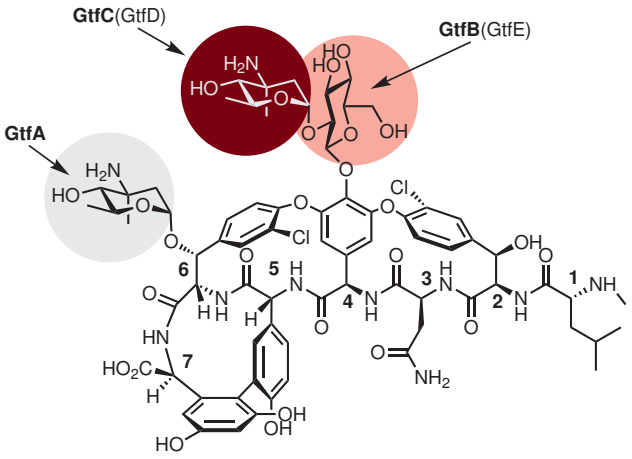


Figure 1.14: Glycosylation pattern of chloroeremomycin. The glycosytransferase catalysing the sugar attachment is shown in bold, with the vancomycin equivalent shown in parentheses.

there being a very low similarity in amino acid sequence, especially the C-terminal domain, which is responsible for the recognition of the sugar donor.[78] Structurally, this class of enzyme is made up of two separate Rossmann domains connected by a flexible linker region, with the catalytic site between the two domains. After substrate binding, a conformation change brings the two domains into closer proximity to catalyse the loading of the sugar onto the substrate and the release of the nucleotide diphosphate (NDP).

1.11.4 Biosynthesis of *epi*-vancosamine

The main structural differences between vancomycin-type glycopeptides are in the incorporation of characteristic sugars. The five enzymes for the formation of the activated deoxyamino sugar *epi*-vancosamine have been over-expressed and characterised.[83] These enzymes convert TDP-4-keto-6-deoxy-D-glucose into TDP-L-*epi*-vancosamine, which is used as a donor molecule by the Gtf. The first step is C-2 deoxygenation carried out *via* EvaA (Orf23) which catalyses the 2,3-elimination of water from

TDP-4-keto-6-deoxy-D-glucose to form the enol, which tautomerises to the diketo substrate for EvaB (Orf25). EvaB then catalyses transamination at the C-3 position to form TDP-3-amino-2,3,6-trideoxy-D-*threo*-hexopyranos-4-ulose). This compound is then methylated at C-3 by EvaC (Orf14) before being epimerised by EvaD (Orf26), switching C-5-methyl substituent from equatorial to axial, an energetically disfavoured transformation, followed by the flipping of the pyranose ring to release strain. The final step is C-4 ketoreduction by EvaE (Orf24) to yield TDP-*epi*-vancosamine (figure 1.15).

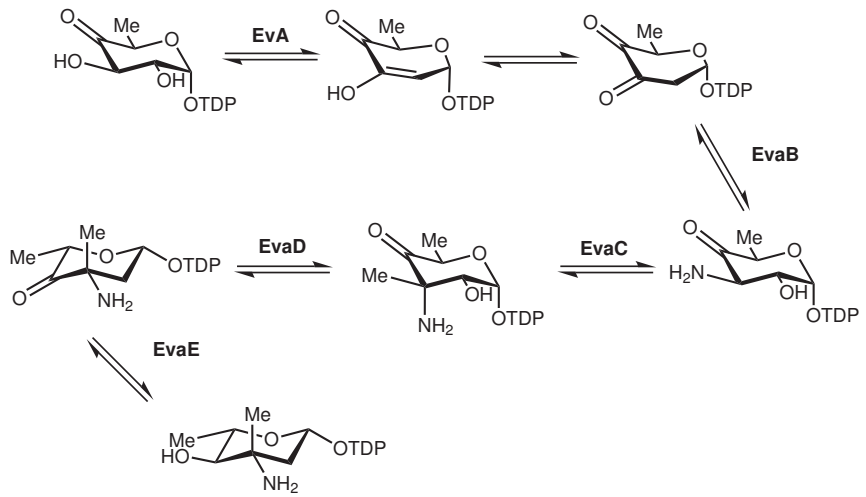


Figure 1.15: Biosynthetic pathway of TDP-L-*epi*-vancosamine

1.12 Bioactivity of glycopeptides

1.12.1 Cell wall biosynthesis

Currently, there are just four main established target sites within bacterial cells that are used to selectively kill bacteria without harming humans.[13] Folic acid and cell wall biosyntheses provide the two most versatile targets sites as there are no equiva-

lent pathways within mammals. The third and fourth target sites are found in mammals as they are the DNA and protein pathways that are crucial for all life, although due to differences between the eukaryote cells and the prokaryote bacterial systems there are differences that can be exploited without affecting the host. The target of penicillin – the first modern antibiotic and a member of the β-lactam antibiotics – is cell wall biosynthesis. The β-lactams inhibit transpeptidase enzymes by irreversibly binding to the active site and thereby stopping the catalysis of the cross-linking reaction between the peptidoglycan chains that make up the rigid cell walls. Glycopeptides also target the cross-linking reaction in the cell wall, but do so by binding to the lipid II mucopeptide precursor, formed as explained below.[84, 85]

Inhibition of cell wall biosynthesis results in cell death, because of internal osmotic pressures of up to 15 atm that are normally contained by the peptidoglycan scaffold around the cell. The biosynthesis of the cell wall of *E. coli* is initiated with the conversion of UDP-*N*-acetyl-glucosamine to UDP-*N*-acetyl-muramyl-L-Ala-γ-D-Glu-L-Lys by the enzymes MurC, D and E. D-Ala is then converted by DdlA/B to the D-Ala-D-Ala dipeptide and attached to the growing chain to form UDP-*N*-acetyl-muramyl-L-Ala-γ-D-Glu-L-Lys-D-Ala-D-Ala. The next step catalysed by MraY, is the transfer of the *N*-acetyl muramyl-pentapeptide to a C_{55} isoprenol-P-carrier to form lipid I. MurG then catalyses the β-1,4-addition of *N*-acetylglucosamine onto lipid I to form lipid II. The lipid II intermediate is then translocated to outside the cell before the trans-glycosylase forms links between adjacent peptidoglycan units. Transpeptidases then catalyse cross-linking between the peptide chains[86] (see figure 1.16).

1.12.2 Mode of action of glycopeptides

The interaction of vancomycin[84] and teicoplanin[85] with the bacteria cell wall lipid II intermediate is achieved by a tight non-covalent binding to the D-Ala-D-Ala terminal

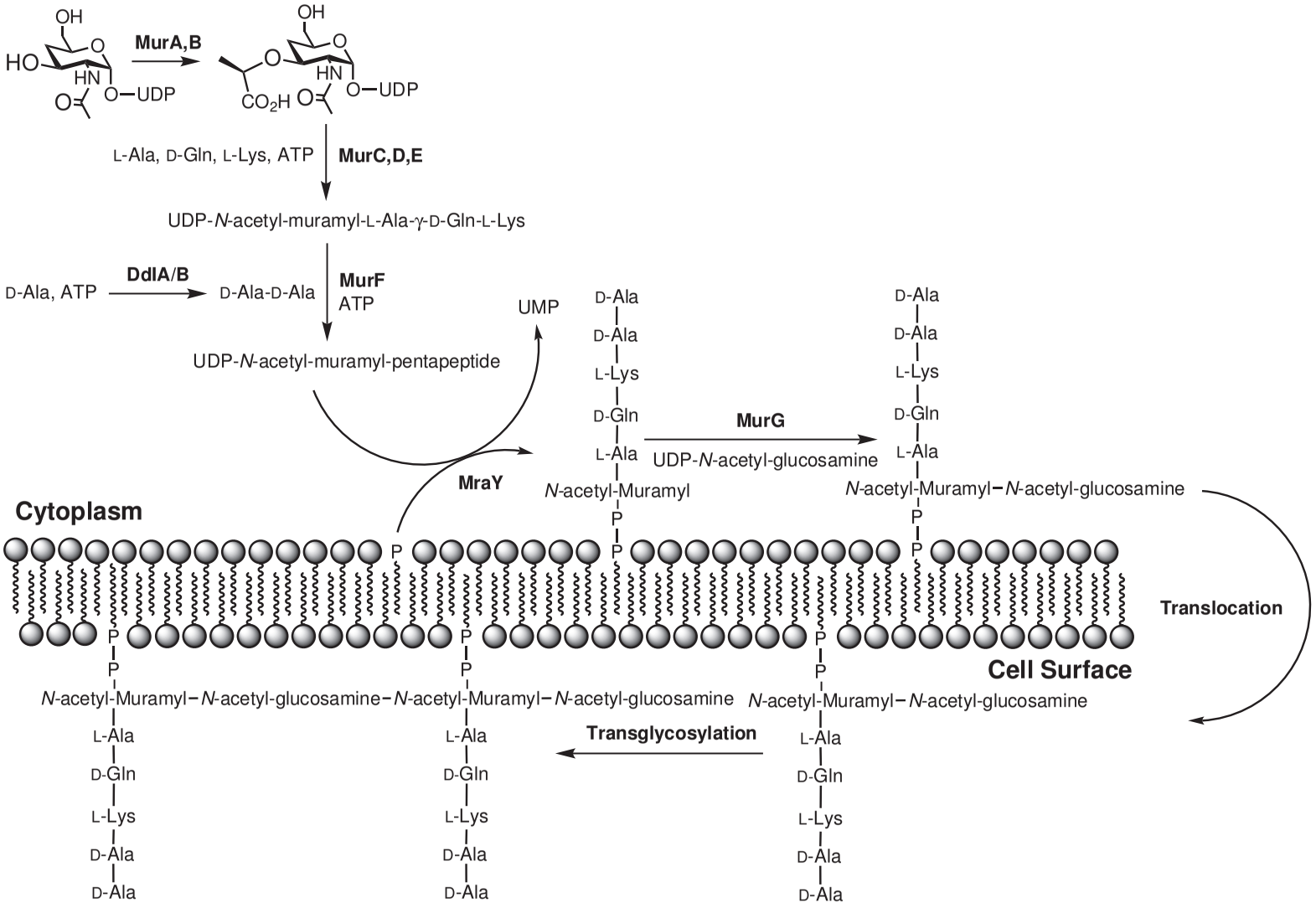


Figure 1.16: Overview of cell wall biosynthesis up to the point of transpeptidation

amino acids of the pentapeptide chain attached to the *N*-acetyl muramic acid unit. The rigid three-dimensonal cup shape of the glycopeptide allows for five hydrogen-bonds between the cell wall precursor and the glycopeptide backbone. This complexation disrupts the formation of the network that makes up the cell wall, leading to cell rupture.

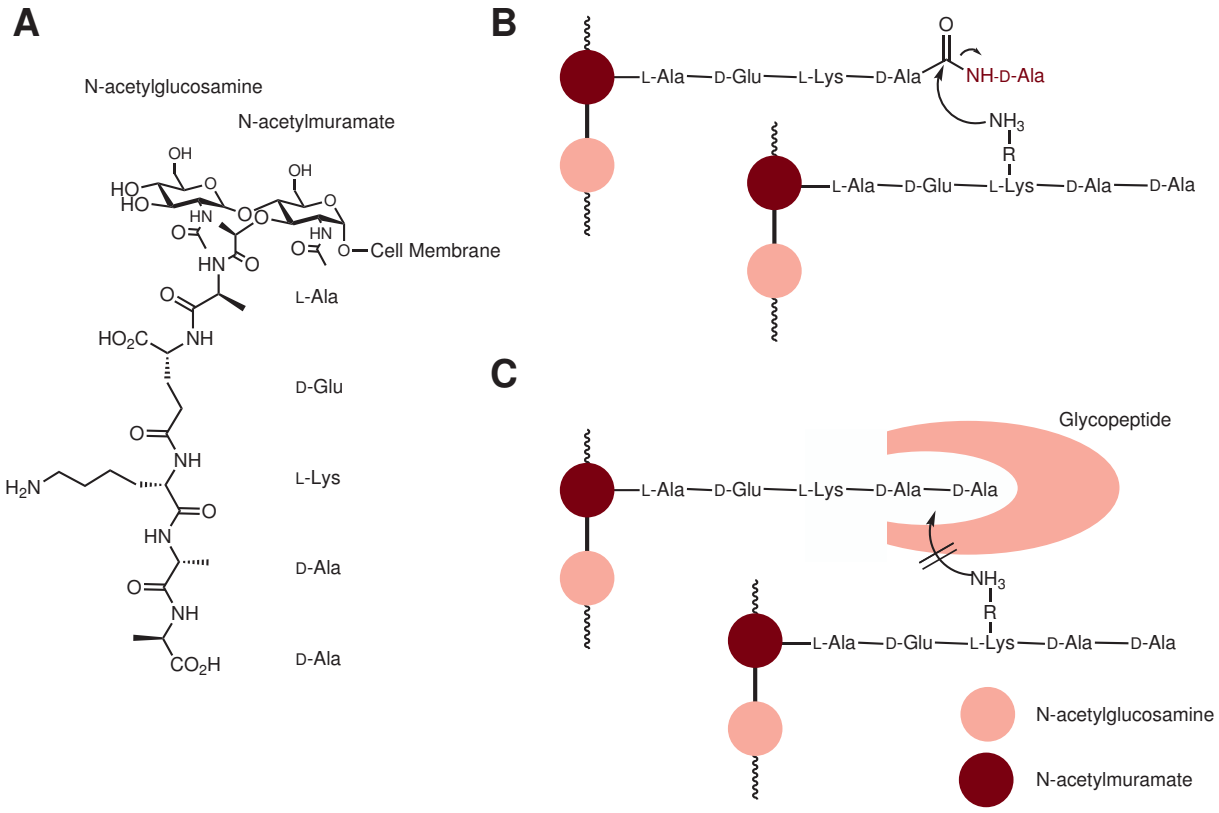


Figure 1.17: Transpeptidation step in cell wall biosynthesis: A) Cell wall mucopeptide precursor; B) Schematic outline of the transpeptidation step in cell wall biosynthesis; C) Action of the glycopeptide.

Dimerisation of glycopeptides has been observed for all the classes of glycopeptides with the exception of teicoplanin and its derivatives. Chloroeremomcyin dimerisation is facilitated by the formation of six hydrogen bonds while bound to D-Ala-

D-Ala in a back-to-back arrangement: four amide-amide hydrogen bonds from the peptide backbone and two ammonium ion-amide carbonyl interactions between the amino-sugar, 4-*epi*-vancosamine and the backbone.[87, 88, 89] All the glycopeptides that dimerise, with the exception of ristocetin A, show a higher dimerisation constant when bound to cell wall precursor analogues than when free, indicating a co-operative effect between dimerisation and the binding of cell wall precursors. [90]

In general, the sugar substituents found on amino acid residues 4 and 6, along with the chlorine atoms on the tyrosine residues at 2 and 6 act to stabilise the dimer.[68] Although eremomycin binds cell wall precursors very weakly, when compared to vancomycin it is more active as an antibiotic because its dimer is more stable when interacting with the cell membrane.[91, 90, 92]

The complexes between lipidated glycopeptides, such as teicoplanin, and the peptidoglycan precursor are not stabilised by dimerisation. Instead the lipid chain is able to enhance glycopeptide binding to the peptidoglycan by forming a second anchor point into the cell membrane.[93]

1.12.3 The vancomycin resistance-cassette

Resistance to glycopeptides was first observed as vancomycin resistance in enterococci in 1986, and it was demonstrated that it could be transferred from the enterococci to *S. aureus* within a laboratory environment.[94] In Japan in 1996, the first naturally occurring strains of *S. aureus* were isolated that showed an intermediate resistance to vancomycin (VISA).[95] This resistance was conferred by a different mechanism to that of enterococci: instead the strains possessed a thicker cell wall, providing resistance up to a minium inhibitory concentration (MIC) of 8 mg/L, much less than that in vancomycin-resistant enterococci, which has a MIC of $>$ 256 mg/L for vancomycin.[96] A whole-genome sequencing study showed that this resistance had grown over time due to single point mutations (35 mutations in 31 loci) rather

than *via* an acquisition of resistance *via* DNA transfer of a gene or genes from another organism. Full vancomycin resistance in *S. aureus* due to the presence of the enterococcal transposon was not seen until 2002.[97]

Vancomycin resistance within enterococci is categorised in seven known phenotypes (VanA–VanG) (table 1.2). VanA and VanB are the most prevalent in clinical environments and confer the greatest protection to the host cell. VanA was the first resistance mechanism characterised and is provided by a transposon. It is this positioning on a mobile DNA element that allowed for the demonstration of transfer of the resistance mechanism from enterococci to *S. aureus*. The phenotype VanB is more commonly located within the bacterial chromosome rather then being contained within a transposon.

Table 1.2: Phenotypes conferring vancomycin resistance on Enterococci[98]

Phenotype	Peptidoglycan Terminus	Resistance (MIC in mg/L)	Source	Induction	Organisms
VanA	D-Ala-D-Lact	Vanc ($\geq$ 64)	Acquired	Inducible	*E. faecium*
		Teic ($\geq$ 16)			*E. faecalis*
VanB	D-Ala-D-Lact	Vanc ($\geq$ 4)	Acquired	Inducible	*E. faecium*
					E. faecalis
VanC	D-Ala-D-Ser	Vanc ($\geq$ 2)	Intrinsic	Constitutive	*E. gallinarum*
				and inducible	*E. caseliflavus*
VanD	D-Ala-D-Lact	Vanc ($\geq$ 16)	Intrinsic	Constitutive	*E. faecium*
		Teic ($\geq$ 2)			
VanE	D-Ala-D-Ser	Vanc (16)	Acquired	Inducible	*E. faecalis*
VanG	D-Ala-D-Ser?	Vanc (16)	?	?	*E. faecalis*

The VanA gene-cassette contains five genes required for resistance: *vanR, vanS, vanH*, and *vanX*. The genes *vanR* and *vanS* encode the regulatory system that induces the expression of *vanH, vanA* and *vanX* genes. The difference between the VanA and the VanB phenotype is found in the *vanS* sensory kinase, as while activation occurs in the presence of both vancomycin and teicoplanin for the VanA phenotype, in the case of VanB it only causes activation in the presence of vancomycin. The remaining genes are responsible for the synthesis of the altered peptidoglycan cell wall. *vanA*

encodes a gene that is homologous to a D-Ala-D-Ala synthase; however its substrate specificity is altered and it synthesises D-Ala-D-Lac preferentially.[99] The gene *vanH* encodes a α-ketoacid reductase for the conversion of pyruvate into D-lactate to provide a substrate for VanA. The third gene is *vanX*, which encodes Zn^{2+}-dependent D-D-dipeptidase responsible for the cleavage of the D-Ala-D-Ala within the cell, reducing its concentration. This, combined with the presence of D-Ala-D-Lac at higher amounts, leads to the preferential incorporation of D-Ala-D-Lac into the cell wall precursor by MurF (figure 1.16). The result is a lipid II precursor containing a modified pentapeptide which reduces the glycopeptide binding energy by 4 kcal/mol.[100] The cause is the loss of a hydrogen-bond between the glycopeptide and the lipid-II precursor as the NH of the bond has been replaced with an oxygen lone pair that produces a repulsive effect to the glycopeptide.

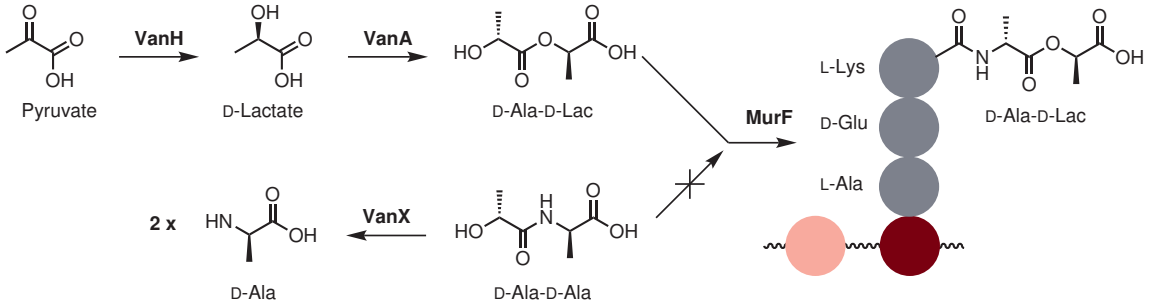


Figure 1.18: VanA and VanH are responsible for the formation of D-Ala-D-Lac from pyruvate, while VanX is responsible for breaking down D-Ala-D-Ala. This results in reducing the overall cell concentration of D-Ala-D-Ala and therefore causes MurF to incorporate D-Ala-D-Lac into the cell wall.

1.13 Directing biosynthesis

1.13.1 Introduction

While cases of vancomycin-resistant infections remain rare, they will inevitably increase with the increased use of the antibiotic. New compounds will be required to bypass the resistance the bacteria have acquired. Synthetic approaches to the problem have provided a poor source of new drugs in comparison to those developed from natural products. Combinatorial chemistry has produced a huge number of compounds after immense investment by the pharmaceutical industry, but has produced disappointing results.[101] This can be explained by the comparatively small chemical space represented by these compounds when compared to that of natural products. Diversity-orientated synthesis has been developed to increase the chemical space captured by synthetic compounds, but this work is currently in its infancy.

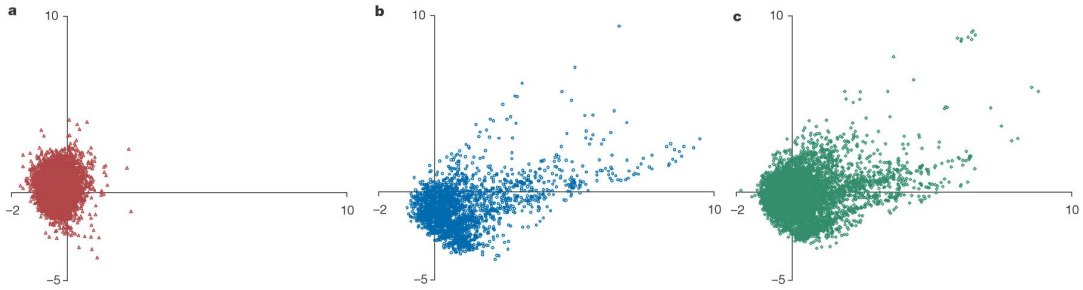


Figure 1.19: Chemical space – (a) Combinatoral chemistry, (b) Successful drugs, (c) Natural products - The diagram is the result of analysis of a large database of compounds. To visualise the diversity of the compounds studied, a statistical approach was used called 'principal component analysis'. Plots show the first two principal components. Note the similarity between the drug library and natural products when compared to the much smaller, more well defined region of chemical space covered by combinatorial chemistry.[102, 103] Figure taken from reference 102.

It is likely therefore that natural products will continue to be a crucial source of

leads in anti-infective drugs. These natural products have traditionally been found by the screening of producing organisms. However it is now possible to sequence the entire genome of a new organism and search for 'cryptic' or silent gene clusters. This also is important as only a small handful of organisms can actually be cultured within a lab, and those which which can be cultured are probably not in the correct conditions to activate the production of all the secondary metabolites that they could produce. With sequencing techniques we now can access these genes expanding the number of unharnessed compounds available to us.

1.13.2 The aims of this work

The aim of 'directed biosynthesis' is to engineer biosynthetic systems that make therapeutically relevant natural products and harness them so that new compounds are produced, which can then be tested alongside conventional products of semi-synthesis. This is a formidable undertaking, because the production line of an NRPS is very complex, and also is tightly linked to the large number of auxiliary enzyme catalysed steps in the pathway. However, significant progress has already been made in this field. For example, broad-specificity glycosyltransferases have been used *in vitro* to attach unnatural sugars to vancomycin aglycone.[104] Another example is the reprogramming of the calcium-dependant antibiotic (CDA) NRPS.[105] This was achieved by a mutation in the conserved-serine residue on the PCP of the first module, inactivating the PCP as it is no longer able to tether the intermediate. *N*-acyl-serinyl *N*-acetylcysteamine thioesters with 'unnatural' chains were able to take the role of this PCP and the acyl chains were then incorporated into the NRPS product generating CDA analogues by 'mutasynthesis'.[105] Other examples of modifications of PKS and NRPS have involved the swapping of domains to generate altered compounds. In the case of balhimycin it has even been possible to add an entire extra module, to produce an extended peptide chain.[106]

All of the above examples, however, have revealed problems: domain swapping often leads to greatly reduced yields due to structural incompatibility, and post-NRPS enzymatic steps are often disrupted. Chapters 2 and 3 of this work are concerned with the mechanism of precursor formation and of post-NRPS modification within the chloroeremomcyin gene cluster, with the hope that a greater understanding of these steps will further our ability to redirect biosynthesis. The work in Chapter 4 describes a new approach to the genetic engineering of A-domains of NRPS modules, which might in future further our ability to modify them and develop new compounds of clinical interest.

2

Substrate Specificity of 4-Hydroxymandelic Acid Synthase

2.1 Introduction to dioxygenases

Dioxygen activation by non-haem, iron-containing oxygenases are an important feature of many living organisms. The oxygenaseses catalyse a wide variety of complex oxidations, including carbon-carbon bond cleavage, monohydroxylation and dihydroxylation.[107, 108] Despite both the high availability of molecular oxygen as an oxidation reagent for aerobic organisms and a large thermodynamic driving force; a large kinetic barrier to reduction of oxygen arises from the triplet ground state of dioxygen, which is spin-forbidden from direct reactions with ground-state singlet molecules.[109] This kinetic barrier is advantageous to the cell as it allows for the control and localisation of dioxygen chemistry within aerobic life. Dioxygen-dependant oxygenases overcome this barrier by promoting the formation of species capable of single-electron chemistry and which promote the spin inversion.

Enzymes that activate and make use of molecular oxygen fall into three main classes of oxidoreductase: oxidases, which catalyses a reaction utilising molecular oxygen as the electron acceptor; dioxygenases, which catalyse the incorporation of both atoms of oxygen from the dioxygen substrate into the substrate; and mono-

oxygenases, which incorporate only a single atom of the oxygen into the substrate with the second atom reduced to form water.

Nearly all enzymes that catalyse the reduction of dioxygen make use of at least one cofactor that is able to mediate the delivery of electrons to molecular oxygen within the active site. Two types of cofactors are generally found within these systems. The first is an organic prosthetic group which is capable of stabilising a radical, an example being the isoalloxazine ring system of flavin adenine dinucleotide (FAD), and the second is a co-ordinated redox-active metal ion such as iron, manganese or copper. Hybrid systems involving both organic cofactors and metallic ions are common too, for example the metalloporphyrin ring structures of haems. The most common subgroup of dioxygenases comprises those collectively known as α-ketoglutarate- or α-keto acid-dependent oxygenases (αKAOs), which are responsible for catalytic reactions that are essential in the biosynthesis of many primary and secondary metabolites.[110, 111] As their name suggests, they are distingished from other non-haem iron enzymes by their dependence on the presence of an α-keto acid, as well as Fe(II) and molecular oxygen, for catalytic turn-over. Examples of these enzymes include: prolyl hydroxylase, which oxidises specific prolyl residues of collagen;[110] clavaminate synthase for the biosynthesis of clavulanic acid (a β-lactamase inhibitor);[112] deacetoxycephalosporin C synthase responsible for the ring expansion of isopenicillin N to deactoxycephalosporin and the subsequent hydroxylation to deacetylcephalosporin C;[113] and 4-hydroxyphenylpyruvate dioxygenase (HppD), which converts 4-hydroxyphenylpyruvate to homogentisate as part of the tyrosine degradation pathway.[114]

Other classes of dioxygen activating non-haem iron-containing enzymes are: catechol dioxygenases, lipoxygenases and Rieske oxygenases. Catechol dioxygenases are responsible for the degradation of aromatic compounds and are of two types, namely intradiol-cleaving enzymes, which utilise Fe(III), and extradiol-cleaving en-

zymes, which utilise Fe(II) or less often Mn(II) (figure 2.1).[115, 116] Lipoxygenases are widely found among plants and animals and are responsible for the catalytic oxidation of unsaturated fatty acids that contain a *cis,cis*-1,4-pentadiene structure which is converted to an 1-hydroperoxy-*trans,cis*-2,4-diene. The mammalian lipoxygenases are of particular interest as they typically act upon arachidonic acid to produce hydroperoxides that are precursors to leukotrienes and lipoxins. These compounds have been linked as potential mediators of inflammation and hence lipoxygenases may provide a target for anti-inflammatory drugs.[117] Rieske oxygenases are involved in the initial steps of the oxidation of aromatic compounds found in the soil by catalysing the formation of diol precursors to the catechol dioxygenase substrates. The Rieske oxygenases require a mononuclear Fe(II) centre and a Fe_2S_2 cluster. The iron-sulphur cluster is responsible for the channelling of single electrons from NADH *via* a reductase. Due to their similarities (such as the requirement for an iron-sulphur cluster) Rieske oxygenases can be thought of as non-haem analogues of cytochrome P450 enzymes.

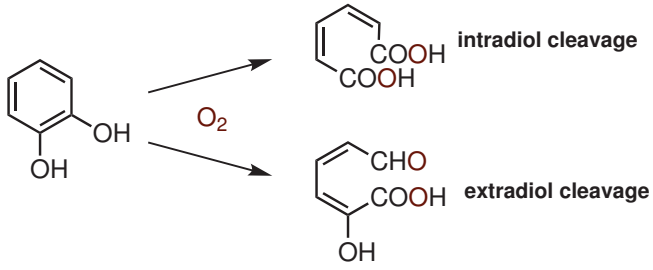


Figure 2.1: Modes of catechol dioxygenase cleavage.

α-keto acid-dependent oxygenases generally require three entities in the active site: α-ketoglutarate, molecular oxygen, and the substrate. Within this class of enzymes however, there are a selection of notable exceptions, those which require two or four substrates for catalytic turnover. Of interest here are the two substrate α-keto acid-dependent oxygenases, where the substrate acts as both the α-keto-acid (αKA)

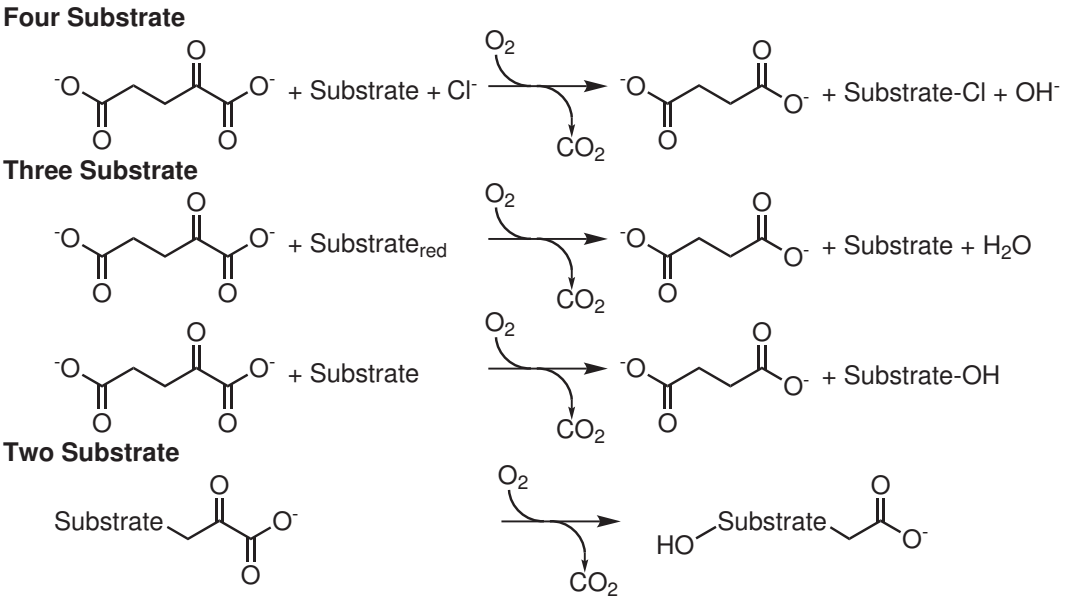


Figure 2.2: Generalised reactions of α-keto acid-dependent oxygenases (αKAOs) and related enzymes.[118]

as well as the substrate for hydroxylation. Four-substrate αKA-dependent halogeneses are structurally similar to the other α-KAO enzymes but use four substrates by an altered ligand arrangement around the metal ion (figure 2.2).[118]

In two-substrate αKAO enzymes, such as those discussed in this work, the iron atom is observed in a mixture of both six- and five-coordinate geometries in both the resting and substrate-bound forms.[119, 120] This differs from the three- and four-substrate αKAO enzymes in which αKA as a bidentate ligand associates with the iron in an octahedral resting state, displacing two water molecules and yielding no change in coordination number. Two substrate αKAO enzymes show a five-coordinate resting state like that of extradiol dioxygenases[121] and also show a greater similarity to the amino acid sequences of extradiol dioxygenases than the other classes of αKAO enzymes. This similarity in sequence is also reflected in their structure, with none of the two-substrate αKAO enzymes displaying the 'jelly-roll fold' of the three- and four-substrate αKAO enzymes.[122, 123, 124] Combined, these

observations suggest that two-substrate α-KAO enzymes may be the result of convergent evolution, in terms of reaction mechanism, with three- and four-substrate αKAO enzymes, yet are the result of divergent evolution from an ancestral extradiol dioxygenase.[118, 125]

The mechanism intially put forward for the αKAO-like enzyme HppD involved the formation of an iron-peroxy (Fe-O_2^{2+}) species.[126] In this model the iron peroxy species reacts with the quinonoid carbanion, formed by the deprotonation of the substrate, to give a peroxyquinone which nucleophilically attacks the keto group of the α-keto acid to form the cyclic dioxolane intermediate. Decarboxylation followed by an NIH rearrangement gives homogentisate (figure 2.3).

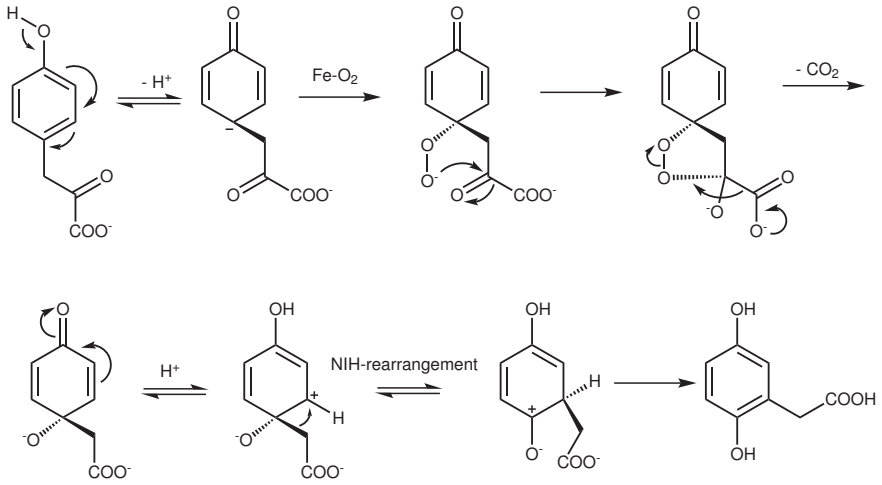


Figure 2.3: General reaction mechanism of an α-KAO-like enzyme according to Lindblad *et al.*

This mechanism however does not account for experimental data which showed that persuccinic acid, thought to be the active species in three-substrate α-KAO enzyme catalysed reactions, cannot substitute for the role of O_2 and α-ketoglutarate in prolyl hydroxylase. Secondly the per-acid did not inhibit the binding of the α-ketoglutarate.[127, 128] More relevant to the catalytic mechanism of the two-substrate

αKAO enzyme was that 2-(4-hydroxyphenyl)peracetic acid did not afford homogentisate with HppD.[129] Instead evidence for the existence of a iron(IV)-oxo species was detected in both thymine hydroxylase and HppD, in that both enzymes were shown to be able to catalyse sulphoxidations.[130, 131] This behaviour, as noted by Thornburg *et al.*,[130] is typical of cytochrome P450 enzymes and suggests the presence of this high valent iron-oxo compound. In conjunction with this, Lindblad *et al.* found that ^{18}O was incorporated from $H_2{}^{18}O$ into the product of HppD reactions. The mechanistic explanation of the predicted iron(IV)-oxo species was first proposed by Crouch *et al.*[132] The suggestion was that the substrate, 4-hydroxyphenylpyruvate (4HPP), and molecular oxygen co-ordinate to the metal and that decarboxylation of the substrate then forms the activated Fe(IV)=O species. This species is then able to convert the benzene ring to an epoxide, which can undergo an NIH-shift leading to the final product, homogentisate.

2.2 Previous work on HmaS

2.2.1 The search for HmaS

(*R*)-4-hydroxyphenylglycine (Hpg), found at positions 4 and 5 of the chloroeremomycin heptapeptide backbone, is one of the many non-proteinogenic amino acids found in NRPS products such as glycopeptides. As there is an epimerase domain associated with both modules responsible for the loading of Hpg onto the peptide chain at both position 4 and 5 respectively, it was assumed that (*S*)-4-hydroxyphenylglycine was the precursor for the (*R*)-4-hydroxyphenylglycine residues. It was noted in feeding experiments that Hpg was formed from (*S*)-tyrosine,[50, 133] and that homogentisate, an intermediate in the tyrosine degradation pathway, had the correct carbon skeleton for (*S*)-Hpg. The enzyme that produced homogentisate, HppD, is a well-characterised enzyme that converts 4-hydroxyphenylpyruvic acid to homogentisic

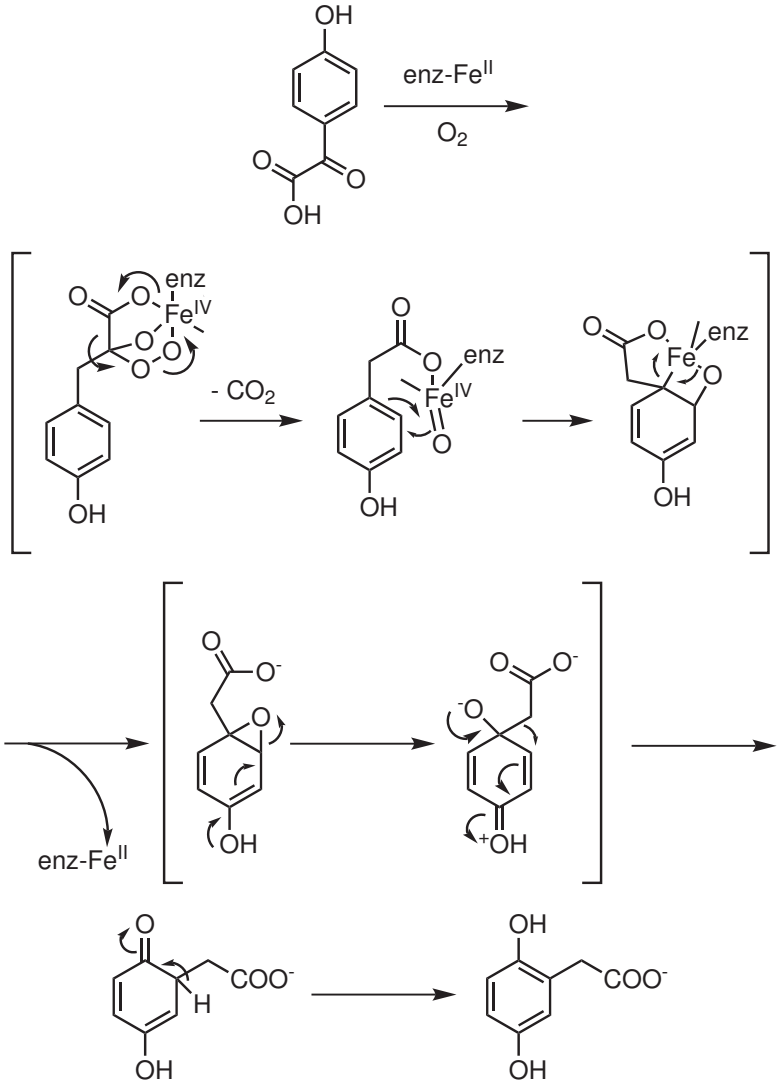


Figure 2.4: Reaction mechanism of HppD involving an iron(IV)-oxo species as proposed by Crouch *et al.*

acid and if hydroxylation were altered to occur at the benzylic position and not at the *ipso*-ring position it would produce 4-hydroxymandelic acid (Hma) (figure 2.5).

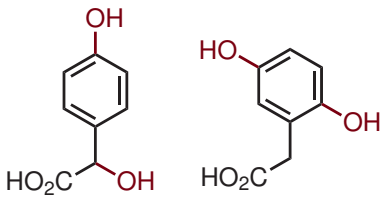


Figure 2.5: Comparison between 4-hydroxymandelic acid (left) and homogentisic acid (right), of note is that the carbon skeleton (black) is identical.

The chloroeremomycin gene cluster was subsequently searched for an open reading frame with similarity to HppD.[134] The use of BLAST searches and sequence alignments revealed that ORF21 within the chloroeremomcyin cluster showed significant similarity and identity to HppD. The highest identity was 54% with the nucleotide sequence of HppD from *Streptomyces coelicolor*.

2.2.2 Characterisation of HmaS[134]

To identify if ORF21 was able to catalyse the conversion of 4-hydroxyphenyl pyruvic acid into Hma, a 1105 bp DNA fragment of orf21 was cloned by Choroba *et al.* and ligated into pET28(+). The recombinant protein was overexpressed in *E. coli* BL21(DE3) and the His_6-tagged protein purified by the use of a metal-affinity chromatography column. The molecular weight of the protein was determined by electrospray mass spectrometry to be 40365 $\pm$ 20 Da, within experimental error of the mass of the protein calculated from its sequence, 40371 Da.

To determine the enzyme activity a reaction mixture containing the uncharacterised enzyme was incubated with 4-hydroxyphenylpyruvate. The reaction mixture was then monitored by reverse-phase HPLC and the product was detected as 4-hydroxymandelic acid at a wavelength of 273 nm. A more detailed analysis of the

product was carried out using GC-MS after extraction of the enzymic reaction mixture into diethyl ether and derivatising it with diazomethane. This result confirmed that the enzyme did not produce homogentisate from 4-hydroxyphenylpyruvic acid as would occur with HppD, instead the product was characterised as 4-hydroxymandelic acid.

2.3 Evolution of 4-Hydroxymandelic Acid Synthase[135]

The work described in section 2.3.2–2.3.6 was carried out by Dr H. M. O'Hare in this laboratory.

2.3.1 Introduction

HppD is responsible for catalysing the first dedicated step of the tyrosine degradation pathway. This enzyme has been the focus of research for many decades, as mutations within the enzyme cause type III tyrosinaemia, a rare inherited disorder. HppD is also of interest as homogentisate is a precursor of plastoquinone (a molecule involved in the electron transport chain during photosynthesis) and of tocopherols in plants. Inhibitors of HppD are therefore important commercially as herbicides.[136] The inhibition of HppD is also important clinically for the treatment of hereditary type I tyrosinaemias caused by the failure of the final enzyme in the tyrosine degradation pathway. By the inhibition of HppD the build up of toxic compounds in the body can be reduced, resulting in less severe symptoms similar to those of type III tyrosinaemia.[137]

HmaS and HppD show a high level of similarity of amino acids (34% identity) and use the same substrate for the reaction they catalyse within the cell. The mechanisms by which they operate also appear to be identical for the first few steps, only differing at the final step of the HmaS pathway. It is proposed that HppD forms the epoxide, instead of oxidation at the benzylic position, before undergoing an unusual

alkyl shift.[138]

The similarity in activity led to the suggestion that HmaS may have evolved from HppD over time.[139] This concept was an interesting one and, if true, would allow for the study of how a system changes over time to produce a different product. This is especially important to the work discussed here, as the overall aim of the work within the group is to manipulate enzymes in the biosynthetic pathway to create novel products, and understanding how nature has controlled and changed a reaction over time in its own manner is of much value.

In the crystal structure of HppD isolated from *Pseudomonas fluorescens*, three amino acids were revealed in coordination with the iron in the active site: histidines 161 and 240, and glutamate 322.[140] A hydrophobic substrate binding pocket was also observed, but the orientation of substrate binding was unknown at the time. The location of the residues coordinating the Fe^{II} helped with the alignment of the sequences from a large number of homologues of HppD. It was hoped that comparison of HppD with HmaS would suggest positions where mutations might result in interconversion of the activities of the enzymes. The choice of target sites was hampered by the fact that, although there were multiple crystal structures for the various homologues of HppD,[140, 141, 142, 143] none of these contained the bound substrate. No crystal structure was available for HmaS, with or without substrate. This meant that there was no direct evidence to identify which residues help bind the substrate. As a result, a saturation mutagenesis approach, combined with a high-throughput screen, was taken to interconvert the two enzymes.[135]

2.3.2 High-throughput activity assays

To screen the libraries of mutants that would be produced, a microplate assay was developed by Dr. H. M. O'Hare to screen for HppD and HmaS activities. *E. coli* lacks the tyrosine catabolic pathway, and as such contains no activity for the generation

or degradation of homogentisate.[144] Thus, expression of HppD in *E. coli* leads to a build up of homogentisate in the culture medium, which then undergoes spontaneous oxidation and polymerisation to ochronotic pigment providing a convenient assay.[145] The HppD genes from both *S. avermitilis* and *P. fluorescens* were used for this study as they both expressed well in *E. coli* and the strains produce concentrations of up to 300 mg/L of homogentisate, well above the detection limit of approximately 10 mg/L.

For HmaS the absorption spectra of hydroxyphenyl pyruvic acid and hydroxymandelic acid are too similar and so HmaS activity cannot be monitored by direct spectrophotometic assay. Fortunately, although Hma is not a secondary metabolite of *E. coli* a presumed oxidase with promiscuous substrate specificity converts Hma into 4-hydroxybenzaldehyde, which shows absorption maxima at 270 and 330 nm.[135] Using this assay the production of Hma in cultures over-expressing HmaS was measured to be about 300 mg/L, well above the detection limit of approximately 10 mg/L.

The mutant libaries were screened by both methods to allow for the identification of any mutants showed either a loss of activity or a switch of activity. The effects of mutagenesis would, it was hoped, give insight into whether (and how) these enzymes may have diverged over time.

2.3.3 Identification of residues crucial for HmaS activity

The five residues chosen for the initial mutations in HmaS were those that were conserved within the nine different HmaS homologues, and also conserved in the corresponding enzymes from the HppD family, but differing between the two enzymes. By looking at the various crystal structures of HppD, the residues were shown to be located in the active site. Four of these five residues were also found to be within close proximity of the inhibitor 2-[2-nitro-4-(trifluoromethyl)benzoyl]-

1,3-cyclohexanedione (NTBC) in the crystal structure of the *Streptomyces avermitilis* HppD.[142] The residues chosen (numbered according to HmaS from *A. orientalis*) were M199, T214, I216, I335 and Y339 and their equivalents L, P, N, F and F in HppD. A library of HmaS genes was constructed by saturation mutagenesis at each of the residues identified, giving five separate libraries each carrying a mutation at a single position. From each libary 96 clones were examined ($P > 0.95$, that a given codon will be sampled) and their activity classified as HmaS wild type, Hma producer or inactive. The results of the mutations of these residues were that, surprisingly, HmaS was found to be very tolerant of the mutations with none of them producing any detectable 4-hydroxyphenylpyruvate.. In particular, 60% of enzymes mutated at M199 and 39% mutated at Y339 retained wild-type activity and only 18% and 52% were completely inactive. This high tolerance of mutation suggested they were not catalytically essential and hence were not a target for further mutation. T214, I216 and I335 were the least tolerant to mutation and were targeted for further work.

2.3.4 Identification of residues crucial for HppD activity

Libraries were constructed for the equivalent residues to T214, I216 and I335 in HmaS (P214, N216 and F335) for both HppD from *S. avermitilis* and *P. fluorescens*. The results for HppD showed it to be much less tolerant of mutagenesis compared to HmaS, with significantly less wild type activity present in the mutants screened. In the libraries of mutants of the *S. avermitilis* enzyme at positions 214 and 216 the proportion of strains that showed wild-type activity was close to the proportion predicted to have no mutation. This result was originally taken to imply that all mutations at these sites resulted in disruption of the production of homogentisic acid.

2.3.5 Screening multiple mutants for altered production

Double mutants of both HppD and HmaS were then produced at other positions using the same set of locations as the first round. Again, no switch of activity was noted with the HmaS mutants. Over 1600 mutants were screened, including triple mutants of HmaS, and 41% displayed wild-type activity at a high level with 12% producing at wild-type yields. Of the 1488 HppD mutants with double mutations, 61 4-hydroxymandelic acid producing enzymes were discovered, showing that interconversion was possible. Even more interesting was that some of these mutants showed a greater production of 4-hydroxymandelic acid than homogentisic acid.

2.3.6 Characterisation of selected clones

The three wild type enzymes and four HppD mutants that displayed HmaS activity were selected for *in vitro* assays, and are listed in table 2.1.

Table 2.1: Reaction rates for selected clones

Enzyme	Mutations	HGA production (min^{-1})	HMA production (min^{-1})
HmaS	Wild type	<0.02	7200
HppD_SA	Wild type	560	<0.02
HppD_PF	Wild type	232	<0.02
SA1	N216F, F335V	0.088	0.404
SA2	N216L, F335V	0.126	0.302
SA8	P214T, N216P	0.187	0.364
PF1	N216G, F335V	<0.02	0.047

The rates of production of both products were determined by coupled spectrophotometric assay. HmaS activity was detected by the conversion of the Hma product to 4-hydroxybenzoylformate by hydroxymandelate oxidase while Hga production was detected by the conversion to 4-maleylacetoacete using homogentisate dioxygenase.

2.3.7 Sequencing of mutant libaries

The work by Dr H. M. O'Hare had successfully created a HppD mutant that displayed more HmaS than HppD activity. However, only a small sample of the most promising mutants had been analysed and after Dr H. M. O'Hare's departure I continued with this work. The mutant libraries were retransformed and screened. The single mutant *S. avermitilis* HppD clones that still displayed wild-type activity were sequenced. This confirmed that all the active mutants still contained proline at position 214 and asparagine at 216. Position 335 was found to tolerate hydrophobic substitutions (valine, leucine and methionine) without affecting homogentisate yield. The *P. fluorescens* HppD libraries showed a much higher proportion of clones with wild type activity, six mutants showing wild-type activity were sequenced from each libary. This revealed a greater diversity in amino acids compared to the *S. avermitilis* HppD mutants with a variety of substitutions: leucine and alanine at position 214, glutamine, methionine and valine at position 216, and leucine and valine at position 335.

2.3.8 Discussion

Wildtype HmaS and HppD, although very similar, display impressive selectivity, with each enzyme producing a single product with greater than 99.9% accuracy. In an assay capable of detecting product formation at a turnover rate of 0.02 min^{-1} (2×10^4-fold lower than the k_{cat} of HppD from *S. avermitilis*), no side products from either of the enzymes could be detected. In contrast to this, mutations introduced into the active site of HppD were able to promote a switch in production from homogentisate to hydroxymandelic acid. Hma production was relatively frequent among libaries of mutants of both homologues of HppD, which is surprising as HppD from *P. fluorescens* has a relatively low amino acid identity of 26% with HmaS and 34% with HppD from *S. avermitilis*.

Independently, whilst this study was underway, Gunsior *et al.* also demonstrated the potential for HppD to catalyse Hma production.[138] Four individual residues of HppD from *S. avermitilis* and their equivalent residues in HmaS were targeted for mutation. The analysis of the products by HPLC showed that two mutants, N216I and F335I, yielded a mixture of products of both Hma and homogentisate. Importantly they identified a third product, oxepinone. This product is particularly interesting as it supports the idea of a benzene oxide intermediate in the proposed HppD reaction mechanism (see figure 2.6).

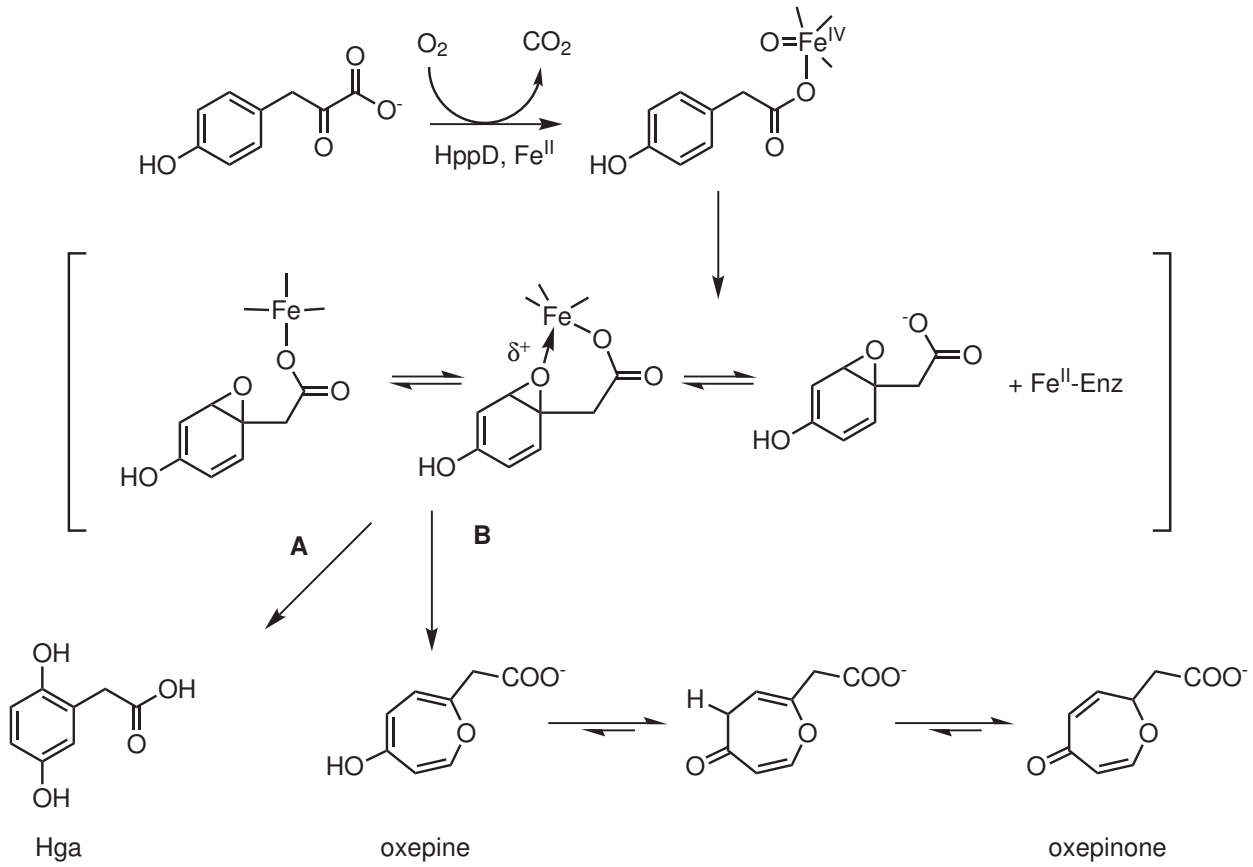


Figure 2.6: Proposed mechanism of Hga and oxepinone formation by Gunsior *et al.*[138]

The resilience of HmaS against mutations resulting in homogentisate production, suggested that the production of homogentisate requires an active site with a much more precise configuration than active sites that produce Hma.

The proposal for the catalytic process of each enzyme was that both enzymes follow the same initial enzymatic steps to give the Fe(IV)=O intermediate (see figure 2.4). This is supported by the results of this study. The more stringent requirements of homogentisate production suggest requirement for a greater level of steric control for the formation of the epoxide intermediate. As over 60 HppD mutants were detected that were capable of formating Hma and as the new residues were not limited to those found in the wild-type HmaS, it is hypothesised that the mutants cause a switch in activity not by mimicking the function of HmaS but by blocking the subsequent steps required for the production of the oxepinone or homogentisate, and as such Hma is simply produced as a side product. The idea of a blocking of wild-type activity fits with the maximum turnover rate of the HppD to HmaS mutant detected as only 0.4 min^{-1}, whereas the rates of wild-type HppD and HmaS are 560 and 7200 min^{-1}, a result of $>$99.9 % loss of activity.

The evolution of genes like HmaS poses a difficult question, as their only role is the synthesis of building blocks for secondary metabolites, in this case glycopeptides, and the gene itself serves no role within primary metabolism. This is also true of many of the other sugars and amino acids whose intermediates also serve no function except for the incorporation into the final antibiotic. It is therefore often difficult to see how these pathways could have evolved. The evolution of new catalytic activity is proposed to arise from evolutionary pressure leading to improvement and selection for a promiscuous activity of an enzyme. It seems likely that HmaS evolved from HppD, and over time has lost the more complex structure required for the latter half of HppD catalysis. This is further reinforced by the result that it only requires two mutations to convert wild-type HppD into an enzyme capable of the formation

of Hma and another two enzymatic steps to convert Hma into the final product, 4-hydroxyphenylglycine. It was found within *E. coli* that a pre-existing enzymatic pathway was able to catalyse the first of these two steps to give millimolar concentrations of 4-hydroxybenzoylformate, reconstituting two of the three steps required for 4-hydroxyphenylglycine biosynthesis.[49]

2.4 Stereochemistry of HmaS products

To further investigate the mechanism of HmaS and HppD, a sample of HmaS was donated to Professor E. Solomon and colleagues. Their EPR based studies of HmaS and HppD confirmed that the oxygen activation, active site structure and the binding of substrate in both HppD and HmaS are very similar.[119] During this work several alternative substrates were tested on these enzymes, one of which was phenylpyruvic acid. This substrate had previously been used in other work to demonstrate that the mechanism of HppD did not require the hydroxyl group on the ring para- to the pyruvic acid side chain. The data generated by Neidig *et al.* produced an unexpected result and led his group to postulate that the phenylpyruvic acid may be binding in a different orientation to that of 4-hydroxyphenylpyruvic acid, and should produce the reverse stereochemistry in any products produced.[146] Should this be true, it would have large implications on the reactivity and mechanism of the enzyme. As a result, it was felt that further investigation was required.

To study the stereochemistry of the enzymatic products, a method was adapted from the initial assay used to identify the activity of HmaS.[139] The protein was incubated in the same conditions with the new substrate (phenylpyruvic acid, Fe^{2+}, DTT and ascorbic acid) at 30 °C. After 2 hours the protein was precipitated and the product was extracted into diethyl ether and esterified with methanol/conc. HCl(aq), 9:1 (v/v), at room temperature for 2 hours. For experimental see section 5.2.1.

Since the derivatization differed from the method used previously,[139] it was first

tested on spectroscopically pure samples of the acid obtained from Sigma-Aldrich. The structure of the product was confirmed by the use of proton NMR and the stereochemistry was analysed by the use of a polarimeter. The results showed good optical purity with a specific rotation of α = +179.2° (*S*) and α = -182.1° (*R*) compared to α = +178°/-176° in the literature.[147] The NMR spectra also matched previous studies.

The stereochemistry of the natural HmaS product had been originally confirmed by the use of a known stereoselective enzyme, the product and substrate of which were known and could be detected.[134] This, however, was not a possibility for detection of the stereochemistry of the product phenylpyruvic acid as the reactivity of the previous test enzyme towards mandelic acid was unknown. Instead, it was decided that the stereochemistry of the product could be determined using a chiral HPLC column. The two standards were both analysed separately and then together using a Chiracel OD column. The two enantiomers showed a 2 min baseline seperation and there were two clearly defined peaks for the mixture of enantiomers.

Disappointingly, while the standards gave consistent retention times, the analysis of the extracted enzyme products was unsuccessful by this method. The resultant mixture was found to have a very high background due to the low product yield, making determination of the products difficult. 'Spiking' the extract with either standard was attempted, but no satisfactory results could be obtained. Due to these difficulties, GC-MS was selected as an alternative method of detection. This had previously been used for the detection of products from similar HmaS assays, however only with an achiral column which did not allow resolution of enantiomers. To resolve this problem the derivatised products were analysed by GC-MS on a chiral column. For experimental see section 5.6.2. Results with the chiral column proved successful and showed a 20 second separation between stereochemically pure standards. The derivatised product of the enzymic reaction was analysed by the same method and produced clear results, with the mass easily detectable and identical fragmentation

patterns to that of the standard (see figure 2.7). The reaction was spiked and analysed once with each of the standards to confirm the stereochemistry of the peak being monitored.

2.4.1 Results and discussion

The detected products showed only ca. 60 % e.e. of (*S*)-mandelic acid, (i.e. that having the same absolute configuration as the product from the natural substrate). This result seems to corroborate the initial hypothesis of Neidig *et al.* that the binding of the substrate was altered, though I did not see the total inversion of stereochemistry that was proposed. It would appear that the lack of the hydroxyl group on the substrate's aromatic ring causes reduction in the specificity of binding, and leads to a mixture of products.

2.5 Crystallisation trials of HmaS and SA1

To further investigate HmaS and HppD and explain their relative reactivities, work was undertaken in an attempt to obtain crystal structures. Several HppD structures already exist, although none of them have been solved with the substrate bound. HmaS, however, had never been crystallised.

Meanwhile, of the HppD structures already available, two of them had been crystallised in the presence of the inhibitor 2-[2-nitro-4-(trifluoromethyl)benzoyl]-1,3-cyclohexanedione (NTBC).[142] It was thought that the crystallisation of HmaS might be promoted if attempted in the presence of this inhibitor (NTBC), as it may stabilise the structure of the enzyme.

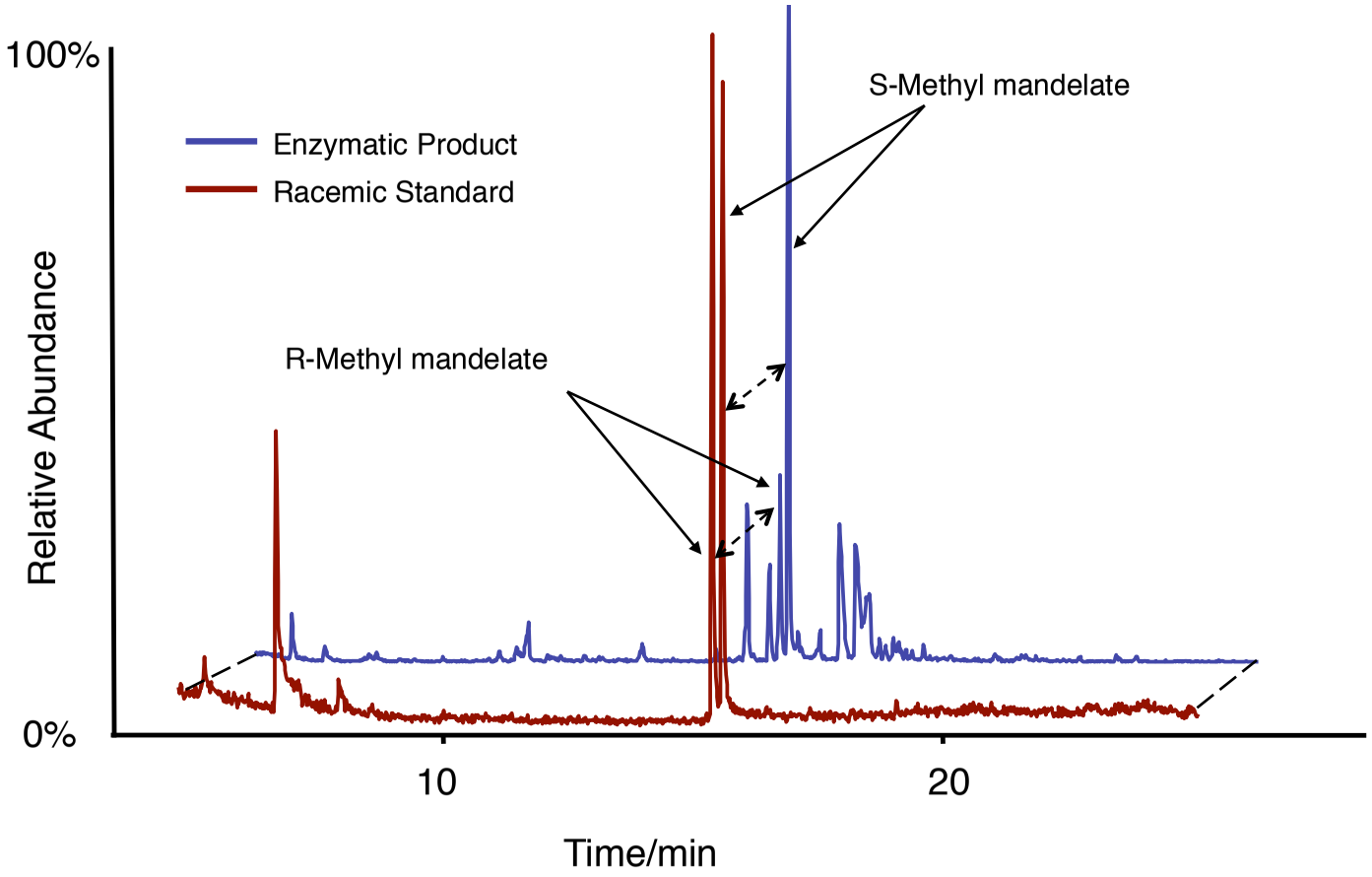


Figure 2.7: Chiral GCMS analysis of the products of the reaction phenylpyruvic acid catalysed by HmaS. The observed peaks in the enzymatic reaction were confirmed as mandate rather than a contaminant by comparison of the fragmentation pattern with that of standards.

2.5.1 Screening trials

The screening trials were undertaken with the use of a crystallisation robot (Cartesian Honeybee, Genomics Solutions). Since crystallisation is very dependent on purity, the protein was further purified by gel filtration. The fractions containing the desired protein were identified by UV and SDS-PAGE, combined, concentrated and then stored at 4 °C, and the remaining fractions discarded. The time taken for the protein to elute from the column also confirmed that it was present as a tetramer as previously documented.[142]

The HmaS protein was first screened under multiple conditions using Hampton I, Hampton II (Hampton Research), Wizard I and Wizard II screens (Emerald Biostructures). The plate was left at room temperature for the droplets to reach equilibrium. The plate was checked multiple times over the following weeks for crystal formation. Out of the 192 initial conditions tested by the screens, only two produced results that were worth investigating further. Both were from the Hampton Screen I: condition B3 produced small needle-shaped crystals, while C7 only produced micro-crystals, neither were large enough for diffraction experiments (see table 2.2 for conditions).

Table 2.2: Crystallisation conditions for HmaS that gave initially positive results

Hampton Screen I B3	Hampton Screen I C7
30% PEG 8000	10% PEG 8000
0.1M Na Cacodylate	10% PEG 1000
0.2M $(NH_4)SO_4$ pH 6.5	
10% PEG 1000	

These two conditions were then scaled up, screened more closely with a selection of variations (which were picked around the target conditions) and set up by hand using the hanging drop method. The plates were checked over the course of several weeks to allow for the well in the plate to equilibrate. For the B3 condition, 24 con-

ditions were investigated, however none produced crystals. Either phase separation occurred within the drop or it remained clear. For the C7 condition, another 24 conditions were checked, however these resulted in either precipitation of the protein within the drop or it remained clear. These conditions were repeated several times with varied conditions, however no protein was successfully crystallised.

2.6 Synthesis of NTBC

2.6.1 Introduction

Two of the wild-type HppD homologues have only been crystallised in the presence of NTBC so it was thought that the presence of the inhibitor would promote the crystallisation of the enzymes. NTBC is administered as Nitisinon (Swedish Orphan Int. AB) for the treatment of type I tyrosinaemia (T1Y). This potentially fatal disease is caused by a failure of tyrosine degradation. In a person with T1Y, the final step of this pathway fails, causing the build up of fumarylacetoacetate, which is then converted into other toxic compounds rather than being broken down. NTBC is an effective treatment for this disorder, as the inhibition of HppD slows down the oxidation of 4-hydroxyphenylpyruvate (which is in equilibrium with L-tyrosine in the presence of tyrosine aminotransferase) and reduces the level of fumarylacetoacetate present in the body. Because of its use as a drug and the patent covering it, the compound is not easily available and attempts to acquire it from Swedish Orphan Int. AB have been unsuccessful. Instead a synthesis of the compound was attempted from commercially available compounds.

2.6.2 Synthesis

Retrosynthetic analysis (see figure 2.8) of the structure of NTBC showed that the compound could be easily formed in a single-step reaction.

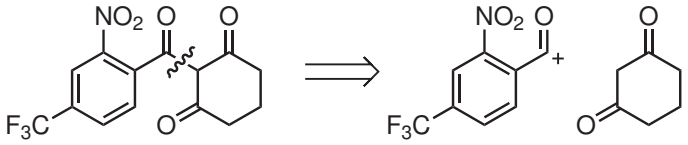


Figure 2.8: Retrosynthetic analysis of NTBC

The coupling of the two reagents was planned *via* activation of the acid and addition of the de-protonated diketone to act as a nucleophile. This follows the same concept that is used in the patent for the synthesis of NTBC, which uses the acyl chloride as the activated form of the acid[148] and a mixture of the diketone and triethylamine as the nucleophile.

The first method attempted for the synthesis of NTBC used 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) to activate the acid in dichloromethane (DCM) before the addition of the nucleophile and catalytic triethylamine (see figure 2.9C). The reason for choosing this method was that it had worked well on similar reactions within the group, despite the known lower activity of the EDC-activated acids compared to the acyl chlorides. The side products can be removed by washing with acidified water, leaving the product in the organic layer. The reaction proved unsuccessful; TLC and ^{1}H NMR spectroscopy showed a mixture of compounds that were not all identified as well as large amounts of starting materials were still present. The reaction was repeated over a longer time but TLC still showed little, if any, of the desired product and the NMR was still dominated by high levels of both starting materials.

As a result the synthetic route was altered to go *via* the acyl chloride. The acyl chloride was formed by treating a solution of the acid in dry DCM with oxalyl chloride and DMF. After this, a mixture of triethylamine and the diketone in DCM was added to the activated acid. Work-up gave what was thought to be the product but in low yield and very impure.

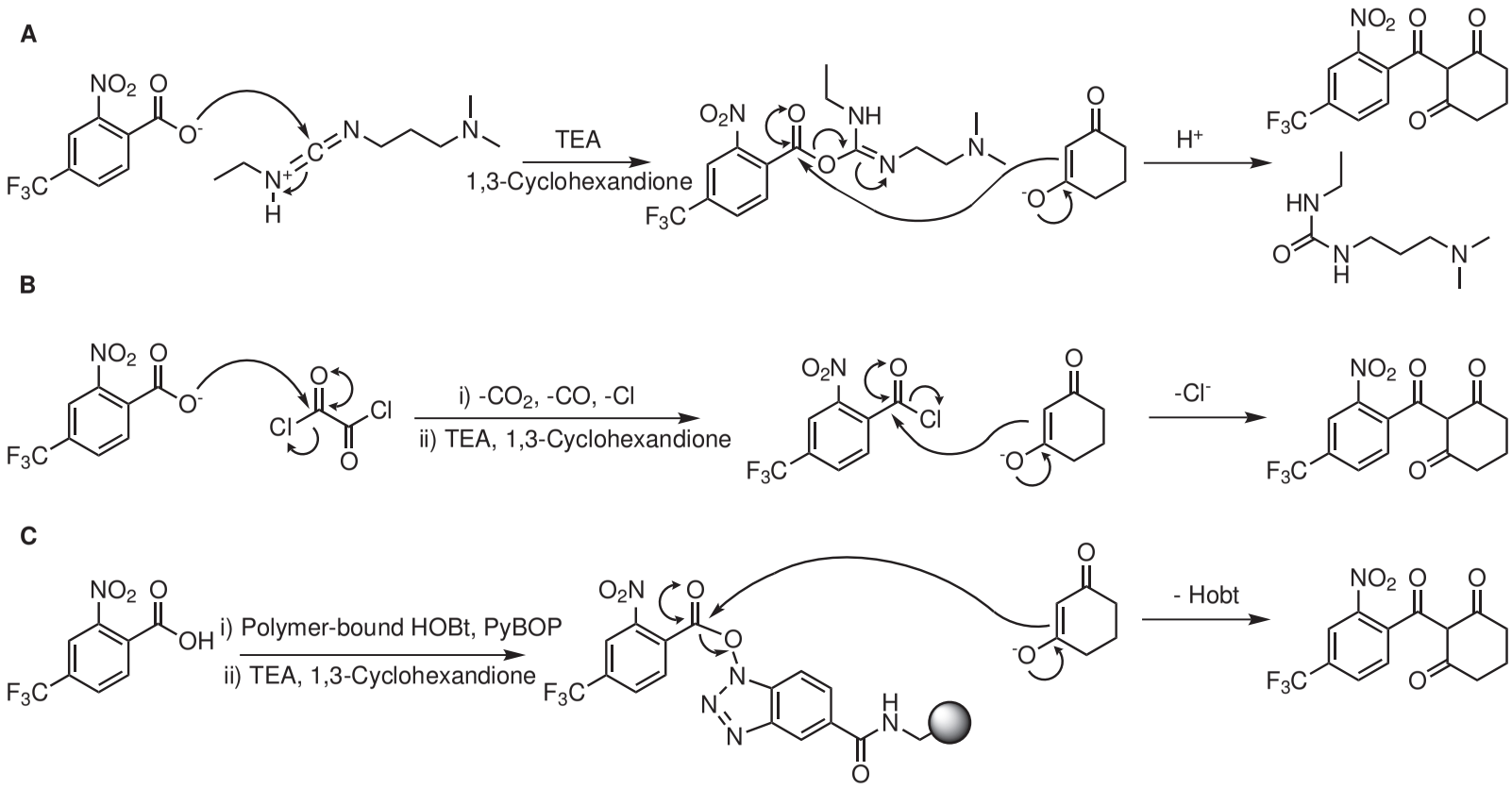
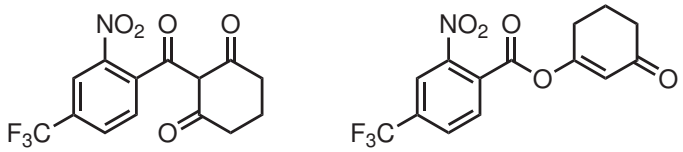


Figure 2.9: Outline of synthetic methods used for the synthesis of NTBC: A) EDC based synthesis of NTBC. B) Oxalyl Chloride based synthesis of NTBC. C) HOBt based synthesis of NTBC.

On discussion with members of the Ley group (Cambridge), it was suggested that a polymer-bound *N*-hydroxybenzotriazole (HOBt) be used. Initially this was attempted in a flow reactor but after several unsuccessful attempts it was decided that the reaction was too slow, so it was performed batch-wise. The polymer-bound HOBt was treated with benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) and *N,N*-diisopropylethylamine and the resulting resin was then washed and used to activate the acid and immobilise it overnight. The resin was then washed and re-suspended in dry DCM and the diketone/base mixture added. After reaction overnight the resin was filtered off and the product, after work-up, was purified on a small column of silica gel and analysed by NMR and HRMS. For experimental see section 5.2.2. The ^{1}H NMR spectrum suggested that the product was the isomer **2** of the desired compound in which the acid is coupled *via* one of the oxygen atoms of the diketone (figure 2.10). The evidence for this was the appearance of three different multiplets for the cyclohexane protons and a 1H singlet at δ 6.07 for the vinylic proton which was shown to be coupled to carbon atom at δ 120 by HMQC.



1 **2**

Figure 2.10: NTBC **1** Major product from the resin-based synthesis **2**

2.7 Concluding remarks

Subsequent to the work described here in sections 2.3.2–2.6.2 the crystal structure of HmaS (PDB:2r5v) was eventually solved by Brownlee *et al.*[149] to a resolution of 2.3 Å.The structure confirmed the importance of the three residues identified through

mutagenesis as critical for catalysis (T214, I216 and I335) and all were found to be involved in binding of Hma. Brownlee *et al.* proposed that the divergence in catalytic pathways between HmaS and HppD occurs after the formation of the reactive Fe(IV)=O intermediate (see figure 2.11). The two reaction mechanisms diverge with the ferryl-oxo species reacting with either the aromatic ring (HppD) or the benzylic position (HmaS), implying that it is the specific binding position of the Hpp and its orientation that is responsible for divergent products in the wild type enzymes. Brownlee *et al.*'s rationalisation of the tolerance of HmaS to mutations whilst maintaining activity and the inability to engineer the HppD phenotype into HmaS was based on the relative size of the substrate binding pocket (30 $Å^3$ for HmaS vs. 63 $Å^3$ for HppD).[149] This difference means that simple mutations are unable to confer the changes in binding position required to proceed *via* a HppD-type mechanism in HmaS as the tighter pockets holds the substrate in position with little flexibility, while the larger active site of HppD allows for small numbers of mutations to promote the shift to the binding required for an HmaS type binding. This is in contrast to our original postulate (see section 2.3.8) that suggested the differences between the active sites meant that some more complex motif of residues had been lost in HmaS that was able to stabilise intermediates in HppD; without this more specific active site the reaction pathway would be blocked and the 'default' product of Hma is produced. To some extent the original idea of a more complex structure change being lost in HmaS over HppD was correct. However, it seems that it is not, as was originally surmised, due to a stabilisation effect in HppD, but rather an evolutionary reduction in HmaS active-site volume that has led the latter to become less versatile and lose the ability to carry out its original catalytic function. HppD on the other hand is unable to reduce the volume of its active site and maintain catalytic activity as it is required to accommodate a number of more varied structures due to the hydroxylation of the aromatic ring and ensuing NIH shift. Even so the fact that wild-

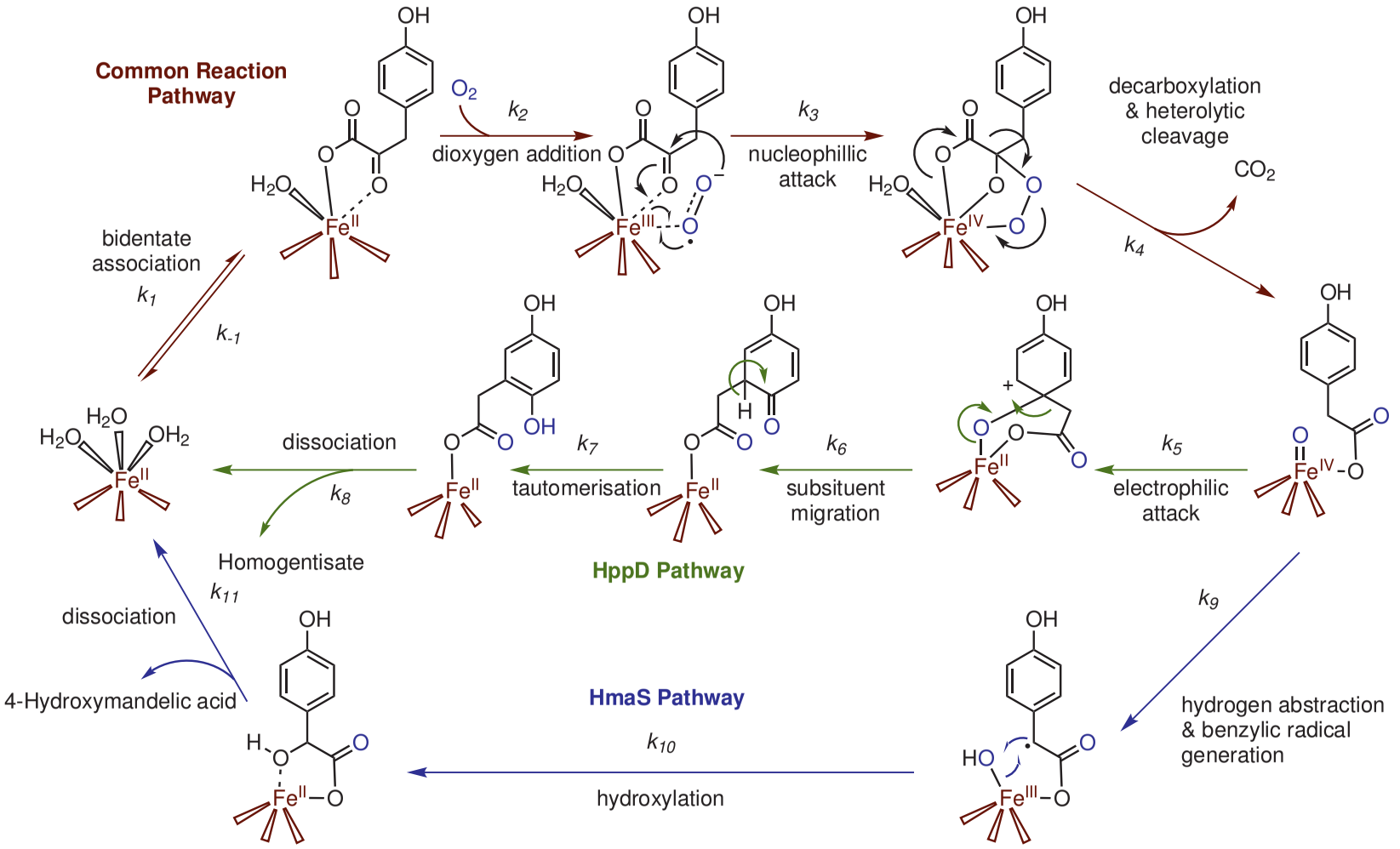


Figure 2.11: Comparison of the HmaS and HppD mechanisms.[149]

type HppD shows no detectable HmaS activity is remarkable when it is considered how similar the two enzymes are, and how few mutations of residues are required to allow the altered binding of 4-hydroxyphenylpyruvic acid.

3

Mechanistic Analysis of Cross-Linking in Glycopeptides

3.1 Introduction

3.1.1 Cross-linking in glycopeptides

The peptides produced by the different glycopeptides NRPSs are all extensively modified both during and after their biosynthesis. Particularly striking features of the peptide backbone of vancomycin-like glycopeptides are the two aryl ether cross-links between the aromatic residues 2-4 and 4-6, and the biaryl crosslink between the aromatic rings of residues 5 and 7. The crosslinking is catalysed by a set of oxidative cytochrome P450 enzymes (OxyA, OxyB and OxyC) to form the tricyclic aglycone structure.[150] It is these cross-links that give the peptide its rigid structure and they are vital for its ability to bind the D-Ala-D-Ala moiety of the nascent bacterial cell wall. It has also been shown by heterologous complementation that the oxidative crosslinking enzymes of *A. orientalis* and the balhimycin producer *Amycolatopsis balhimycina* are functionally equivalent, as those from balhimycin are able to form the cross-links between aromatic residues in the structurally equivalent peptide chain formed by the chloroeremomycin NRPS.[150] However, despite a large amount of research into the field of cytochrome P450 enzymes, the mechanism by which these

cross-links are formed is comparatively poorly understood.

3.1.2 P450 enzymes

Cytochrome P450 is from the family of enzymes called oxygenases. Oxygenases are a group of enzymes that add one or more oxygen atoms to a substrate. The first P450 enzyme was isolated from rat liver in 1958.[151] These enzymes were given their name because they exhibited an unusual red-shift at 450 nm in their absorption spectrum,[151] and as such were originally reported as a novel 'microsomal carbon monoxide-binding pigment'.[152, 153] The eventual elucidation of its function as an oxygenase in 1963 triggered a rapid expansion of research into this haemoprotein. Annual numbers of papers dealing with cyctochrome P450 have increased rapidly ever since and now cytochrome P450 is regarded as the collective name of a large number of haem-containing proteins. The pigment was eventually shown to be a *b*-type cytochrome containing an iron-protoporphyrin IX prosthetic group (see figure 3.1) by Omura and Sato[154, 155] who named it for the first time as a cytochrome P450.

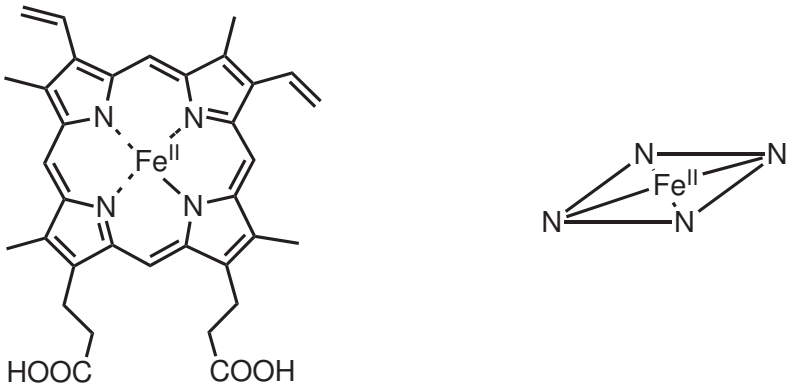


Figure 3.1: Structure of the iron-protoporphyrn IX (haem) (left) and its schematic form (right).

The characteristic shift of the Soret band from 394 nm for the high-spin ferric species or 418 nm for the ferric aquo complex to 450 nm on the formation of the ferrous carbon monoxide complex, was rather unusual when compared to other Fe-

porphryin containing proteins such as peroxidases and myoglobin. It was Mason *et al.* who first suggested that this could be due to the proximal haem ligand, a cysteine sulphur (figure 3.2),[156] and this now been confirmed by molecular orbital calculations.[156]

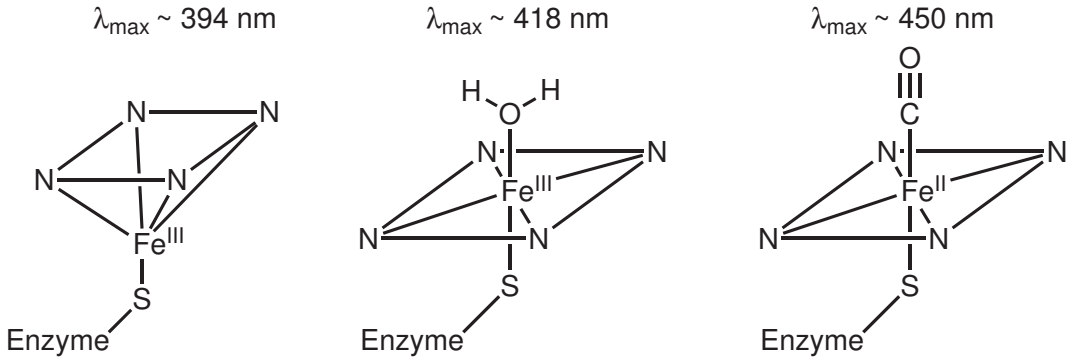


Figure 3.2: Schematic representations of iron-porphyrin complexes, Fe^{III} high-spin (left), Fe^{III}-aquo high-spin (middle), Fe^{II}-carbonyl low-spin (right)

Since then cytochrome P450-dependent mono-oxygenase enzymes have been found in nearly all living organisms[157, 158, 159] in which they catalyse the oxidation of a wide variety of substrates. In animals, for example, P450 enzymes and isozymes have been detected in almost all tissues that have been examined, including liver, kidney, lung, nasal membrane, brain, intestinal mucosa, bladder, testis, adrenal gland, aorta and many others (see reference Black *et al.*[160]). There are over 3000 known P450 enzymes, and some are highly selective, such as those used in the formation of steroids, while others are far less selective including ones which convert toxins to more soluble forms that are readily removed from the organism. Despite this wide range of activities they all seem to have diversified from a single ancestral protein during the course of biological evolution. It is an interesting contrast that in eukaryotes P450 enzymes are found mostly in a membrane-bound form, whereas bacterial P450s have no obvious membrane association.

Cytochrome P450 enzymes mostly catalyse oxidation reactions often with cleav-

age of kinetically and thermodynamically stable bonds, such as C–H. The activated oxyferryl-species formed by the enzyme is represented in multiple different forms, the simplest form being represented as 'Fe^{V}=O'. More commonly the depiction of the activated form involves the whole iron-porphyrin complex, whereby the electron distribution around the iron corresponds to an Fe^{IV} d^{4} system with a porphryin π-cation radical as its ligand, which can be written as 'Por$^{\bullet+}$/Fe^{IV}=O'. [161]

Monooxygenation (*e.g.* hydroxylation) is the main catalytic activity displayed by cytochromes P450. Other chemical transformations that have been characterised include: epoxidation, peroxygenation, N-, S- and O-dealkylation, desulphurisation, reduction of nitro and azo groups and formation of N-oxides and peroxides, deamination, dehalogenation, isomerization and non-hydrolytic carbon-carbon bond cleavage.[162, 163] Bacterial and mitochondrial P450s are generally found to be a more substrate specific group of enzymes with much higher regio- and stereospecificity compared to the microsomal P450 enzymes extracted from eukaryotes.

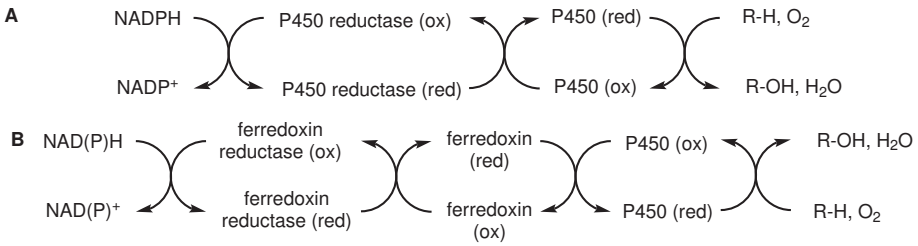


Figure 3.3: Supply of electrons to P450s. A) Microsomal P450s require a NADPH-dependent cytochrome P450 reductase. B) Mitochondrial or bacterial P450s require a ferredoxin and a NAD(P)H-dependent ferredoxin reductase: NADH is utilised by bacteria, NADPH by mitochondria. R-H is the substrate to be hydroxylated.

For the catalytic turnover of cytochrome P450-dependent enzymes a cofactor and reductive system is required. The microsomal system requires a NADPH-cytochrome P450 reductase and NADPH in addition to the haem-containing protein (figure 3.3.A), while the bacterial and mitochondrial systems need another step in the electron chain,

with electrons shuttled from NADH (in bacterial systems) or NADPH (in the mitochondrial systems). The microsomal reductive protein contains one riboflavin-5'-phosphate (FMN) prosthetic group at its *C*-terminus and a flavin adenine dinucleotide (FAD). In bacterial systems the ferredoxin reductase, another FAD-containing enzyme, accepts two electrons from NADH and then provides a single electron reduction of ferredoxin, an iron-sulphur cluster-containing protein. The ferredoxin is then able to catalyse a single electron transfer to the P450 as is required by the P450 catalytic cycle.[164, 165] In this manner the membrane-bound enzyme within the microsomal redox chain is replaced by the membrane-bound ferredoxin-reductase/ferredoxin pair in mitochondria or the cytoplasmic equivalent in bacteria.

Some of the earliest crystal structures of cytochrome P450s were those of $P450_{cam}$ from *Pseudomonas putida*,[166] $P450_{BM-3}$ from *Bacillus megaterium*[167] and $P450_{eryF}$ from *Saccharopolyspora erythraea*.[168] Their substrates and preferred position of hydroxylation are shown in figure 3.4. The most extensively studied of these and of all the P450 superfamily is $P450_{cam}$, which catalyses the hydroxylation of camphor regio- and sterospecifically to form 5-*exo*-hydroxycamphor. Extensive studies, including a variety of spectroscopic techniques, have provided a wealth of information on the catalytic behaviour of $P450_{cam}$, making it a model system for understanding the behaviour of P450 enzymes. Some of the techniques used include: optical,[169] EPR,[170] resonance Raman,[171, 172, 173] infrared,[174, 175] Fourier transform infrared,[176] NMR,[177, 178, 179] EXAFS,[180] and MCD[181] spectroscopy. These data combined with knowledge about other enzymes, such as horseradish peroxidase, have provided a deep insight into the reaction mechanism of P450s.

P450 mechanism

The catalytic cycle can be thought of as being divided into two main parts. The first is oxygen activation to form the Fe^{V}=O species. The second part is the interaction

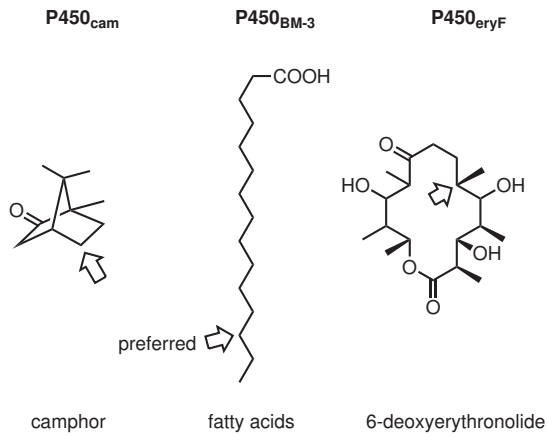


Figure 3.4: Natural substrates for P450$_{cam}$, P450$_{BM-3}$ and P450$_{eryF}$. Arrows mark the position of hydroxylation.

with and oxidation of the substrate.

The resting state of P450$_{cam}$ is the ferric-aquo complex shown in figure 3.2 with the porphyrin ring providing 4 equatorial ligands and the conserved cysteine from the protein backbone providing a proximal sulphur ligand. Within the active site of P450$_{cam}$ there are six water molecules, of which one is bound to the top of the haem, and the iron atom is out of the plane of the porphyrin ring towards the cysteine below. With this distal bound water the haem is six-coordinated and low spin. Evidence to confirm the presence of this bound distal-water was achieved by the use of ESR spectroscopy as well as computation studies.[182] The low-spin ground state of the ferric aquo-haem complex was unprecedented within the shell of an intact protein.[183] Semi-empirical calculations show the isolated form of the aquo-porphyrino-methylthiolato-iron(III) (S–haem–OH_2) complex to be high-spin. In P450s it is believed that the electrostatic field from the protein around this complex helps stabilise the low-spin state.[184, 185, 186] A distal water molecule alone is not thought to be enough to drive the complex to adopt a low-spin configuration and it is believed that hydrogen bonding between the active site water molecules is also required.

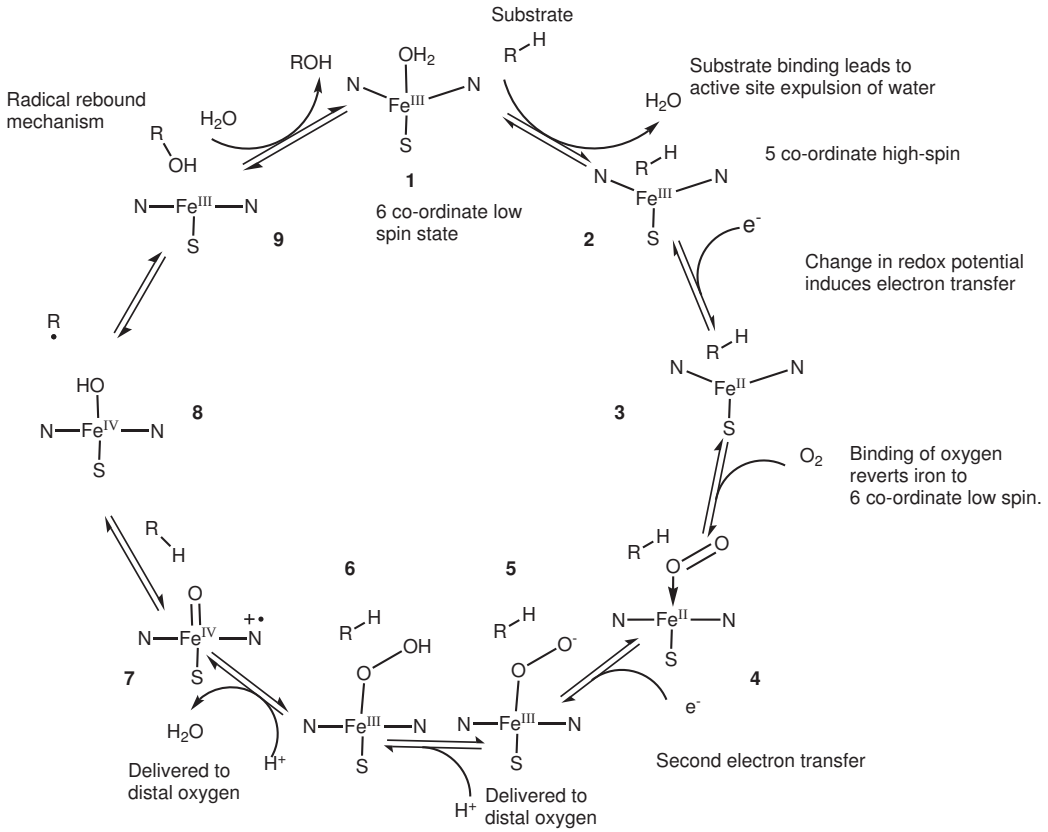


Figure 3.5: Overview of standard P450 cycle. Species **1–4** have been crystalgraphically determined, **5–6** have been spectroscopically charactised, and **7–9** are proposed intermediates

Binding of the substrate within the active site generally results in the expulsion of all the water molecules that occupy this space (figure 3.5 – step **1**). The expulsion of water and the binding of substrate results in a change of spin state of the Fe^{III} from low-spin to high-spin as shown by spectroscopic measurements including ESR and optical spectra and X-ray crystallography on both substrate bound $P450_{cam}$ and $P450_{eryF}$.[168, 187] The change of the iron centre, becoming five co-ordinate, is accompanied by an increase in the enzymes reduction potential from -300 mV to -170 mV. [188, 183] It has been traditionally thought that it is this increase in reduction potential that permits putidaredoxin (PdX) (-240 mV) to donate a single electron (step **2**), and hence it was thought to act as nature's gating mechanism to stop the enzyme wasting the resources of the host organisms.[189] Unlike the reduction in the presence of substrate in which the iron stays in the same conformation, reduction of the substrate-free form of P450cam requires a conformation change from six co-ordinate to five co-ordinate. According to Marcus Theory the larger reorganisation energy for the substrate-free form results in slower electron transfer.[189]

Once the ferric haem centre has been reduced to the ferrous form it is capable of binding a dioxygen molecule reversibly to form the oxyferrous sate, or bind carbon monoxide to generate the carbonylporphyrinatoiron(II) inhibitor complex. Calculations show that the oxygen binding in physiological concentrations increases the reduction protential, making electron transfer from PdX thermodynamically favourable. The oxyferrous complex can then either undergo the next electron transfer and form the ferric-peroxo (V) species or undergo auto-oxidation to form species **2** again and release superoxide (O_2^-). The superoxide has a half-life of a few minutes at room temperature, before disproportionation to water and peroxide.[190] The ferric-peroxo species is then rapidly protonated in step **5** to form the hydroperoxo intermediate (VI) which is protonated at the distal oxygen for a second time leading to O–O bond cleavage *via* the loss of water and formation of the proposed active intermediate **7**.

In comparison to the species **1**–**6** which have been characterised crystallographically, **7**–**9** are all proposed intermediates which have not been observed to date, however Laue diffraction studies on cytochrome c peroxidase have shown that the proposed active species **7**(or compound I) does exist in that system.[191] Other evidence for the formation of this active species comes from the study of P450$_{LM2}$ found in live microsomes where it generates hydrogen peroxide when incubated with dioxygen and NADPH in the absence of substrate. In addition hydrogen peroxide is found to be able to form the active species without the requirement of dioxygen or a reducing agent.[192] This is called the peroxide shunt pathway. These combined results support the theory of a *mono* oxygen species formed in the hydroxylation reactions and the involvement of a peroxo-iron intermediate.

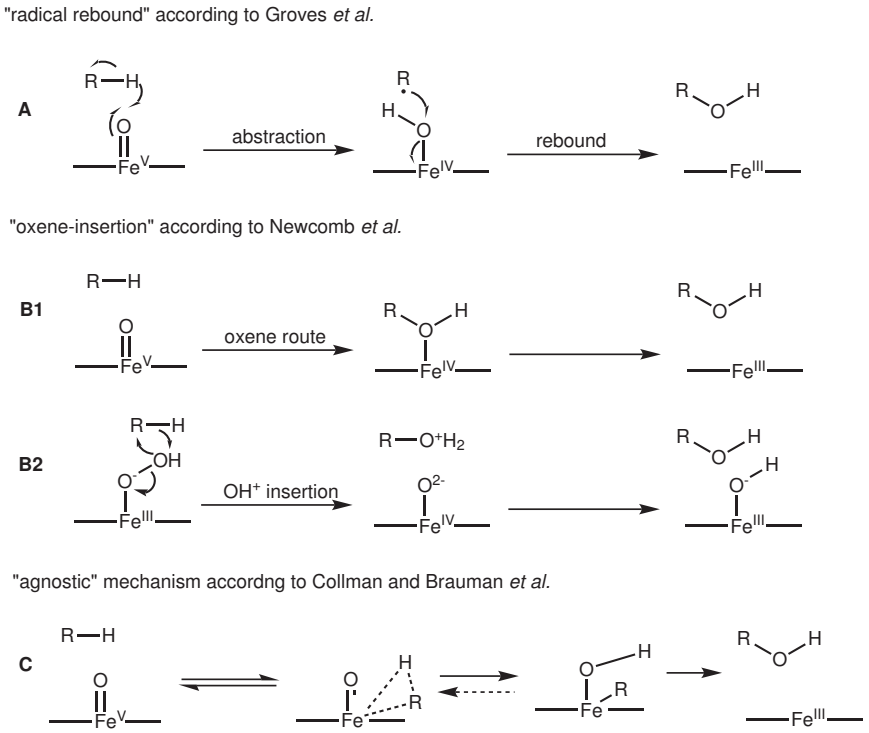


Figure 3.6: Three possible mechanisms for the second half of the P450 cycle.

The reaction of activated species has long been thought to proceed *via* a radical

mechanism.[161, 193] In the mechanism of OxyA–C it may very well be that steps 7, 8 and 9 follow the traditionally accepted radical rebound mechanism (figure 3.6A),[194, 195, 196] to perform the second part of the catalytic cycle, substrate oxidation. However, studies such as Newcomb's use of radical probes have led to the suggestion that an alternative mechanism may be possible, such as oxene-insertion[197, 198] (figure 3.6B1) or the agnostic mechanism proposed by Collman and Brauman *et al.* (see figure 3.6C).[199] Finally, the discovery of a second oxidizing species within the P450 active site in addition to the 'Fe^{V}=O species, a OH^{+} cation, have led Newcomb *et al.* to propose a possible 'OH^{+}' insertion mechanism (figure 3.6B2).[200]

In the radical rebound mechanism (figure 3.6A) intermediate VII abstracts a hydrogen atom from the substrate, to form VIII and the substrate-derived radical held in the active site. This radical then reacts with the Fe^{IV}–OH intermediate to form the product alcohol which can then leave the active site and allow the formation of **9**. It is unknown whether the P450 enzyme can go directly from **9** to **2** without the need for water to displace the product or if instead, it is required to go *via* the resting state **1**.

3.1.3 The mechanism of phenolic cross-linking reactions

One of the earliest examples of cytochrome P450 enzymes that catalyse an oxidative phenol coupling was an enzyme isolated from *Berberis stolonifera* plant cell cultures. It was shown to be responsible for the dimerisation of (*S*)-coclaurine and both the *R*- and *S*- enantiomers of *N*-methylcoclaurine, both regio- and stereospecifically *via* oxidative phenolic coupling to form three distinct dimeric products: (*R*,*S*)-berbamunine, (*R*,*S*)-2'-norberbamunine and (*R*,*R*)-guattagaumerine, (see figure 3.7).

The relevance of this early work is that the mechanism was proposed to proceed in accordance with the proposal of Barton,[201] that is, through the formation of two resonance stabilised phenoxy-radicals (see figure 3.9, mechanism A). This is essen-

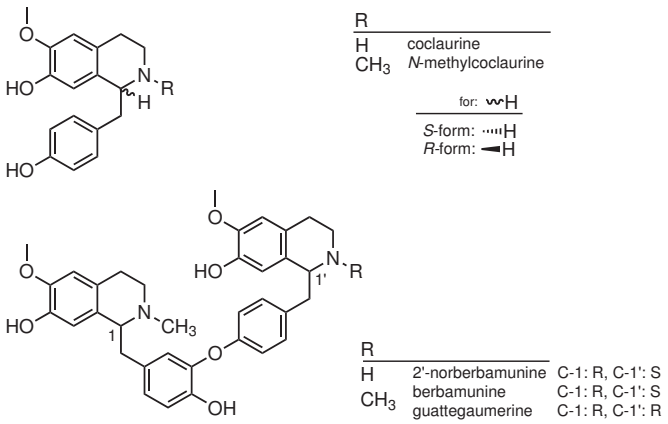


Figure 3.7: Structure of benzyltetrahydroisoquinolines and the dimeric bisbenzylisoquinoline alkaloids produced by oxidative phenolic coupling.

tially the same as the mechanism by which the phenolic cross-linking enzymes found in glycopeptide biosynthesis are thought to proceed.[150] Despite the widespread acceptance of this as a plausible mechanism, very little work has been undertaken to move it beyond a simple conjecture.

The only P450 enzyme that catalyses phenolic coupling to have been successfully crystallised with its substrate bound is CYP158A2[202] which catalyses the cross-coupling of two flaviolin molecules to form dimeric products. The diradical mechanism seems awkward in the context of this structure, as the phenol group of the upper molecule is blocked from the haem by the lower molecule (figure 3.8), an arrangement that would prevent direct abstraction of a hydrogen atom from the upper molecule by an Fe-O species. This apparent incongruity prompted us to consider other possible mechanisms for phenolic coupling that might be more consistent with this crystal structure. Unfortunately, the high-quality crystal structures available for both OxyB[203] and OxyC[204], do not contain substrate, and therefore do not help on this point.

As previously mentioned, P450 enzymes catalyse a wide range of reactions, including epoxidation of aromatic compounds. If epoxidation of one of the aromatic

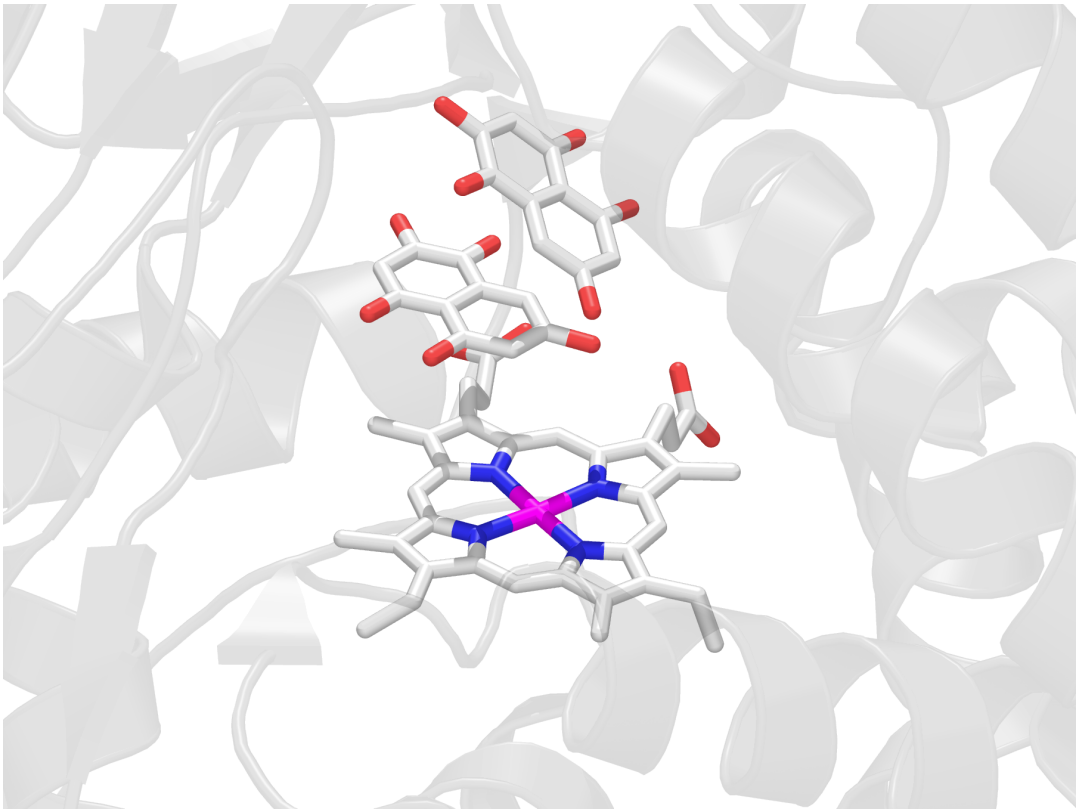


Figure 3.8: The binding of two flaviolin molecules in the active site of CYP158A2 shows the difficulty of the formation of a radical on the upper flaviolin molecule because of the large distance to the haem. (Generated from PDB code 1T93)

rings cross-linked during the biosynthesis of chloroeremomycin occurred, then phenolic coupling could take place by nucleophilic attack of the other aromatic ring on this epoxide to give a gem-diol intermediate (see figure 3.9, mechanism B). This mechanism is attractive because it only requires one of the aromatic amino acids to be close to the haem, as suggested by the crystal structure of CYP158A2. This might be experimentally probed by labelling the hydroxyl group on the aromatic ring with oxygen-18. Dehydration of the gem-diol intermediate could lead statistically to the loss of half of this label, and if this were the case it would provide evidence for this proposed mechanism. During the course of this study, independently, Robinson *et al.* also suggested, as one possibility, this mechanism with an epoxide as an intermediate and carried out an *in vitro* experiment under an atmosphere of oxygen-18 with OxyB, which catalyses the first phenolic coupling of the linear hexapeptide attached to the peptidyl carrier protein (PCP).[74, 205, 206] They found no incorporation of label into the product, suggesting that either the epoxide was not formed or that the elimination of water from the gem-diol is stereoselective. This chapter investigates this further using an *in vivo* study using an oxygen-18 atmosphere or oxygen-18 labelling of the peptide. This approach has the advantage of being able to probe all three phenolic coupling steps catalysed by OxyA, OxyB and OxyC.

In addition to the epoxide mechanism and the diradical mechanism, there are other possible mechanisms, which to our knowledge have not been proposed before. Mechanism C, shown in figure 3.9, involves initial formation of a radical on one of the aromatic amino acids. This radical then attacks the other aromatic moiety, and the newly formed radical abstracts the hydroxyl group from the haem in a similar manner to the rebound mechanism that is proposed for the hydroxylation of many compounds by P450 enzymes. This rebound mechanism would proceed through the same gem-diol intermediate as that proposed in the epoxide mechanism (mechanism B in figure 3.9); therefore these two mechanisms cannot be distinguished solely by the proposed

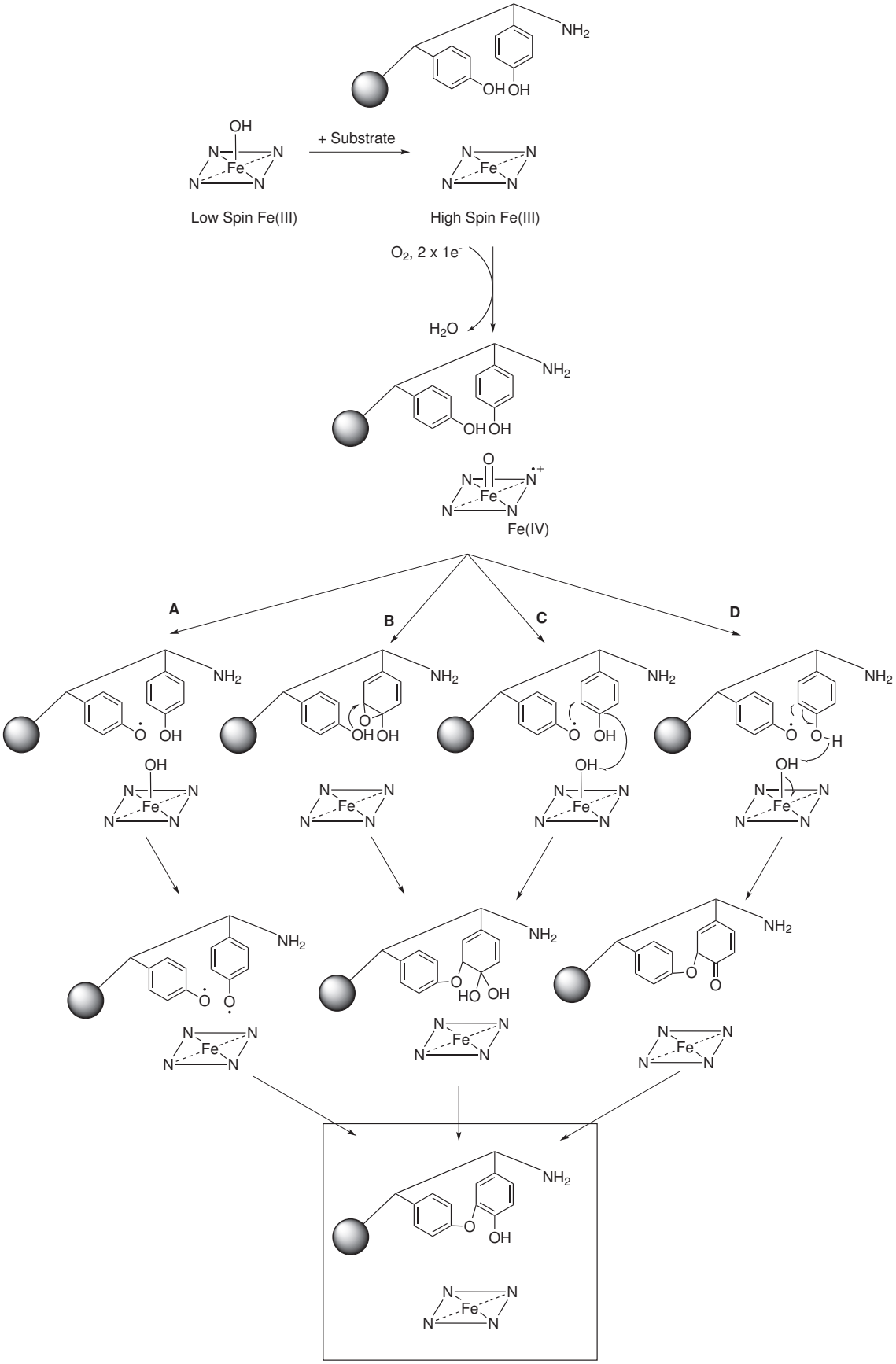


Figure 3.9: Summary of possible mechanistic pathways discussed in this work.

oxygen-18 labelling experiments. This mechanism, like the diradical mechanism, would require the catalytic iron to be close to both aromatic rings. Although the crystal structure of CYP158A2 suggests this is difficult, it may not truly reflect the flexibility of the active site and any changes in conformation as the reaction proceeds. Lastly, mechanism D (figure 3.9) is similar to mechanism A, except that the hydrogen atom of the phenol moiety is lost to the Fe-OH species as part of the coupling step rather than in a separate step. It is predicted this mechanism would, like the diradical mechanism, not involve any exchange of label from the phenolic hydroxyl group.

3.2 Growth of A. orientalis in an oxygen-18 atmosphere

To investigate these proposed mechanisms we aimed to grow the producing strain of chloroeremomycin in a nitrogen/oxygen-18 atmosphere. This would mean that any oxygen atoms incorporated from the atmosphere during the cross-linking reactions should be labelled. The labelled atoms could then be detected by the use of carbon-13 NMR as the presence of oxygen-18 should produce an upfield shift of the peak from the neighbouring carbon. This method had previously been used to discover the origin of the oxygen atoms within the structure of monensin.[207] In the case of 4-hydroxymandelic acid this shift was shown to be 2.4 Hz, when tested on our 500 MHz spectrometer (the sample was synthesized as described in section 5.2.4). The positioning and abundance of oxygen-18 incorporation could then be used to analyse the mechanism of the phenolic coupling reactions. To assign its ^{13}C NMR spectrum, chloroeremomycin was purified by HPLC from a lab standard and carbon-13 spectra in DMSO were obtained using a 500 MHz cryo-probe. To assign the carbon-13 spectra a partial assignment of the ^{1}H NMR spectrum by Dr. Peter Kirkpatrick[208] was used combined with HMBC and HMQC spectra of the compound. The assignment of critical peaks on the heptapeptide core was achieved. To confirm the assignment of these key residues, L-tyrosine with a carbon-13 label in the phenolic position was

fed to a culture. Tyrosine was chosen as, while it does not appear as the unmodified amino acid in the final glycopeptide and was commercially available, the amino acid is incorporated after modification as both 4-hydroxyphenylglycine at positions 4 and 5, and as β-hydroxytyrosine at both positions 2 and 6. The labeled tyrosine in water was fed to a culture of *A. orientalis* in four 10 mg portions over a period between 36-48 hours of growth. The culture was then harvested at 110 hours and purified by the use of a D-Ala-D-Ala column (see section 5.4.7) before submission for NMR. As expected, the phenolic carbons on the rings corresponding to the second, fourth, fifth and sixth residues (δ: 149.7, 133.92, 155.8 and 149.2 respectively) showed a five-fold increase in intensity.

For the growth of *A. orientalis* under an oxygen-18 atmosphere the seed culture was grown as usual (see section 5.4.6), before being spun down and re-suspended in a minimal quantity of fresh media to remove any chloroeremomcyin already produced. 50 mL of FermB media was then inoculated with 1 mL of re-suspended cell pellet. The system was sealed and air cycled through 100 mL of sat. NaOH in a 250 mL flask and a 2 L flask as an air reservoir, using a peristaltic pump to maintain flow. The system successfully allowed the culture to grow, but not consistently, and on runs where air was replaced with a 19% oxygen-18 in nitrogen mix, only a very small amount of glycopeptide (1% of normal production) was isolated. Analysis of these samples by MS showed no incorporation of oxygen-18, but the result was considered inconclusive. From further work described in this chapter (see 3.4) and work published by other groups independently[206]these results were in fact consistent with the other evidence.[209]

3.3 Investigation into P450$_{mel}$

With the difficulties of growing and analysing the *in vivo* system, efforts to look for a less complex *in vitro* system were investigated. Several simple cross-linked

peptides exist, for example K-13 or piperazinomycin (see figure 3.10).[210] These at first seem an attractive choice as they are similar to the compounds we originally wished to investigate, and many have already been the subject of published total syntheses. However, disappointingly, we were unable to obtain the genes for the enzymes which catalyse the cross-linking in these cases.

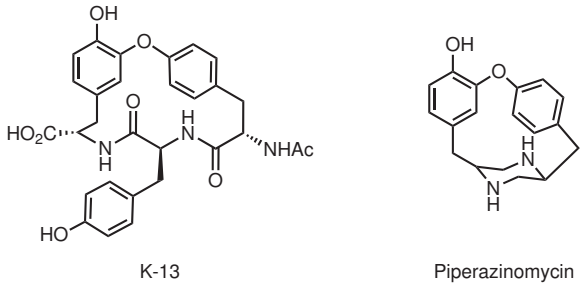


Figure 3.10: K-13 and piperazinomycin

The candidate eventually selected was found in the biosynthesis of hexahydroxyterpylenequinone (HPQ) melanin in *Streptomyces griseus*. Most species within the *Streptomyces* genus produce melanins, high molecular-weight dark brown pigments, by the polymerisation of dihydroxyphenylalanine (DOPA). DOPA is formed by oxidation of tyrosine, and as tyrosine is universally found in living organisms this tyrosinase pathway is the one responsible for melanin production within plants and animals. *Streptomyces griseus* was shown to form an 'albino' phenotype when a mutation occurred within the type III polyketide synthase (PKS) gene *rppA*, which produces 1,3,6,8-tetrahydroxynaphthalene (THN). This suggests that THN is a precursor for a novel type of melanin.[211] Screening of these 'albino' phenotypes is more difficult as *S. griseus* is also able to produce DOPA melanin. Fortunately if the bacterium is cultured on copper-free media no DOPA melanin is produced, as the DOPA melanin biosynthesis involves copper containing mono-oxygenases, and screening for *rppA* mutants is simplified.

The biosynthetic pathway was eventually deduced as summarised in figure 3.11.[211]

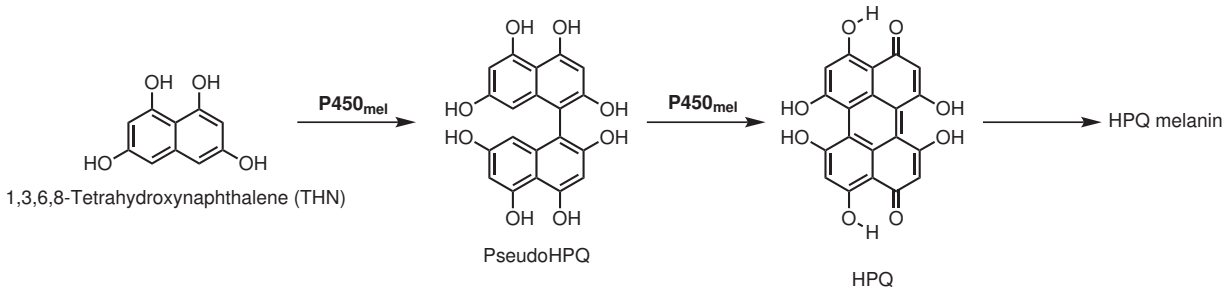


Figure 3.11: The role of P450$_{mel}$ in HPQ melanin biosynthesis

The polyketide synthase, *rppA*, catalyses the production of THN. Two THN molecules then undergo dimerisation with formation of two aryl cross-links between the 4- and 5-positions on each molecule. The product of this dimerisation is HPQ, the melanin precursor. The aryl cross-links within this biosynthesis are both formed by a single enzyme, P450$_{mel}$, and due to the simple nature of the substrate it was thought to be a suitable target for the investigation of the P450 coupling mechanism.

3.3.1 Introduction to polyketide synthases[212]

Polyketide synthases are responsible for producing a vast and diverse selection of natural products in both prokaryotes and eukaryotes. PKSs operate with a similar molecular machinery to that utilised by NRPSs as can be seen in figure 3.12. A notable difference is that while the A-domain of the NRPS can load a wide variety of substrates onto the growing chain, the equivalent acyl transferase domain in the PKS, excepting the first module, has a much more limited pool of substrates to call on. This difference however is compensated by the much more extensive use of tailoring domains compared to the NRPS epimerisation or methylation domains. Each module requires three main domains to catalyse a chain extension; a ketosynthase (KS), an acyltransferase (AT) and an acylcarrier protein (ACP). In addition to this, a

variable set of three main tailoring domains exist in polyketide synthases that carry out keto-modification: dehydratase (DH), enoyl reductase (ER) and ketoreductase (KR).[213, 214] When combined these domains can vary the oxidation level, producing either the ketone, alcohol, alkene or alkane equivalent.

The catalytic cycle of the main three domains parallels that of the NRPS. The first step is the loading of the starter acyl unit onto the first ACP domain. Normally this is an acetyl group from acetyl coenzyme A (AcCoA). From the ACP the acyl group is loaded onto a cysteine thiol of the first KS domain, forming a thioester linkage. A malonate extender unit is then loaded onto the neighbouring ACP of module 1 and undergoes a decarboxylative condensation with the KS loaded acyl chain to extend the chain. This process then repeats to extend the chain, with any tailoring domain present operating after each elongation step.

PKSs are subdivided into 3 classes, labelled Type I, II and III. Type I PKS systems are then further sub-divided into iterative and modular systems. Modular Type I PKSs are similar to the erythromycin system as shown in figure 3.12 and are large multiple domain proteins that contain the entire set of active sites required to produce the polyketide product. Iterative Type I PKSs are similar in this respect, but only contain one catalytic domain for each type of step, and these are reused for each elongation. Any change in oxidation state is thought to be controlled by the length of the growing chain. Type I iterative PKSs are usually found in fungal systems and produce compounds such as aflatoxin and 6-methylsalicylic acid and show the greatest similarity to fatty acid synthases (FAS).

Type II PKSs are iterative in their function and each domain is used for each elongation step, but unlike Type I iterative PKSs each domain is a separate protein. Type II polyketides are generally cyclised and aromatised but do not show extensive reduction or dehydration.

Type III PKSs are again iterative in their function, repeatedly using the same

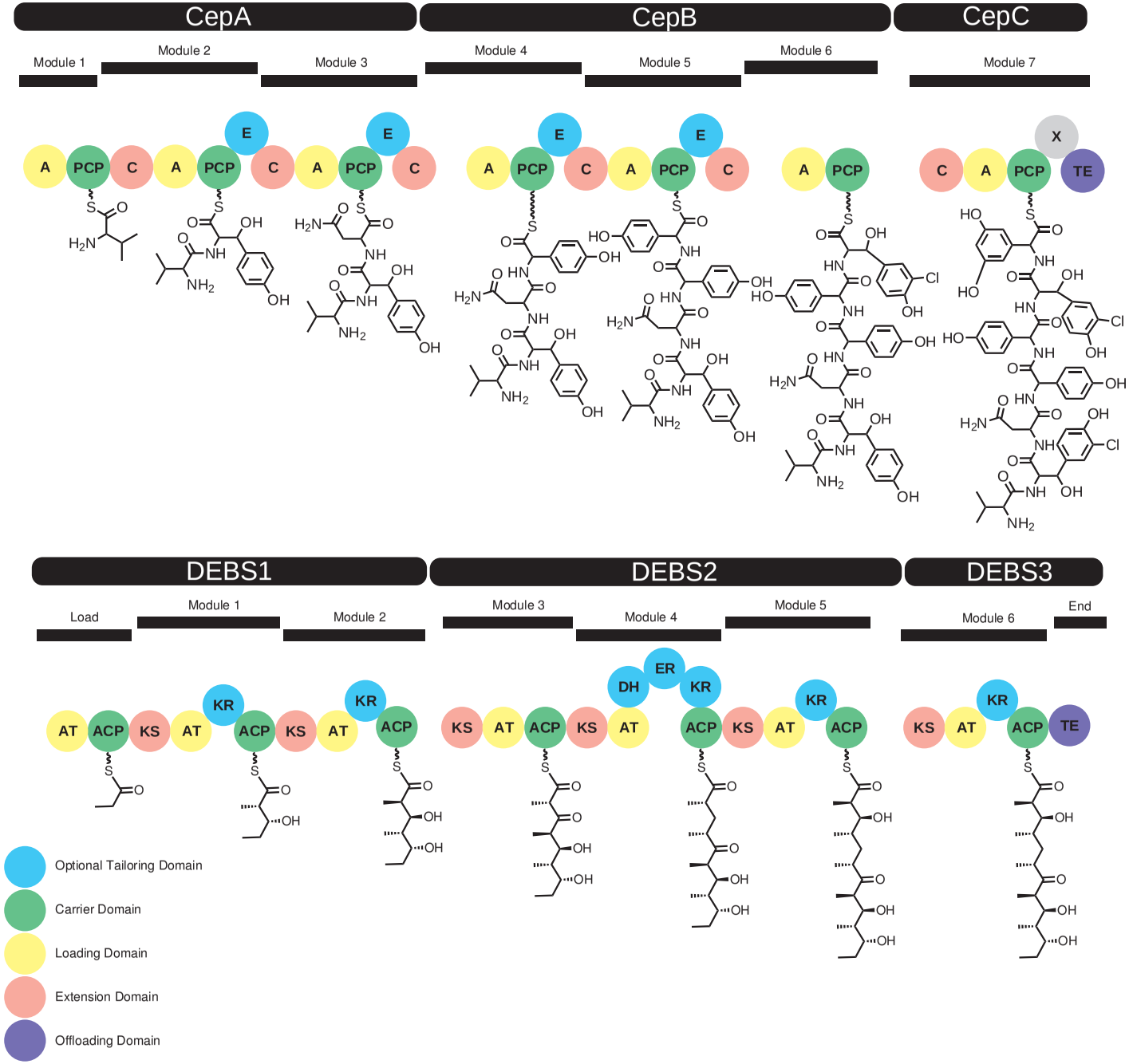


Figure 3.12: Comparison of PKS and NRPSs.

domains, but differ from Type I and Type II PKSs by being able to directly utilise the CoA-bound extender unit without having to off-load it onto an ACP. THN is produced by this class of PKS.

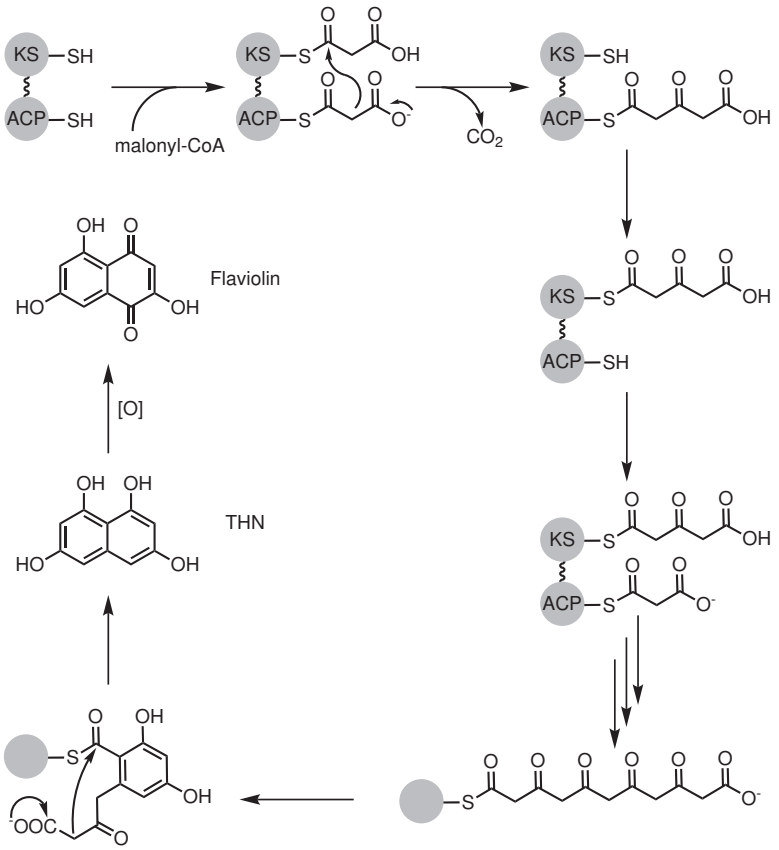


Figure 3.13: The biosynthetic pathway to 1,3,6,8-tetrahydroxynaphthalene (THN). All steps unless otherwise marked are carried out by RppA.

3.3.2 Synthesis of 1,3,6,8-tetrahydroxynaphthalene (THN)

The initial aim was to synthesise the P450$_{mel}$ substrate, THN, and to use it with the expressed protein *in vitro* within an oxygen-18 environment. The product could then be analysed by both mass spectroscopy and carbon-13 NMR, which would each detect an incorporation of the isotopic label. A second requirement was the ability to

label the phenolic oxygen atoms in the substrate molecule, as results from the previous work in this chapter study had suggested, inconclusively, that no molecular oxygen was incorporated by the P450 into the product. In the literature there exists a convenient two-step synthesis from methyl 3,5-dihydroxyphenylacetate.[215, 216] It was reported that methyl 3,5-dihydroxyphenylacetate reacts with acetic anhydride in the presence of boron trifluoride to produce methyl curvulinate (methyl 2-acetyl-3,5-dihydroxyphenylacetate) in 70% yield, presumably *via* an intermediate such as methyl-3-acetoxy-5-hydroxyphenylacetate which undergoes a Fries rearrangement. This method would also offer a way of incorporating oxygen-18 into the final product *via* the use of labelled acetic anhydride. Disappointingly neither the product nor the acetylated intermediate was isolated via this method, despite many attempts and varied conditions. The difficulties in repeating this work have been noted elsewhere in the literature[217] and so an alternative synthesis was followed. The work of Ichinose *et al.*[217] had involved a nine-step synthesis for the production of THN, starting from methyl 3,5-dihydroxybenzoate (see figure 3.14). This method of production had also been used previously to produce carbon-13 labelled compounds and was suitable for the incorporation of oxygen-18. The first 4 steps of this synthesis could be bypassed as instead of starting from 3,5-dihydroxybenzoate, methyl 3,5-dihydroxyphenylacetate, which is commercially available, could provide a shortcut to the protected methyl ester. In attempting to repeat the work the protection of the phenol groups was found to be essential for acetylation in contrast to a possible Fries rearrangement mechanism for the acetylation. The same conditions as those used for the protection of methyl 3,5-dihydroxybenzoate were found to work with methyl 3,5-dihydroxyphenylacetate. All subsequent steps were repeated as undertaken in the original synthesis except for the replacement of acid anhydride and perchloric acid with trifluoroacetic anhydride and acetic acid as it was found to give a more reliable yield. For experimental see section 5.2.2.

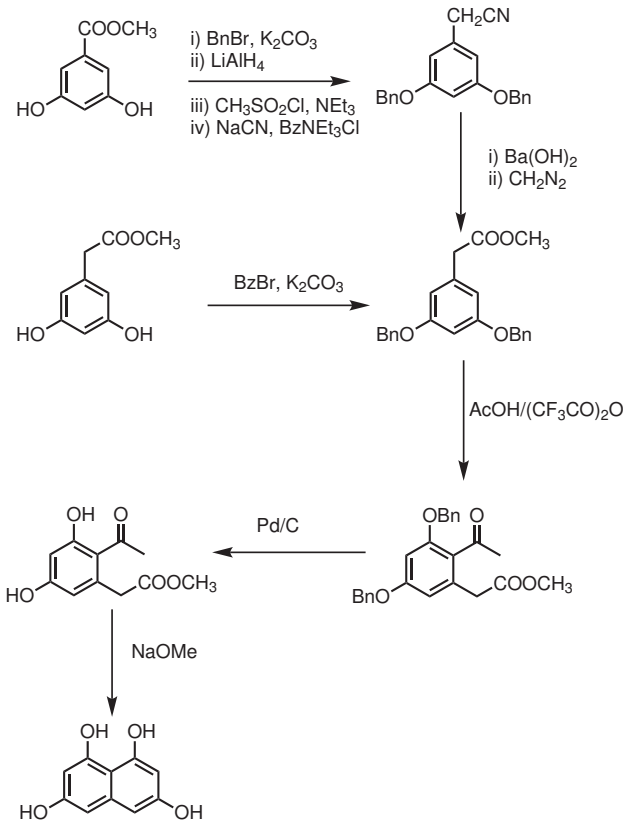


Figure 3.14: Synthetic strategy to obtain 1,3,6,8-tetrahydroxynaphthalene

The gene for the P450$_{mel}$ had previously been cloned into a pET28a vector by Dr. Fanglu Huang. The vector was sucessfully transformed into BL21 competent cells and found to express well on a 1L culture scale after induction with IPTG. Reduction of the purified enzyme with sodium dithionite and subsequent saturation of the buffer with CO showed the appearance of an absorbance at 450 nm, indicative of a functional haem group within the enzyme.

Difficulties arose with the isolation of THN. While the intermediate compounds could be isolated reliably, the oxygen sensitive nature of the final product gave difficulties as it quickly oxidised to form a polymeric dark brown melanin during purification. The use of anaerobic conditions allowed for the isolation of THN, but rapid oxidation gave concerns over the viability of investigating the mechanism with this method. The issue arose that for the catalytic turnover of P450$_{mel}$ an oxygen-containing buffer must be used to provide the molecular oxygen substrate for the enzyme. While the product was stable for 24 hours using argon-flushed solvents, when dissolved in solvents that were previously at equilibrium with oxygen in the air, and not flushed with inert gas, the oxidation reaction occurred rapidly (minutes) to form a dark polymeric compound, thought to be the melanin product. It was therefore going to be a problem that even if a result was obtained, it would be prejudiced by the non-enzymatic oxidation. This was in contrast to the original work from Funa *et al.*[211] who claimed to only see 'insignificant' auto-oxidation in the absence of protein during a 30 minute incubation. No explanation was found for this difference in observed activity.

3.4 Oxygen-18 labelling of the peptide[209]

In parallel to the strategy outlined in section 3.3, it was considered that instead of incorporating oxygen into the mechanism *via* the atmosphere the culture was grown in, it should be possible to synthesise an oxygen-18 labelled tyrosine and feed it to

the growing culture (as had previously been shown to work with carbon-13 labelled tyrosine). A synthetic route to [4-^{18}O]tyrosine was, however, not clear. While it was possible to reduce the aromatic ring *via* a Birch reduction, the resultant product is known to cyclise on addition of acid on the now reduced aromatic ring.[218] To solve this problem the targeted residues were changed. Of the four residues of tyrosine that are incorporated, the two 4-hydroxyphenylglycine (Hpg) residues at positions 4 and 5 (see figure 1.7 of chloroeremomcyin) are involved in all three phenolic coupling steps, whereas the two positions where tyrosine is incorporated as β-hydroxytyrosine are only involved in two of the coupling reactions. Therefore, labelling this Hpg with oxygen-18 still provides a probe for investigating the mechanism of OxyB, OxyA and OxyC. To obtain the appropriately labelled Hpg, we could then take advantage of the dedicated biosynthetic pathway to this amino acid, which goes through 4-hydroxymandelic acid (Hma) (see figure 1.9), which would require less synthetic steps compared to the labelling of tyrosine.

The [^{18}O,^{2}H$_4$]-labelled racemic Hma was synthesised from deuterated aniline *via* ^{18}O-labelled phenol as shown in figure 3.15. For experimental see section 5.2.4. The deuterium label was included so that, after the incorporation experiment, the chloroeremomycin molecules that had incorporated the labelled precursor but lost ^{18}O would be distinguished in the mass spectrum from those derived solely from unlabelled precursors. Hma was produced in an overall yield of 20% from aniline, with an ^{18}O enrichment of 87% and deuterium enrichment in excess of 99% at the appropriate positions.

3.4.1 Incorporation of [^{18}O]4-hydroxymandelic acid

Initially, feeding of [^{2}H$_4$]Hma to a culture of wild-type *A. orientalis* was used to establish the level of incorporation of exogenous labelled substrate into chloroeremomycin. Three chlorination states of the glycopeptide are produced by the strain –

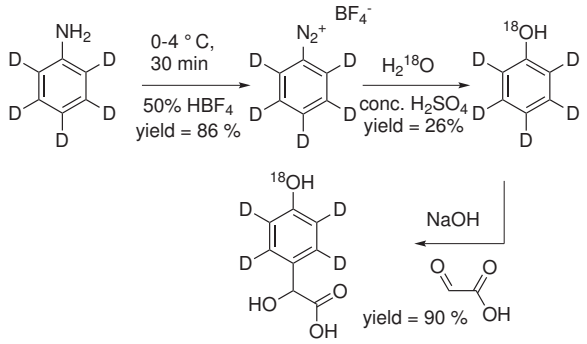


Figure 3.15: [^{2}H $_{4}$, ^{18}O]4-hydroxymandelic acid synthesis.

doubly-chlorinated chloroeremomycin, singly-chlorinated eremomycin, and an unchlorinated product, dechloroeremomycin. For unknown reasons dechloroeremomycin was found to have the highest incorporation of deuterium. A possible explanation of this is that the feeding of labelled 4-hydroxymandelic acid coincided with a reduction in halogenase activity. An alternative hypothesis is that the halogenase is affected by the kinetic isotope effect, as the position of halogenation would require loss of a deuterium. This seems unlikely though as one would expect the rate-determining step to be the addition of chlorine and not the loss of deuterium. Nevertheless the higher incorporation into dechloroeremomycin was beneficial as it simplified the analysis of results, because the unchlorinated product presents a simpler isotope pattern by mass spectrometry than the two chlorinated products. Incorporation of [4-^{2}H$_4$]4-hydroxymandelic acid resulted in three main additional isotopic species compared with natural abundance by mass spectroscopy (see table 3.1): a compound with an M+2 parent ion (retention of two deuterium atoms) from incorporation at the 4th amino acid residue; an M+3 parent ion (retention of three deuterium atoms) from incorporation at the 5th amino acid residue, and an M+5 parent ion from incorporation at both the 4th and 5th positions.

After the successful incorporation of [^{2}H$_4$]Hma, the experiment was repeated using [^{2}H$_4$,^{18}O]Hma. The resulting mass spectral data (see table 3.1) showed mul-

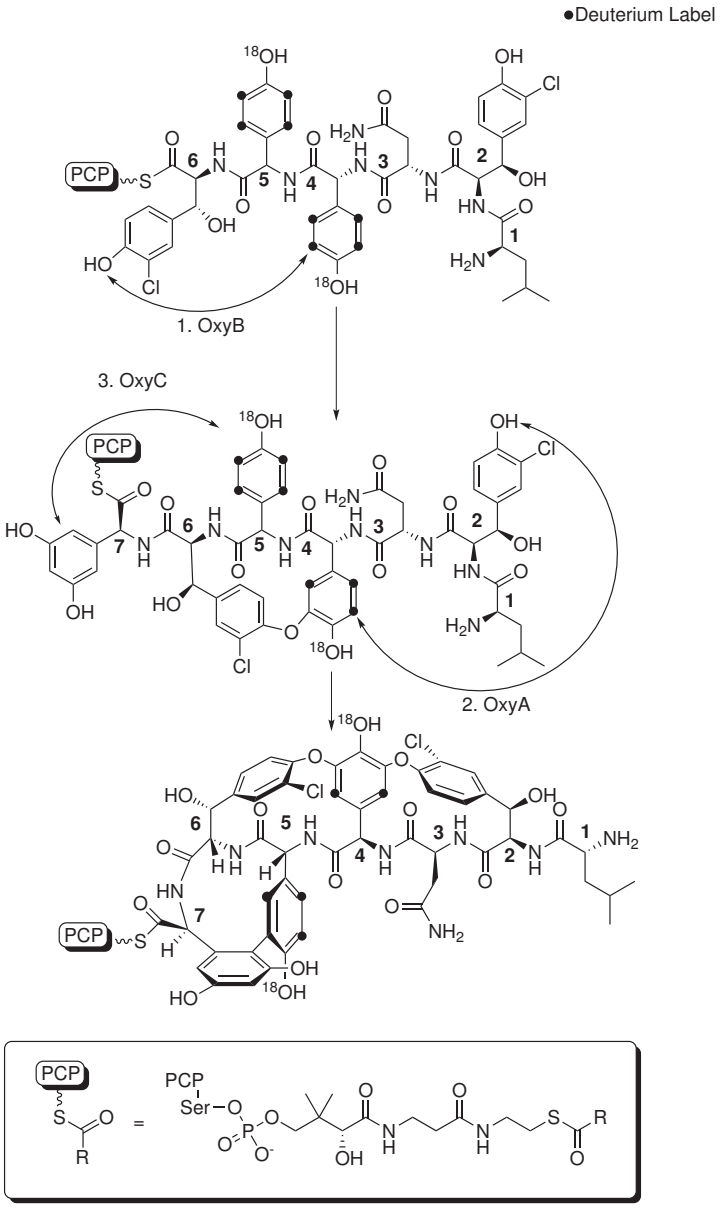


Figure 3.16: Expected labelling from cross-linking in the chloroeremomcyin aglycone assuming no oxygen label is lost. Deuteriated carbon atoms are indicated by dots. Actual timing of the action of the enzymes OxyA and OxyC is currently unknown.

Table 3.1: Mass spectra of dechloroeremomycin after incorporation of labelled Hma.

Mass	Unlabelled	[$^2H_4$]Hma	[$^2H_4$,^{18}O]Hma	Base Peak
+0	100%	100%	100%	Unlabelled
+1	84%	80%	82%	
+2	36%	52%	36%	[$^2H_2$]
+3	11%	44%	12%	[$^2H_3$]
+4	2%	26%	21%	[$^2H_2$,^{18}O]
+5		63%	28%	[$^2H_5$] and [$^2H_3$,^{18}O]
+6		43%	17%	
+7		22%	10%	[$^2H_5$,^{18}O]
+8			8%	
+9			21%	[$^2H_5$,$^{18}O_2$]
+10			14%	
+11			5%	

tiple labelled species due to the two different incorporation sites and the use of [$^2H_4$,^{18}O]Hma that was only 87% enriched in ^{18}O. Even with full retention of the oxygen label, we would expect to see the appearance of eight main species. Two of these species, [$^2H_5$] and [$^2H_3$,^{18}O]dechloroeremomycin, would give the same mass. To dissect the contributions of all 8 species, the standard isotope pattern of dechloroeremomycin was used to predict from the base peak of each species how much it should contribute through natural abundance to each subsequent peak. Starting with the lowest mass species, this could be repeated until all the species had been separated out, with the exception of [$^2H_5$] and [$^2H_3$,^{18}O] dechloroeremomycin, which were treated as a single species, as the ratio of the compounds would not affect the magnitude of subsequent daughter peaks. The final result is shown in table 3.2 and the calculations are provided as supplementary material in table 6.2.

Table 3.2: Relative amounts of species.

Species	Mass Increase	Relative Abundance
Unlabeled	0	100%
[$^2H_2$]	+2	0.5%
[$^2H_3$]	+3	0.6%
[$^2H_2$,^{18}O]	+4	18.3%
[$^2H_5$]	+5	12.1%
[$^2H_3$,^{18}O]	+5	
[$^2H_5$,^{18}O]	+7	5.0%
[$^2H_5$,$^{18}O_2$]	+9	17.5%

In principle, it should be possible to calculate the degree of retention of oxygen-18 by determining the ratio of either the amount of [$^2H_2$]dechloroeremomycin to [$^2H_2$,^{18}O]dechloroeremomycin or that between [$^2H_5$,^{18}O] and [$^2H_5$,$^{18}O_2$]dechloroeremomycin. The former comparison is likely to be less accurate, as the M+2 peak is dominated by the natural isotopic abundance of unlabelled dechloroeremomycin; therefore, the latter ratio was used. The resulting ratio of [$^2H_5$,$^{18}O_2$] to [$^2H_5$,^{18}O] dechloroeremomycin over three separate experiments (3.5:1) was in fact slightly higher than the ratio one would expect on statistical grounds if the ^{18}O enrichment remained unchanged at 87% (which would be 3.35:1), thus indicating essentially total retention of the label. This demonstrates that a negligible amount of the oxygen label was lost during the biosynthesis of the natural product, agreeing with the previous *in vitro* study of OxyB and providing new data for both OxyA and OxyC. The complete retention of the phenolic hydroxyl oxygen suggests that pathways B and C, which go through a gem-diol, are the least likely of the four proposed mechanisms. However, the elimination of water from such an intermediate could be completely stereoselective, as the catalytic iron that donated the hydroxyl group would be in a position to

facilitate its removal. This is not unlikely as there are precedents for stereospecific loss of oxygen from a gem-diol, for example aromatase, also a P450 enzyme, in its second oxidation step (–CH_2OH to –CHO) is proposed to proceed *via* a gem-diol[219] and diol dehydratase (a B12-dependent enzyme).[220] Of the other two mechanisms consistent with the labelling studies, the diradical mechanism A seems more awkward as both radicals have to be generated before coupling takes place, but if the rings are stacked as in the crystal structure it is still possible that the Fe=O abstracts the H from the closer –OH and then that –O· radical abstracts the H from the further –OH. The single radical mechanism D is more appealing, but would still require both phenolic hydroxyl groups to be positioned close to the iron centre.

In conclusion, the evidence is that oxygen labels on both the Hpg residues of the nascent peptide chain are retained during the biosynthesis of dechloroeremomycin. These results suggest that all three oxidative cross-linking enzymes react via a similar mechanism; this is of particular interest in the case of OxyC as it catalyses a carbon-carbon coupling rather than a carbon-oxygen coupling as catalysed by OxyA and OxyB (figure 3.16). Finally, these results suggest that the continuing study to elucidate the mechanisms of all 3 enzymes remains important if we are to fully understand the biosynthesis of this important class of antibiotic.

3.5 Future work

While the results of this study have helped to clarify the mechanism of phenolic coupling enzymes, the scope for future investigations of P450 coupling enzymes is still wide, with the complete picture of the mechanism of phenolic coupling reactions still to be resolved. It would be interesting to investigate, assuming a solution to the problems of auto-oxidation can be found, whether $P450_{mel}$ (section 3.3) catalyses reactions by a similar mechanism to that of OxyA-C. Despite their very different substrates, $P450_{mel}$ and OxyC share a 32% identity in amino acid sequence. It has

been previously reported that it is possible to monitor the turnover of $P450_{mel}$ and it was possible to synthesise the substrate in a manner that would allow for labelling of the phenol groups. A second strategy would be to investigate the catalysis of OxyB *in vivo* with altered PCP-bound peptides to see if coupling reactions are possible in the presence of facile leaving groups, such as halogens, in place of the hydroxyl group. This would prevent the generation of a radical by removal of the phenolic hydrogen, but would allow for the reaction to proceed *via* the generation of an epoxide. This strategy requires the synthesis of a modified PCP-bound pentapeptide, a task which is already made considerably more difficult by the number of amino acids that are not commercially available within the peptide chain. An alternative to the synthesis of several complex peptides would be an *in vivo* method by mutation of the NRPS so that it loads an alternative amino acid in this position, something that is possible by such techniques as module swapping or protein engineering. A more attractive possibility though would be to mutate the active-site serine within the PCP of a glycopeptide NRPS to inactivate it; it should then be possible to provide an SNAC-bound amino acid to substitute for the PCP amino acid donor as a substrate for the condensation domain as has previously be shown to work in CDA.[105] It was these ideas that partially provided the driving force towards the work in chapter 4 with the hope that it could eventually lead to the possibility of testing some of these hypotheses, while in the short-term providing us with a powerful tool for also investigating NRPSs and developing new 'non-natural' products.

4

High-Throughput Screening of Non-Ribosomal Peptide Synthetases

4.1 Introduction

Bacterial non-ribosomal peptide synthetases (NRPSs) produce a large range of natural products. Many of these secondary metabolites are already used clinically as antibiotics including drugs of last resort such as vancomycin. With the continuing need to find new antibiotics in the battle against antibiotic resistance, an enhanced ability to exploit NRPSs would present an interesting opportunity for the discovery of novel compounds.

Three types of catalytic domains on the NRPS are required for peptide synthesis:[221] adenylation domains, peptidyl carrier proteins and condensation domains. The adenylation domains or A-domains can be considered the gatekeepers of the system; they select the amino acid (or another starter or extension molecule) to be incorporated at each step of the biosynthesis and create the amino acyl adenylate (or equivalent).[222, 223] It is this recognition step that is the first hurdle for the inclusion of an amino acid into the peptide chain. The role of an A-domain is two-fold: the selection and activation of the amino acid substrate, followed by the transfer of the activated aminoacyl group to the downstream PCP. The PCP transports the activated amino acid to the

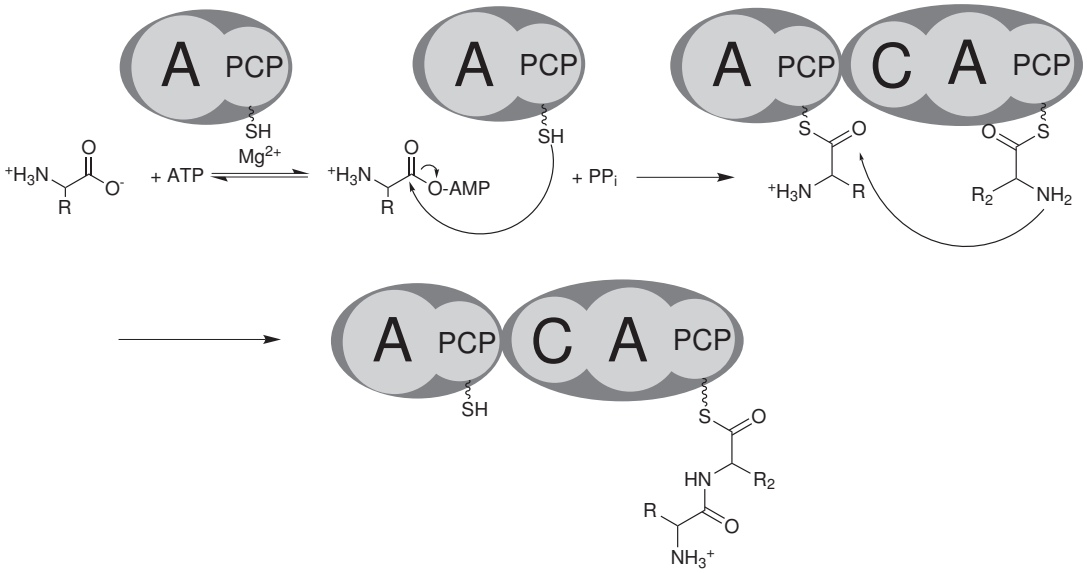


Figure 4.1: Example of a modular NRPS system with the adenlyation domains marked A; condensation domain marked C; and the peptidyl carrier proteins labelled PCP.

condensation-domain or C-domain which catalyses the coupling of the amino acid derivative to the growing peptidyl chain (see figure 4.1). For a more detailed explanation of NRPSs see section 1.5.

The modular nature of NPRS systems provides an attractive basis for the development of combinatorial biosynthetic methods. Efforts have been made to swap modules in the hope of producing non-natural variants of the original NRPS products with novel biological properties. Although this goal has been achieved, it is not without great difficulty and careful choice of target.[36, 35] The resultant chimeric NRPS is often crippled and produces a low yield, if at all.[224, 225, 226] Some success in improving the activity of these chimeric assembly lines has been achieved by directed evolution;[227] however, little can be learned from the sites of the mutations within the enzymes. The purpose of the directed evolution experiment was to increase the yield rather than to change the specificity of the A-domain, and the mutations that had this

effect were mostly located at the external surface of the domain. This implies that the result is brought about by improvements in the docking interactions within the system.

An alternative strategy to swapping the A-domain with that from another NRPS would be to undertake rational redesign upon the A-domain to alter the substrate specificity. This method would maintain the superstructure of the NRPS, and therefore not interfere with the interactions between domains. The large amount of data on A-domains obtained from sequencing and crystal structures enabled a list to be generated of residues that generally occur at eight specific positions in A-domains that have a particular substrate specificity. This allowed the prediction of the substrate specificity of A-domains whose substrate had not been determined.[30, 29] Despite some initial success using this 'NRPS code' to make substrate changes with little penalty to the kinetics of the domain,[228] it quickly became apparent that it was not as simple as had been hoped. Altering residues to give an altered substrate specificity in line with the predictions has given very few examples of success; of these, many display orders of magnitude slower kinetics in comparison to the wild-type.[228, 229] The validity of this specificity-conferring code has also been hampered with examples of A-domains where the NRPS code does not predict the substrate correctly.[230] More recently, Li *et al.* demonstrated the direct activation of D-Ala during fusaricidin biosynthesis by the Fus-A6 domain of *Paenibacullus polymyxa* PKB1, for which no substrate could be predicted, further demonstrating the limitations of the current model.[40] With these results, it is becoming evident that, while a small set of eight conserved residues are able to help to predict the specificity of the domain, there are other residues of importance in dictating the substrate activated by a specific A-domain.

Previously, directed evolution studies have been very successful at identifying residues responsible for catalysis and substrate binding, such as the work in chap-

ter 2. Directed evolution has also been used within the combinatorial biosynthesis field for the identification of key residues in ketoreductase domains[231] and interactions between NRPS modules.[232, 233] The most crucial part of a directed evolution study is the throughput that can be provided by the screening process. The assay has also to be able to detect very low levels of enzymatic activity that might occur from mutagenesis. The traditional method for the screening of the substrate specificity of A-domains is the use of a discontinuous adenosine triphosphate/radio-labelled pyrophosphate exchange (ATP/$^{32}PP_i$) assay.[234] The reliability and robustness of this assay is demonstrated by a 30 year legacy in the determination of activity of NRPS A-domains and the activity of tRNA synthetases. These assays rely on the equilibrium nature of the adenylation reaction. The amino acid to be screened, ATP and the A-domain are incubated together with an excess of radio-labelled $^{32}PP_i$ and the mixture is allowed to equilibrate. Since the $^{32}PP_i$ is in excess over the unlabelled PP_i produced by the enzyme, the reverse reaction allows for the build up of phosphorus-32 labelled ATP. This can then be extracted *via* the use of charcoal and the levels of ATP quantified *via* the detection of the radioactive decay of the labelled ATP. This process is both highly time-consuming and has all the difficulties arising from handling radioactive material.

Interest in developing new techniques for the screening of A-domains increased recently, with several new assays developed to screen A-domains in a high-throughput manner, and over half of these methods have been published during the last twelve months.[235, 236, 237, 238, 239]

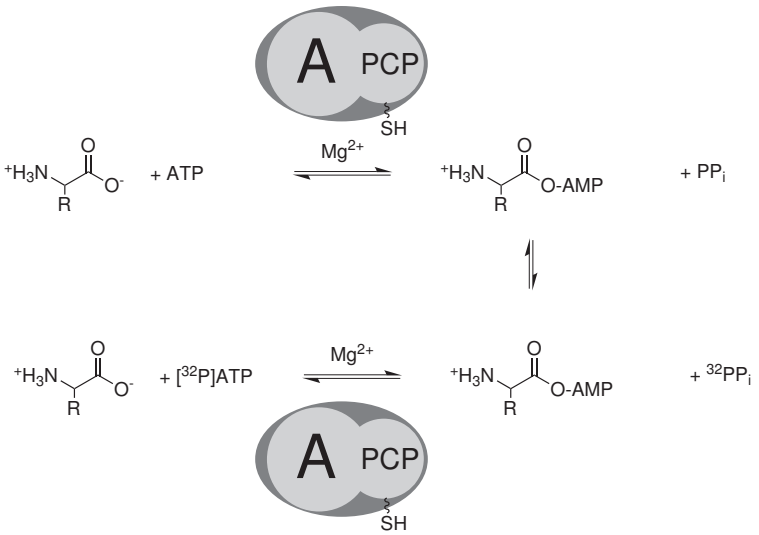


Figure 4.2: Overview of typical ATP/$^{32}PP_i$ assay. The PCP may or may not be present depending on the enzyme being assayed.

4.2 Existing high-throughput screening methods

4.2.1 Fourier-transform mass spectroscopy

The high resolving power of Fourier-Transform Mass Spectroscopy (FTMS) provides an alternative probe for the *in vitro* characterisation of NRPS (and PKS) substrate specificity.[235] The aim of developing such an assay was to avoid the use of radio-labels and to allow the parallel screening of multiple substrates. Secondly, as the assay is carried out on fragments of the native protein this allows the timing of tailoring events to be monitored.

The general method is that over-expressed protein containing a carrier protein is purified and converted to the *holo*-form by incubation with Sfp, a phosphopantetheinyl transferase from *B. subtilis* (see section 1.6.2),[240] and CoA to form the active PCP. This is then combined with an A-domain (if this is not already part of the same protein as the PCP) and incubated with a substrate pool, for example a selection of amino acids. Following an incubation period, the protein is digested to release

the PCP from the rest of the protein (if required), and purified by HPLC prior to ESI-FTMS analysis (see figure 4.3).

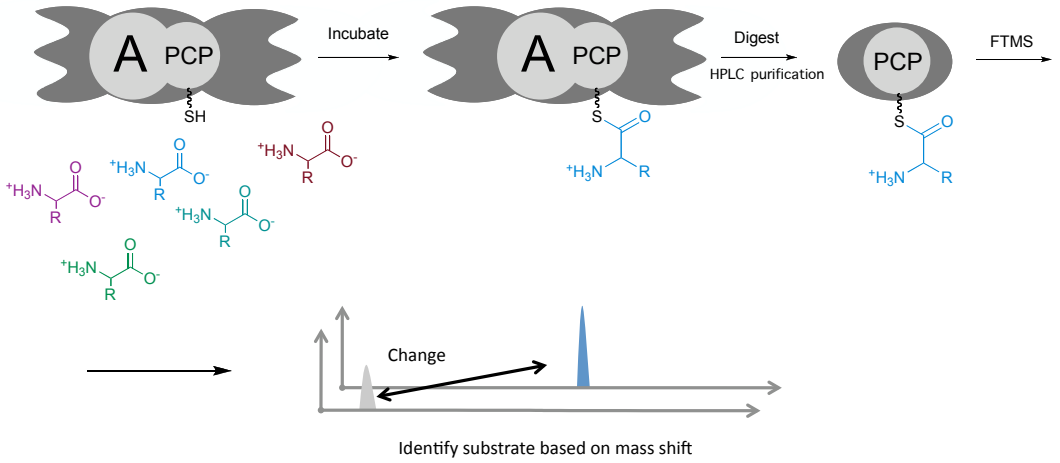


Figure 4.3: General approach of using FTMS to screen for acylation of a carrier domain using multiple substrates

This method has many practical uses; the first is for identification of a substrate from a substrate pool. Negligible loading of all but the natural substrate (if present) was observed in the proof of concept experiments that were carried out. The only unexplained result of these initial studies was the loading of a +199 Da species by the EntE A-domain onto the EntB PCP domain in the absence of 2,3-dihydroxybenzoic acid (its natural substrate). This matched none of the amino acids tested as substrates. It was assumed that this was due to the 1–2% impurities found within the commercially prepared amino acids used. In support of this hypothesis, there was insufficient of this unknown substrate to allow for complete acylation of the carrier protein, whereas in presence of the natural substrate the PCP was fully loaded (see figure 4.4).

The same experiment in the case of other A-domain/PCP combined systems allowed for the identification of previously unknown activities in the absence of known

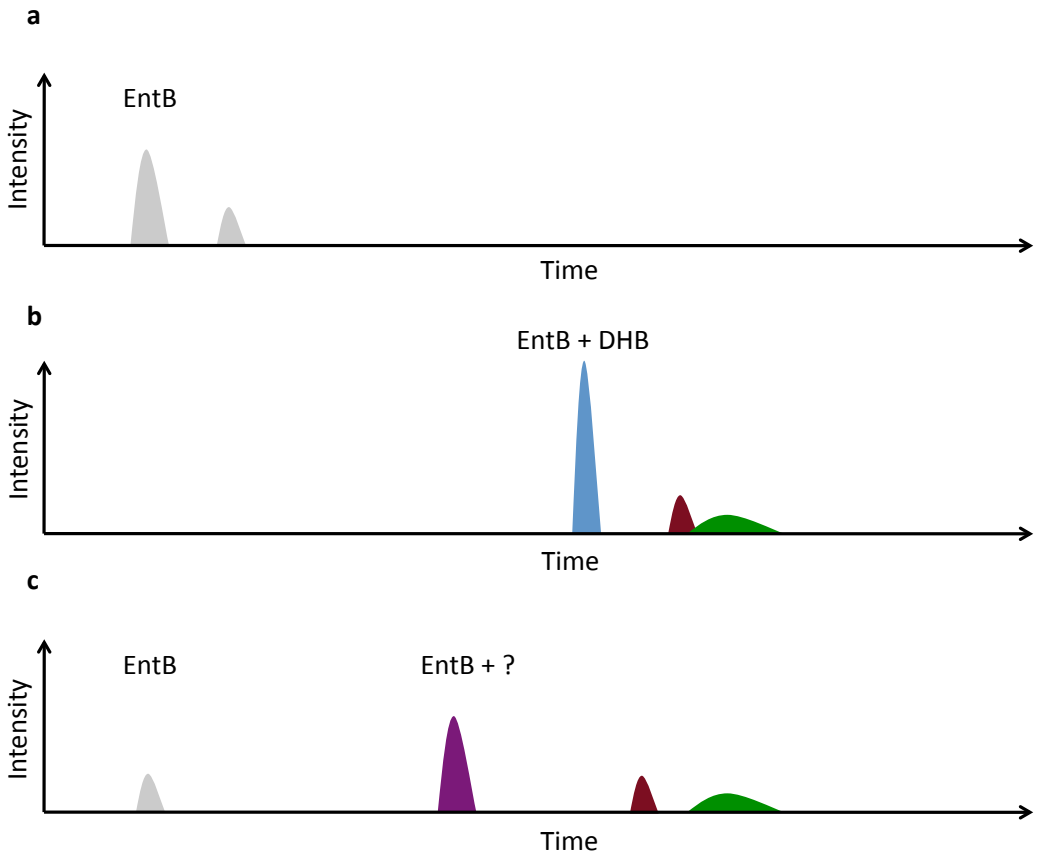


Figure 4.4: Illustrations of published LC-FTMS results. The total ion current against time from the analysis of several reaction mixtures are shown: (a) Holo-EntB; (b) Holo-EntB incubated with ATP and all 19 proteinogenic L-amino-acids, glycine, L-selenocysteine, L-cystine, 4-trans-hydroxy-L-proline and 2,3-dihydroxybenzoic acid; (c) identical to (b) but 2,3-dihydydroxybenzoic acid omitted.

substrates by characterisation of the mass increase. This would allow for the identification of minimal activities that could then be nurtured to create new substrate specificity within the enzyme. The example put forward within the study was that replacing the pyrrole on chlorobiocin or coumermycin may allow for the development of more soluble aminocoumarin antibiotics.[241] To investigate if other amino acids would load in place of proline onto CloN5 or CouN5, a substrate screen was carried out in the absence of proline including the remaining 18 proteinogenic L-amino-acids, glycine, L-selenocysteine, L-cystine, and 4-trans-hydroxy-L-proline. The resulting mass shift was identified as the loading of 4-trans-hydroxy-L-proline. This was then confirmed by the traditional ATP/$^{32}PP_i$ exchange assay, suggesting that 4-trans-hydroxy-L-proline would make a good candidate for developing novel chlorobiocin and coumermycin antibiotics.

As expected, the use of FTMS also allows for the investigation of other functions involved in the maturation of NRPS products such as the tailoring steps. To identify the timing of bromination in the jamaicamide biosynthetic pathway, the PCP (holo-JamC) and A-domain (JamA) were incubated with 6-bromo-5-hexynoic acid, 5-hexenoic acid and ATP. The results showed only the loading of the hexenoic acid, and no loading of the brominated substrate even in the absence of the hexenoic acid. This confirmed that bromination has to occur after loading onto JamC. JamA was also found to load 5-hexanoic acid and 5-hexynoic acid in agreement with the classical assay.[242]

A further use of FTMS in NRPS characterisation is the investigation of orphan or cryptic gene clusters as it allows assays with a wide range of substrates simultaneously. The initial example used in the study was an orphan NRPS/PKS gene cluster from *Bacillus subtilis*.[243, 244] Even though the cluster was known to contain NRPS modules, it was proposed that the cluster produced difficidin, which does not contain a single amino acid, and there is little experimental evidence to support the

suggestion that the cluster is responsible for the production of difficidin. To confirm that the two NRPS didomains of the cluster (PksJ and PksN) were in fact active, they were investigated by the FTMS assay developed in the study. In summary, PksJ and PksN were found to load glycine and alanine respectively. In the case of PksN this was in contrast to the predicted cysteine or serine. It was noted that in the absence of L-alanine, L-serine was loaded, however as alanine out-competed the predicted substrate the study concluded that alanine is the natural substrate. This led to the important realisation that both NRPS didomains were active, which suggests that, as difficidin contains no amino acids or nitrogen atoms, the gene cluster is not responsible for its production. It was noted that they could not exclude the possibility that the cluster produced a modified form of difficidin or analogous natural product.

4.2.2 Optimised ATP/$^{32}PP_i$ exchange assay in 96 well format

Based on the long-running success of the ATP/$^{32}PP_i$ assay (see figure 4.2), Otten *et al.* looked to modify the assay with the aim of increasing its throughput and applying it to a 96-well format.[236] The objective of the work was to reduce the final volume of the assay, the volume after quenching of the enzymatic reaction and addition of scintillation fluid, from 4 mL to a volume suitable for analysis in a 96-well format (200 μL).[245] The activity could then be monitored conveniently by the use of 96-well liquid scintillation counter. Two possible methods for the determination of activity were proposed, the first of which was a precise miniaturisation of the original ATP/$^{32}PP_i$ exchange assay, with the amount of radio-labelled-ATP retained on the charcoal measured. The second procedure developed was the counting of the flowthrough after filtering off the charcoal. This should contain a reduced level of radioactivity if the exchange reaction had occurred.

The study developed both assay formats into a viable method of high-throughput screening. The measurement of the radio-labelled ATP bound to the charcoal was

found to be 10–100 fold more sensitive that the measurement of the eluant: however, the measurement of bound ATP had the disadvantage that it was more time consuming. Compared with the traditional method, the assay also involved much lower volumes and fewer pipetting steps, thus minimising the handling of radioactive material. The charcoal counting method was able to measure activity levels over three orders of magnitude, with the data generated from a 0.1 nM concentration of enzyme showing only slight scatter. This detection of very low levels of activity is very desirable in directed evolution studies where activities of alternative substrates might be very low.

The scaled down procedure was also capable of producing reliable kinetic data. Due to the discontinuous nature of the assay and the extra handling that would be required if the reaction in each well had to be stopped at different times, the measurements were recorded at a single time point using wells with different substrate concentrations. The kinetic data produced was an excellent fit to the expected Michaelis-Menten equation with errors below 5% for the calculation of k_{cat} for the enzymes and mutants used in the study.

To make use of the miniaturisation of the ATP/$^{32}PP_i$-exchange assay a method was required for the high throughput isolation of protein. Ideally, the crude cell lysate could be used from 1 mL cultures of *E. coli* containing the different mutant genes for the protein. However no ATP/$^{32}PP_i$-exchange was detected by this method. The explanation given for this was that ATPase activity in the cell lysate was removing the generated ATP.[246] Purification of the protein was hence necessary, and carried out in TALON 96-well plates. Elution from the plate led to 200 μL enzyme at 1 μM, enough for 20 ATP/$^{32}PP_i$-exchange assays. The purity of eluted protein was confirmed by SDS-page and Coomassie staining. As the concentration of the protein would affect the detected rate it was necessary that the production level of the enzyme was consistent across the plate. A high-throughput assay was undertaken with a

large number of known proteins, that were purified and then measured by ATP/$^{32}PP_i$-exchange assay. The detected activities of the eluted proteins did indeed confirm that the level of protein production was reasonably uniform across the plate.

It is reported that realistically it should be possible to screen 10 plates a week *via* this method (1000 mutants) and that by using mixtures of substrates, a combinatorial screen is possible to increase the number of substrates tested with each plate. An overview of the entire process is shown in figure 4.5.

4.2.3 Phage display of A-domains

Phage display takes advantage of genetic modification to the gene encoding part of the outer protein shell of the phage so that a desired protein is expressed on the surface. This establishes a direct linkage between the active protein and its encoding gene, which is within the same phage particle.[247] The connection between the enzyme and gene has provided a general platform for screening large libraries of proteins for desired binding interactions[248, 249] or catalytic activities.[250, 251, 252] Zou *et al.*[237] developed a screen based on this technology for the screening of NRPS A-domains. To achieve this, they displayed A-domain/PCP didomains on the surface of the phage. To screen for activity, it was necessary to develop an affinity tag to selectively detect and isolate the active phage particles. The solution was to use substrate analogues that contained an azide- or alkyne-functionalised tether. The A-domain would then process the substrate analogues as the normal substrate and load this onto the neighbouring PCP. The tether was then available for the attachment of an affinity tag by 'click' chemistry.[253, 254] This allowed for the direct coupling of the gene encoding the active mutants to a biotin affinity tag, which could then be selected by the use of streptavidin affinity (see figure 4.6).

To test the process, the N-terminal didomain RifA from the rifamycin synthetase of *Amycolatopsis mediterranei* was displayed onto the surface of an M13 phage

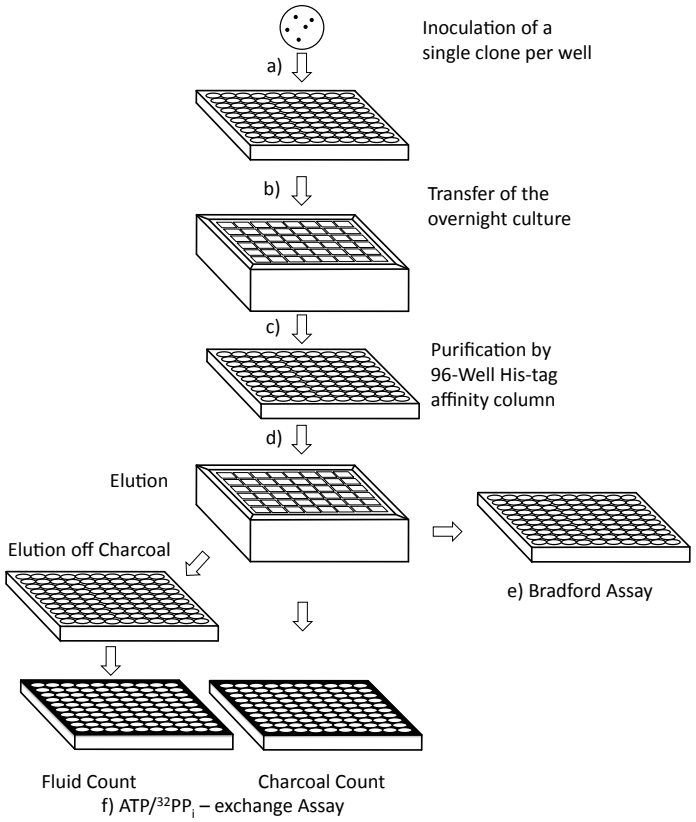


Figure 4.5: Overview of Otten *et al.* A-domain protein production and screening method – adapted from original paper: a) Single clones of *E. coli* (containing plasmid of protein expression) are inoculated into growth medium. b) An aliquot of the overnight cultures are transferred in fresh growth medium and cells are grown for 2 hours before protein expression is induced by IPTG. After 24 hours, cells are lysed. c) After centrifugation, the cleared lysate is transferred on to a TALON plate containing a His-tag affinity resin. d) The His-tagged protein is eluted using an imidazole containing buffer. e) Protein concentration can be evaluated by Bradford assay. f) The purified protein is then assayed by ATP/$^{32}PP_i$. For charcoal counting the assay is performed in a Black and White (B&W) plate, after which scintillation fluid is added to the washed charcoal that contains the radioactive ATP to be monitored. For fluid counting the samples are centrifuged in a conical plate and the fluid containing the remaining radioactive PP_i is eluted into a B&W plate containing scintillation fluid. Radioactivity is then measured with a 96 channel scintillation counter.

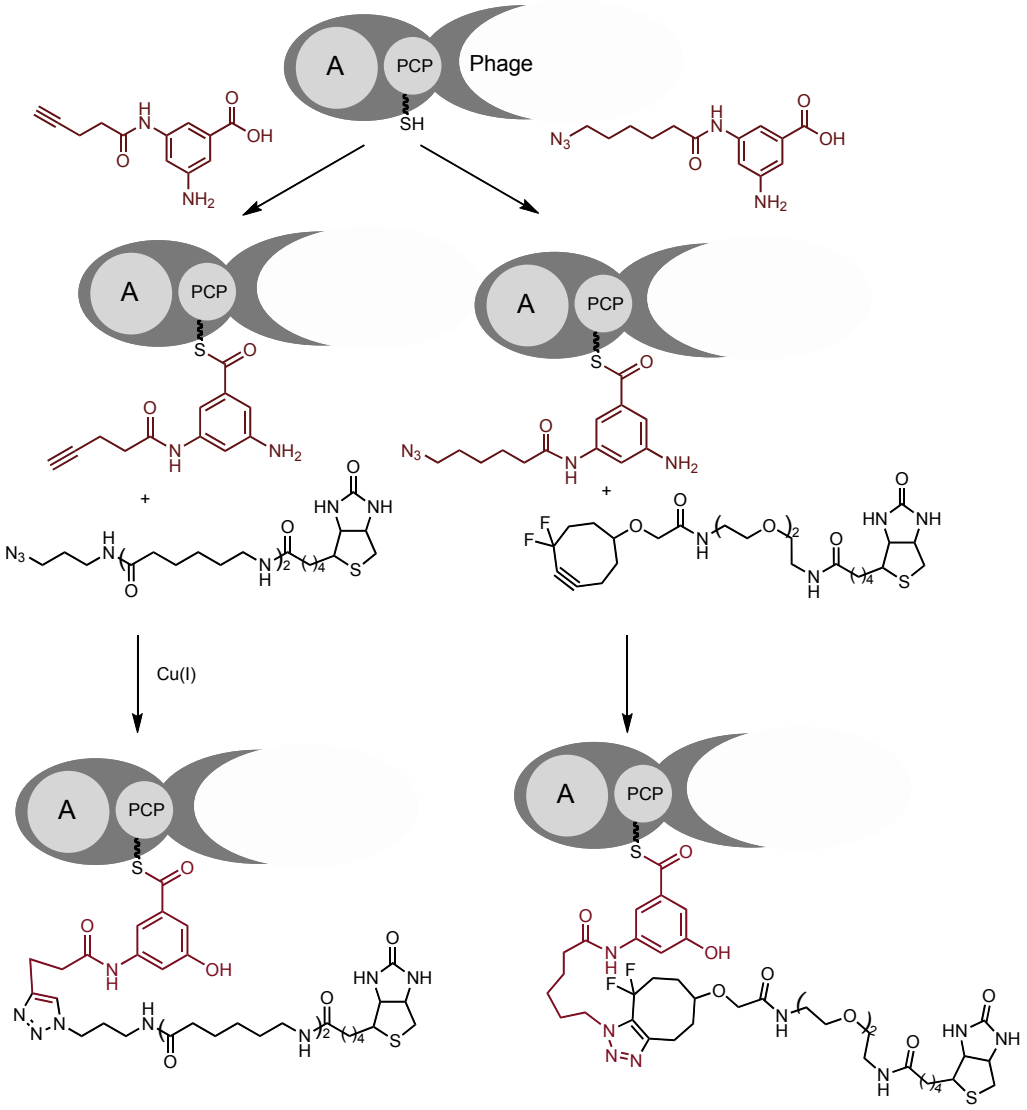


Figure 4.6: Phage selection is carried out by the loading of a alkyne/azide containing substrate onto the phage bound PCP by the A-domain. This can then be coupled to a biotin tag using 'click' chemistry. The biotin affinity tag can then be used for enrichment of the phages displaying A-domain mutants that load the tested substrate.

by fusion with phage capsid protein pIII. To assay the display efficiency of the A-domain/PCP on the surface of the phage, Sfp was used to load a biotin-CoA onto the PCP. This provided the PCP with a biotin-Ppant arm. A Western blot of the reaction mixture was probed with streptavidin-horseradish peroxidase (HRP) conjugate, and the appearance of a band at 84 kDa confirmed that the A-domain/PCP didomain from RifA was displayed on the phages with good efficiency.

To confirm that the A-domain would be active with the alkyne and azide functionalised substrate analogues, the activities were assayed by ATP/$^{32}PP_i$ exchange. The activity assays confirmed that both substrates were successfully activated by the A-domain, although they exhibited, as expected, lower activity than the natural substrate 3-amino-5-hydroxybenzoate (Ahb). To confirm that the alkyne substrate could be loaded onto the PCP it was incubated with the M18 phages displaying the catalytic didomain. After the substrate was loaded the phage particles were precipitated by polyethylene glycol before coupling to the biotin-azide in the presence of CuBr. Western blot was then used to showed the the presence of the newly added biotin label to the A-domain/PCP didomain-phage capsid protein pIII fusion. The control reaction with no substrate was found to give no biotin-labelled protein as expected. Unfortunately, the phage infectivity was decreased substantially by the exposure to Cu^+ ions (10 μM for 30 min). This decrease in infectivity (100-fold) resulted in the phages being unable to infect the *E. coli* cells for future rounds of selection, thus making the use of Cu^+ catalysis in the presence of the phage not viable.

To avoid inactivation of the phages, the substrate was changed from an alkyne- to an azide-containing substrate as the cycloaddition reaction between an azide and biotin-DIFO does not require copper catalysis.[255] Western blot was again used to confirm the biotin attachment and dependancy on the substrate analogue; the isolated phages also showed no loss in infectivity. This method showed a 10-fold increase in phage enrichment compared to the control screen of phages expressing only a PCP

with no active A-domain. In addition phages displaying the PCP but without a Ppant arm, were found to give no phage enrichment over the control phage reactions.

4.2.4 Substrate fluorescence polarisation[238]

Many pathogenic organisms produce siderophores to acquire iron from the host organism and disruption of this production is a possible route to new antibiotic compounds. Siderophores are synthesised by two operationally different mechanisms; the better understood of these two mechanisms is NRPS-catalysed biosynthesis, and the A-domain of these synthetases provides a possible target for an inhibitor. To develop a high-throughput screen with A-domains in search of possible new inhibitors, Neres *et al.* successfully designed and synthesised a fluorescence polarisation probe, Fl-Sal-AMS. The design of Fl-Sal-AMS was based on the Sal-AMS and DHB-AMS inhibitors, in which the labile acyl phosphate linkage of the acyl adenylate intermediate is replaced by a bioisosteric and chemically stable acyl sulphate (see figure 4.7).

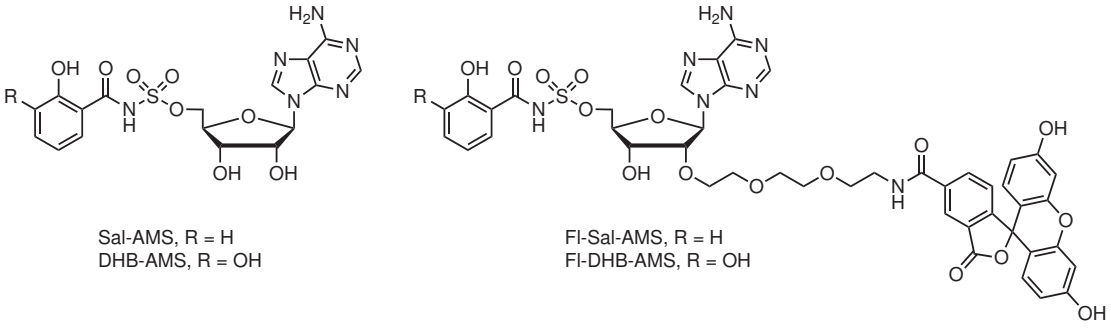


Figure 4.7: Chemical structures of Sal-AMS, 2,3DHB-AMS and Fl-Sal-AMS.

The monitoring of substrate binding by fluorescence polarisation is achieved by exciting a molecule with polarised light, the intensities of the fluorescence that is polarised parallel and perpendicular to the polarisation of the excitation light are then measured to give the fluorescence anisotropy (see equation 4.1. A is the measure

of anisotropy, G is the instrument specific G-factor, $I^{\parallel}$ and $I^{\perp}$ are the intensities measured with parallel and perpendicular polarisation respectively).

$$A = \frac{I^{\parallel} - GI^{\perp}}{I^{\parallel} + GI^{\perp}} \tag{4.1}$$

The fluorescence anisotropy is dependent on the rate of tumbling in solution; therefore as Fl-Sal-AMS binds to the protein the rate of tumbling will decrease and the anisotropy will increase. From the anisotropy at various concentrations a binding constant can be calculated by non-linear curve fit to equation 4.2. In equation 4.2, A_{OBS} is the experimentally measured anisotropy, Q is the ratio of the fluorescence intensity of the probe in the bound and free states, F_{SB} is the fraction of bound substrate, A_B and A_F represent the anisotropies of the bound and free state respectively. In equation 4.3, K_{D1} is the equilibrium dissociation constant of the compound measured, L_{ST} is the concentration of the substrate and R_T is the concentration of the enzyme.

$$A_{OBS} = \frac{QF_{SB}A_B + (1 - F_{SB})A_F}{1 - (1 - Q)F_{SB}} \tag{4.2}$$

$$F_{SB} = \frac{K_{D1} + L_{ST} + R_T - \sqrt{(K_{D1} + L_{ST} + R_T)^2 - 4L_{ST}R_T}}{2L_{ST}} \tag{4.3}$$

The initial results demonstrated that the Fl-Sal-AMS fluorescence probe worked successfully with a selection of salicylic and 2,3-DHB activating A-domains (MbtA, YbtE, EntE, VibE, BasE and DhbE). The fluorescence was measured after incubation with the enzyme for 30 minutes (although the signal was found to still be reliable 24 hours later), and K_D calculated by serial dilutions of substrate from 200 nM to 20 nM. These results were later confirmed by isothermal titration calorimetry (ITC).

As the original aim of the study was to screen for novel inhibitors, the process was scaled down into 384-well format for the use of competitive displacement assays (calculations explained in the original paper)[238] on BasE, the A-domain from *A.*

baumannii responsible for the activation of 2,3-dihydroxybenzoic acid in the acinetobactin biosynthetic pathway. A total 85 000 compounds were screened in a search for novel inhibitors and several hits for BasE inhibitors with low K_D values were found.

While the use of fluorescence probes within this work was not to search for alternative possible substrates of A-domains, the original authors note that it is able to probe for substrate specificity. It was shown that displacement experiments using their natural substrates, such as salicylate, 2-3-dihydroxybenzoate and AMP, were able to give K_D constants for all six of the A-domains investigated in this study. It was also noted that Fl-Sal-AMS was ineffective at binding to AsbC, which uses 3,4-dihydroxybenzoic acid.

4.2.5 Nonradioactive high-throughput assay

As the traditional method of analysing A-domains required the handling of radioactive material, McQuade *et al.* sought to develop an assay that could be used in facilities that were not equipped to handle such material.[239] Instead, inorganic pyrophosphatase was used to convert PP_i produced during the activation of the amino acid substrate of the A-domain to orthophosphate (P_i). The concentration of P_i was then measured colorimetrically by the use of a molybdate/malachite green reagent added after the desired timeframe, stopping the reaction. After 15 minutes to allow the colour to develop, the absorbance of the malachite green was quantified at 600 nm.[256] This technique results in the measurement of a different parameter to that of the traditional ATP/$^{32}PP_i$ exchange assay, which measures the reversible phosphodiester bond cleavage, whereas the molybdate detects the release of PP_i. Of more concern, however, is the substantially reduced k_{cat} observed in the release of PP*i* compared to that measured by ATP/$^{32}PP_i$ assay. This is rationalised as the release of the amino acyl adenylate product being rate limiting. However, the parameter mea-

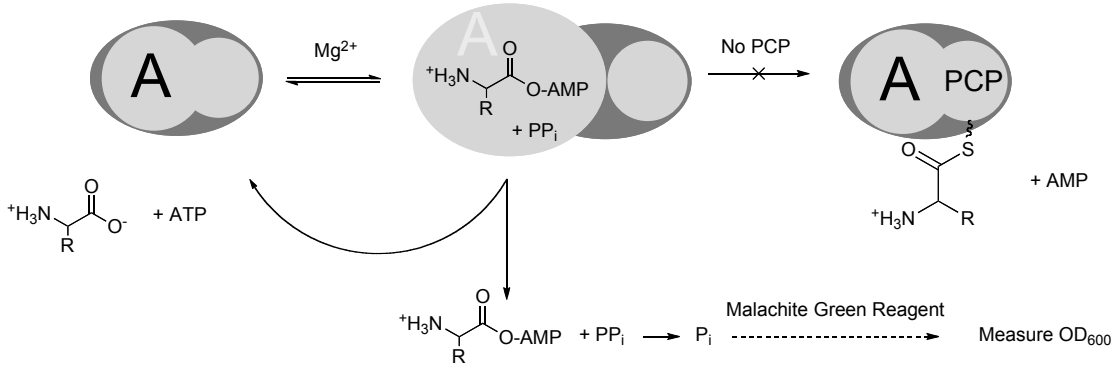


Figure 4.8: Reaction schematic for the detection of enzyme kinetics by malachite green.

sured is the steady-state release of PP_i in the assay, and it has previously been shown that turnover rates are up to 200-fold faster in the presence of a *holo*-PCP domain to accept the activated amino acid.[257] This question could be solved by investigating the effects that the native *holo*-PCP would have on the results gained from the assay; however, no such results were presented in the original work.

4.2.6 Discussion

The aim of this study was to develop a new general method for the high-thoughput screening of A-domains. This limits the use of such techniques as current phage display methods in the literature for A-domains that require the use of an azide group on the substrate to be screened. Other less specific phage display methods are generally more suitable for the directed evolution of binding interactions, because the binding interaction can be used in the selection of the successful mutants. It is much more difficult to use these technologies for evolution of catalytic activity as there is no easy way to identify the active mutants. While the inclusion of an azide group as an

accessible functional group on the resultant NRPS product would open interesting avenues for semi-synthetic products, it makes the technology incompatible with the original aim of a general method. Fluorescence polarisation, despite its robustness, as an assay requires the synthesis of complex probes (seven steps) and only provides information on binding. This produces a risk that without very careful thought in a directed evolution study the optimisation rounds will only optimise for the parameter that is being screened. The FTMS assay described for high-throughput screening sadly is also not relevant as, while it allows for the screening of a large range of substrates, it involves lengthy HPLC purification followed by injection into an FT mass spectrometer. Furthermore, if the PCP is part of a multi-domain enzyme, a way of cleaving off the PCP from the rest of the protein must be found. In the original study, in all examples analysed the PCP was either at the end of the protein chain, so a single cleavage reaction could release it, or a discrete domain. Cleavage and isolation of the PCP would be much more difficult if the PCP was in the middle of the protein, as is generally the case in NRPSs.

The two remaining assays to be considered are the optimised ATP/$^{32}PP_i$ exchange assay in 96-well format and the use of malachite green. While the use of the ATP/$^{32}PP_i$ exchange assay has been well proven, the facilities to work with radioactivity are not easily available and to achieve the assay in a 96-well format requires uncommon equipment such as a 96-well scintillation counter. The work of McQuade *el al.* was in part meant to solve this problem; however, the work had not been published at the time that this work was started. Furthermore while it resolves the issues of the handling radioactive material, the statement that the release of PP_i is rate-limiting in A-domain activity seems invalid. The rate of PP_i release measured was 600- to 2700-fold slower than that measured by the exchange assay in the absence of an acceptor substrate. Therefore the rate-limiting step in these conditions, must be the rate of unbinding of the other product, the amino acyl adenylate. This

intermediate is not released at all under normal conditions when an acceptor PCP is present, but in the absence of the PCP it does slowly leak out of the active site of the A-domain. The problem with this is demonstrated in the case of EntE (from the enterobactin gene cluster). There is a similarly slow release of PP_i (0.7 min^{-1}) for its natural substrate in the absence of a PCP, whereas the substrate analogue salicylate displays a markedly faster leak rate (9 min^{-1}) under the same conditions.[258] In this case, one might have assumed that salicylate was the natural substrate; however, either on addition of the PCP or by using the traditional ATP/$^{32}PP_i$, the result returns to that of the natural substrate being the most active. Had the use of the malachite green assay been published before this work, it would have provided a possible starting point to adapt the assay to take these factors into account.

As a final remark, while the ATP/$^{32}PP_i$ exchange assay appears to be the most reliable and practical method for the screening of A-domains, if one is to ignore the difficulties raised by the use of radio-labelling, the assay only measures the first half of the reaction of the A-domain. At the beginning of this chapter, it was noted that the role of the A-domain was two-fold and that the second activity was transfer of the activated amino acyl group to the down/-stream PCP. Therefore, it is believed that any assay used in the evolution of an A-domain ought to also take into account this second part of the process. If it does not then optimisation of an A-domains specificity to a novel substrate, may be at the expense of its ability to transfer the amino acyl group onto the PCP.

4.3 Enterobactin

4.3.1 Introduction

The word ‘siderophore’ came from the Greek ‘iron carrier’ and, while they form a structurally diverse family of compounds, they all are high-affinity chelating com-

pounds secreted by bacteria and fungi under low-iron stress to scavenge iron.[259] Research into siderophores was stimulated by the realisation that nearly all aerobic and facultative anaerobic organisms (organisms that can undergo both aerobic respiration in the presence of oxygen and anaerobic fermentation without oxygen) synthesise at least one siderophore. Despite being one of the most abundant elements in the Earth's crust, the development of an aerobic atmosphere caused surface iron to become locked into sparingly soluble oxides and limited its bioavailability.[260] At physiological conditions, the uncomplexed ferric ion in solution is probably at a concentration of no greater than one attomolar while a level of one micromolar of iron is needed for optimal bacterial growth. This is because aerobic bacteria need iron for many reactions that use atmospheric oxygen (see section 2.1) and as such is essential for growth. These environmental pressures have led to formation of biological molecules that can compete with hydroxyl ions in solution for the scavenging of ferric state of iron, an abundant but trapped resource.

Siderophores are also important for pathogenic bacteria in mammalian hosts. Within the mammalian host, uncomplexed iron is strongly regulated by such proteins as transferrin at a level of about 10^{-24} M, well below the levels required by pathogenic bacteria.[261] The importance of iron-limiting microbial growth is displayed by the remarkable effect excess iron has on increasing the virulence of a wide range of organisms,[262, 263] as demonstrated by non-lethal injections of *E. coli* into guinea pigs, which become lethal if the addition of iron is great enough to saturate the transferrin.[264]

Enterobactin was simultaneously isolated from *Salmonella typhimurium* and named as enterobactin by Pollack and Neilands, and from *E. coli* by O'Brien and Gibson, who named it enterochelin. The O'Brien and Gibson paper was submitted first, but the Pollack and Neilands paper was published first; as such, both names were adopted.[265, 266] Enterobactin (shown in figure 4.9) is a macrocyclic compound com-

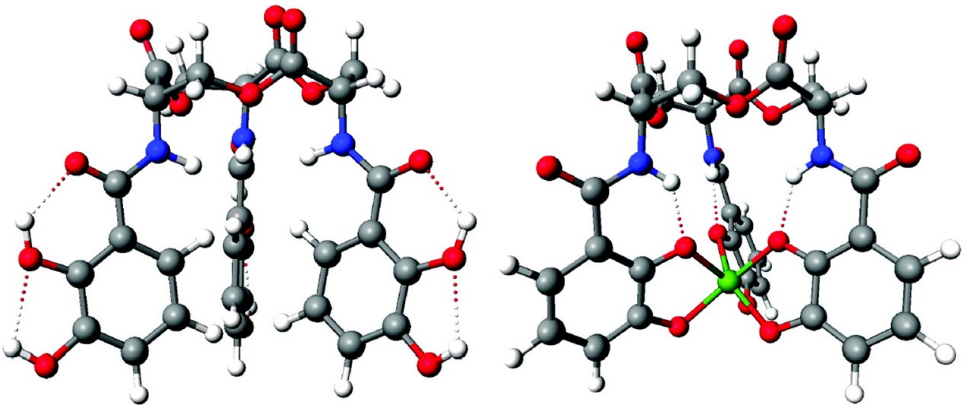


Figure 4.9: Enterobactin: space filling structures, left unbound, right bound.[267]

posed of three 2,3-dihydroxybenzoic acids bound by amide bonds to a cyclic structure made from three serine residues connected by ester linkages. The formation constant of the complex with ferric iron is $10^{52}M^{-1}$, the highest ever reported for a siderophore.[268] Within the neutral unbound compound, the *ortho*-hydroxy proton is able to form a hydrogen bond to the adjacent amide oxygen atom and upon deprotonation it changes so the amide proton hydrogen-bonds to the *ortho*-hydroxy oxygen atom rotating the phenyl ring through 180 degrees. In physiological conditions this results in about 50% of all the catecholate units having their hydroxy groups oriented inline with the amide proton. The presence of outward facing hydroxy groups in the absence of iron allows the free ligand to rapidly undertake the initial binding of iron followed by a slower conformational change that leads to hexadentate binding by the catechol ligands. This gives the molecule the role of both searching and sequestering ferric iron. It is worth noting that at pH values lower than 6, protonation of the *meta*-hydroxyl oxygen atoms changes the ligation into the salicylate mode, in which the ferric iron is bound by the *ortho*-hydroxyl group and the carbonyl oxygen of the amide group.[268]

4.3.2 Biosynthesis of enterobactin

The gene cluster for enterobactin can be separated into three sets responsible for different functions: the first set of enzymes catalyse the biosynthesis of enterobactin (EntA–F), which is discussed below; the second set are responsible for transport of enterobactin back into the cell (FepA–G); and the third allows the utilisation of the bound iron (Fes) by the hydrolysis of the lactone linkers to release ferric iron within the cell (see figure 4.10). Furthermore, the biosynthesis can itself be subdivided into two parts, one for the synthesis of 2,3-dihydroxybenzoic acid (2,3-Dhb) and the other for NRPS-catalysed formation of the macrocycle. The biosynthesis of 2,3-Dhb is catalysed by enzymes (EntA–C) from chorismic acid (see figure 4.11). The first step is catalysed by the isochorismate synthase, EntC, that catalyses an unusual

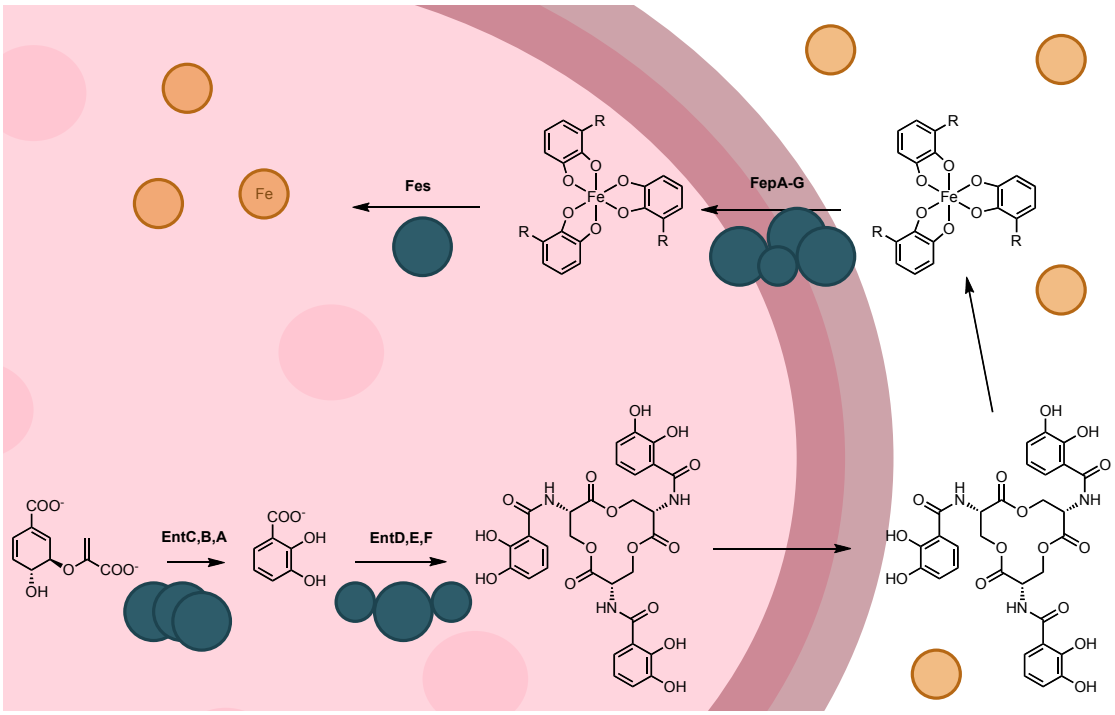


Figure 4.10: Components for enterobactin utilization in *E. coli*: The proteins required for synthesis (Ent), transport (Fep), and utilisation (Fes) of enterobactin are shown.

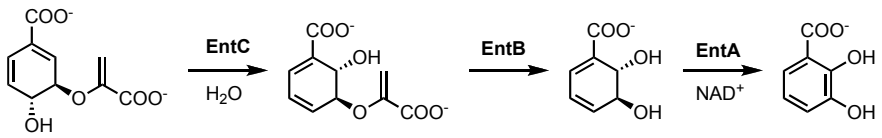


Figure 4.11: Biosynthesis of 2,3-dihydyroxybenzoic acid

1,5-addition/elimination reaction. Analysis by $H_2{}^{18}O$ showed the oxygen atom at position 2 was sourced from the solvent and not from the hydroxyl group lost from the 4-position. It is believed that the mechanism proceeds *via* a Michael-type addition of water, forming an enzyme-stabilised enolate, followed by elimination of the 4-hydroxyl group. This is then followed by the hydrolysis by isochorismatase, which is one domain of EntB, liberating 2,3-dihydro-2,3-dihydroxybenzoate. Finally, EntA acts as an alcohol dehydrogenase, but the unstable cyclohexadienone tautomerises to the more stable catechol, 2,3-dihydroxybenzoic acid.

The formation of the macrocycle from 2,3-Dhb and L-serine is catalysed by a NRPS consisting of EntE, EntB and EntF. In addition EntD is a 4'-phosphopantetheinyl transferase required for the post-translational modification of EntB and EntF.[99] EntE is the first A-domain within the NRPS and is notable for the fact that the domain is a discrete enzyme. EntE activates 2,3-Dhb and catalyses the loading of the acid onto the carrier domain of EntB (the other domain being the isochorismatase responsible for biosynthesis of 2,3-Dhb). EntF contains the remainder of the NRPS, including the A-domain for the activation of L-serine, a PCP domain and C-domain responsible for forming the 2,3-Dhb-L-serine dimer. The terminal thioesterase domain of EntF later releases enterobactin after cleavage of the thioesters of three molecules of DHB-Ser-S-PCP by intermolecular cyclization.[269]

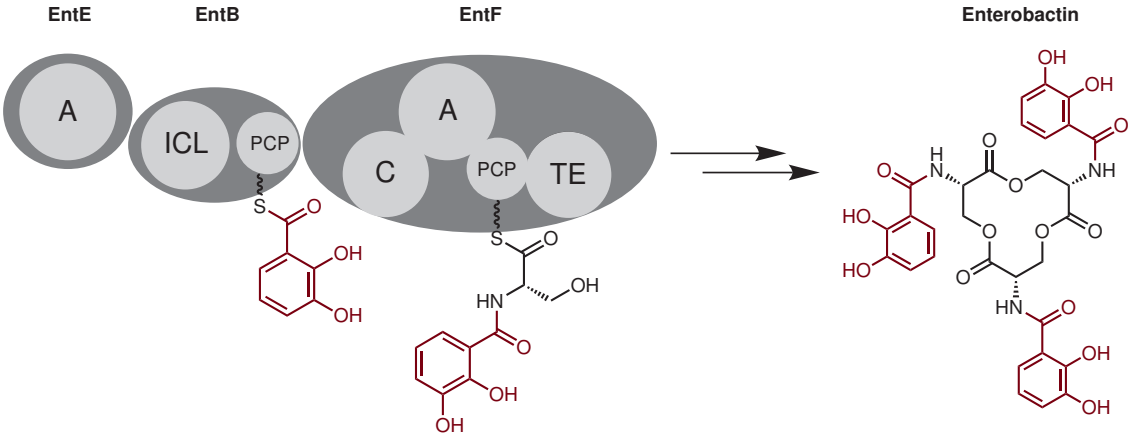


Figure 4.12: The enterobactin NRPS

4.4 Investigating the use of luciferase to determine A-domain activity

Beetle luciferase from *Photinus pyralis* (firefly) has been used in a wide range of *in vitro* and *in vivo* studies due to the high quantum yield of the bioluminescence reaction. Examples include the ultra-sensitive detection of pathogens, DNA sequencing and as a genetic reporter for imaging.[270, 271] The appeal of luciferase as a reaction probe is due to the fact that one substrate is ATP, allowing it to be coupled to a wide range of biological processes, and the high sensitivity of the reaction. In addition to this, recombinant luciferase is known to be fully functional when expressed in a wide variety of cells. In a typical assay, the level of ATP is monitored by the output of light from the reaction shown in figure 4.13.

$$\text{D-Luciferin} + \text{ATP-Mg}^{2+} + \text{O}_2 \xrightarrow{\text{Luciferase}} \text{Oxyluciferin} + \text{AMP} + \text{PP}_i + \text{CO}_2 + \text{photon}$$

Figure 4.13: Overview of the luciferase reaction

The luciferase used for this work was the X5 mutant,[272] which was developed because the wild-type enzyme displays a number of undesirable characteristics, such

as inactivation at elevated temperatures and red-shift in its bioluminescence emission at lower pH, affecting the sensitivity of the recording photomultiplier tube (PMT).[273, 274, 275]

Two methods of monitoring the reaction with luciferase end-point assays were tested. The first was the use of luciferase to measure ATP depletion at a chosen time point. The reaction was quenched on ice, and the ATP concentration recorded by detecting the intensity of the initial flash of light produced by the injection of luciferin into the combined luciferase and reaction product. This assay, while avoiding the need of radio-labelling, did not provide the benefit of measuring both parts of the A-domain's catalytic function; however, all the materials for the assay were readily available, which allowed for a rapid preliminary study.

The second method considered was that of utilising the PCP-bound product and incubating it with both AMP and PP_i in the presence of the A-domain to be screened. In theory, the amino acid would be released and ATP generated. As the thioester linkage in the product is of similar energy to the phosphate-phosphate linkage in the ATP starting material and catalysts do not alter the position of equilibrium, it can be expected that the A-domain will also catalyse the back reaction, formation of ATP from an excess of the products. While it is possible to synthetically create the acyl-PCP product, it has been shown that *N*-acetylcysteamine (SNAC) thioesters of amino acids can also be loaded onto NRPSs, thus providing a simple product mimic that can be used in place of the acyl-PCP.[105] As shown in figure 4.14, the reason for the success of SNAC thioesters as surrogate NRPS substrates is because they mimic the end of the phosphopantetheine arm **1** of the PCP.

For the initial feasibility studies, EntE was chosen from the enterobactin gene cluster of *E. coli* K12 because it is a well studied enzyme. It was known to express well and is a discrete domain, avoiding problems that may be caused by the presence of other domains or the loss of activity that may result from expression of a single

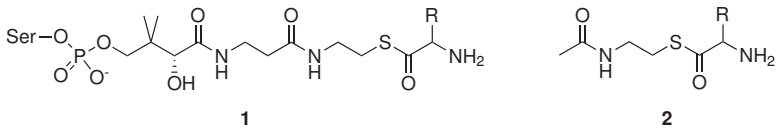


Figure 4.14: Comparison of compound **1**, the phosphopantetheine arm bound to the PCP serine and compound **2**, the SNAC amino acid derivative.

A-domain from a multidomain protein. The kinetics of EntE have also been studied for several substrates using the traditional ATP/$^{32}PP_i$ exchange assay, allowing for a comparison of results between the two methods.

4.4.1 ATP depletion assay monitored by luciferase

For the depletion assay, EntE was incubated with its substrate (27 μM) and ATP (100 μM) at 37 °C. After one and a half hours, the ATP was found to have depleted to 40% of that found in same mixture kept on ice. This cannot be solely due to the utilisation of ATP in the formation of the acyl-AMP as the initial concentration of 2,3-Dhb was only 27% of that compared to the starting ATP concentration. The controls with no enzyme or substrate showed a reduction of ATP to 87% and 86% respectively when compared to the same mixtures kept on ice. This latter reduction in ATP over time can be explained by the hydrolysis of the ATP in water. The ATPase activity in the absence of the substrate was no higher than that of an enzyme-free control. This background is not large enough to explain loss of 60% (46% when corrected to background) of ATP found in the assay. A possible explanation would be that the enzyme, due to the lack of a suitable donor, hydrolyses the acyl-adenylate. The released acid could then be recycled as a substrate for the enzyme. The alternative of spontaneous hydrolysis in solution seems unlikely when Kellerman *et al.* reported at 37 °C and pH values between 0 and 8 no hydrolysis of benzoyl adenylate occurs.[276]

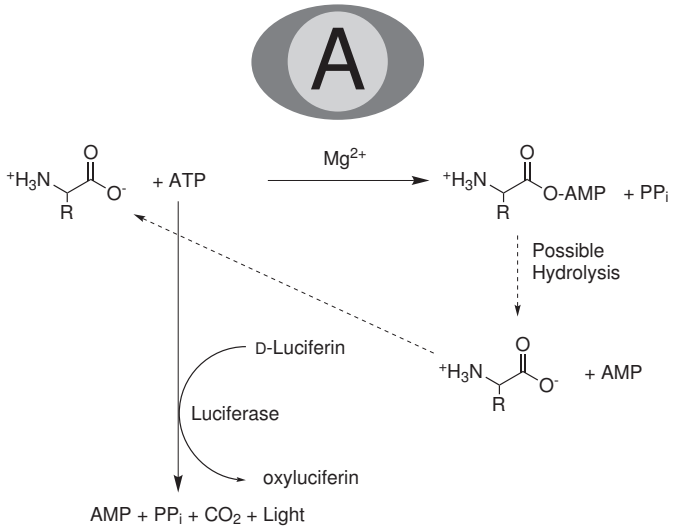


Figure 4.15: ATP depletion based assay. Luciferase and D-Luciferin added once a chosen time point is reached. The concentration of ATP can then be measured by the amount of light emitted.

4.4.2 ATP generation assay monitored by luciferase

For the reverse-reaction assay, EntE was incubated with the SNAC derivative of its substrate (23 μM) (synthesised from 2,3-dihydroxybenzoic acid and SNAC),[277] PP_i and AMP (100 μM) (see figure 4.16). The reaction mixture was allowed to incubate for one and a half hours at 37 °C, the reaction was screened for ATP and compared to a control under the same conditions but without the inclusion of the enzyme. The amount of light emitted was found to be about 370% higher in the enzyme/substrate incubation.

This initial result looked promising and so several other SNAC derivatives of potential substrates were synthesised, initially from 2,4-dihydroxybenzoic acid (2,4-Dhb), 2,5-dihydroxybenzoic acid (2,5-Dhb), 2,3,4-trihydroxybenzoic acid (2,3,4-Thb) and benzoic acid. For experimental see section 5.2.5. Previously, all of the acids had been tested with EntE as possible substrates by the use of the traditional ATP/$^{32}PP_i$ assay and kinetics published with respect to the forward reaction, with the

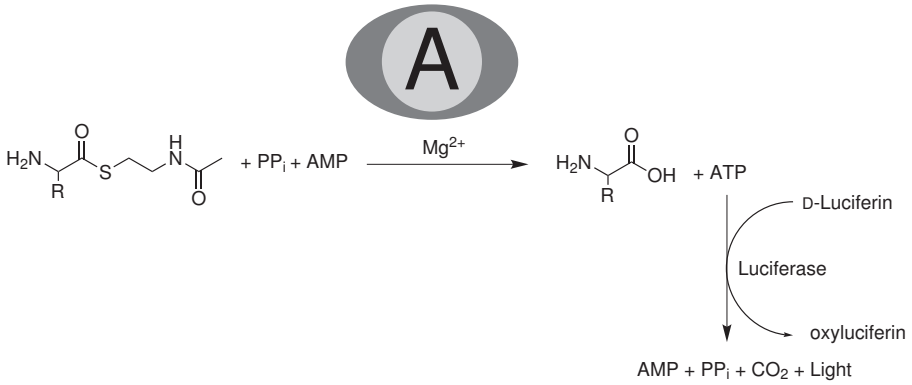


Figure 4.16: Luciferase based ATP production based assay. Lucferase and luciferin are added once the chosen time point is reached. The concentration of ATP can then be measured by the amount of light emitted.

exception of benzoic acid, which was believed to be inactive. Initial results showed no ATP generated by any substrate other than 2,3-Dhb-SNAC. Due to solubility issues 2,3,4-Thb-SNAC was not screened.

It was first assumed that the lack of detectable ATP formation with the substrate analogues could be due to the low sensitivity of the assay; therefore a series of studies were undertaken to optimise the assay including varying the pH and the substrate, imidazole and DMSO concentrations. The results from these assays showed that sensitivity could be increased by the equivalent increase in concentration of SNAC substrate (to 172.5 μM); however, any change in pH or increase of imidazole concentration was found to greatly impair the intensity of the chemiluminescence. Whether this effect is due to the change in conditions altering the rate of reaction of the A-domain or inhibiting the luciferase is unclear. The higher concentration of substrate was used in subsequent assays and is the method reported in the experimental section.

As shown in figure 4.17, even with these optimised conditions, the detected enzyme activity for EntE with any substrate other than 2,3-Dhb produced no significant enhancement of signal over that detected for the control reaction in the absence of

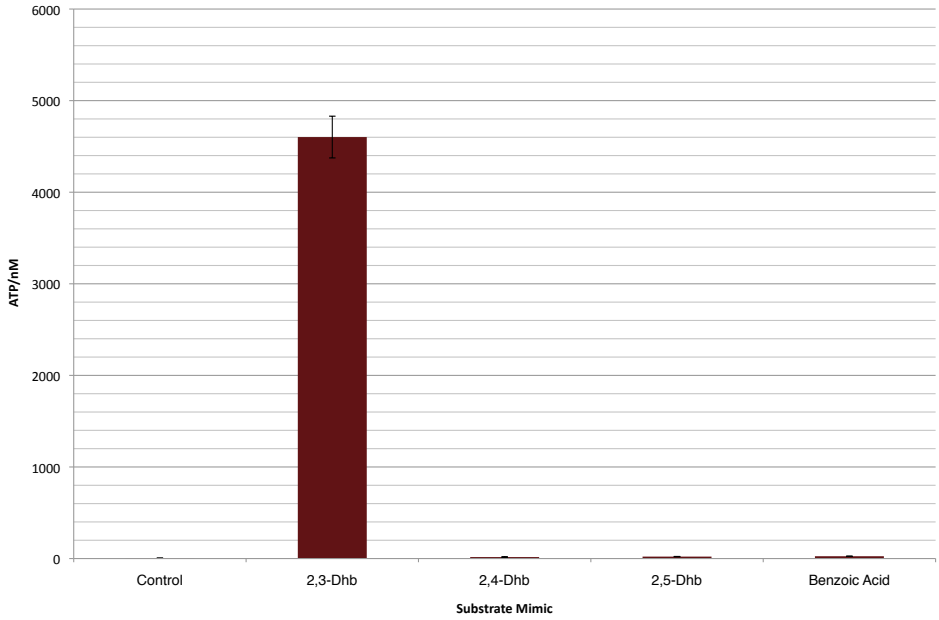


Figure 4.17: Results from measurement of the A-domain reverse reaction by luciferase, after optimisation of signal. Substrates are given as the acid from which the SNAC-thioester substrate is created.

substrate ($<$ 30 nM ATP generated). Extending the incubation time made no difference.

The kinetic parameters obtained for the EntE-catalysed reverse-reaction using 2,3-Dhb-SNAC ($K_{m,rev} = 536\ \mu$M, $k_{cat,rev} = 0.0018\ \text{min}^{-1}$) were compared with those from the traditional ATP/^{32}PP$_i$ exchange assay ($K_{m,ex} = 2.7\ \mu$M, $k_{cat,ex} = 330\ \text{min}^{-1}$) (see figure 4.18).

The high K_M for the reverse reaction and shape of the plot in figure 4.18 indicates the assays were not at saturating conditions for the enzyme with respect to the substrate concentration. This could not be solved as the substrate was not soluble at higher concentrations, and as such the data is extrapolated over a greater range of values than one would generally wish for reliable data. If the low rate and binding constant for the alternative substrates was the cause of no discernible result then one would expect that extending the assay time would have allowed for this. To

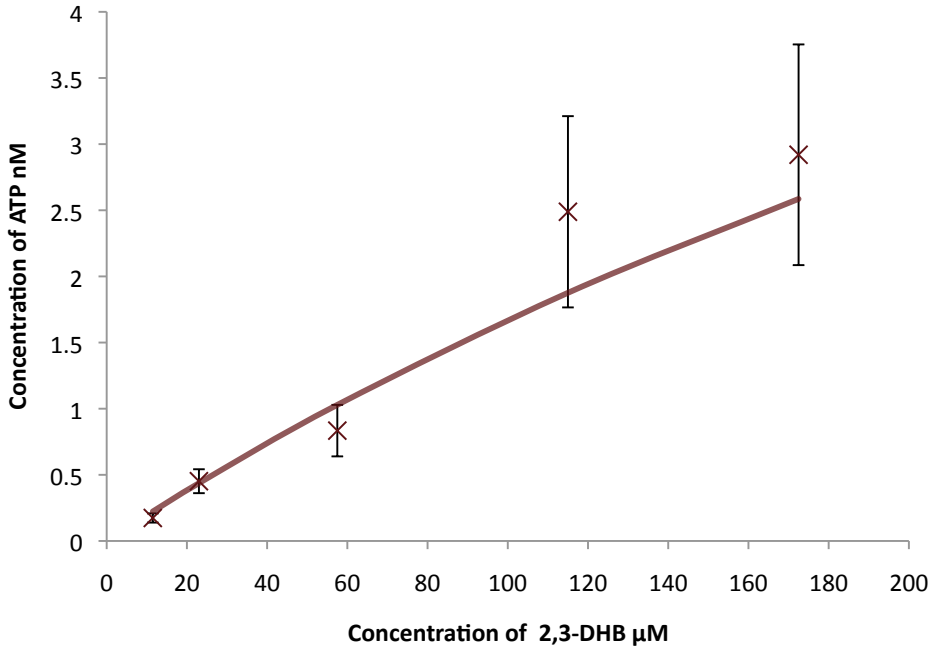


Figure 4.18: Michaelis-Menten plot of reverse A-domain reaction. Range of data was limited as the substrate was at saturation at the highest concentration shown.

see if the lack of detected signal with other substrates was purely a result of lower turnover rate the SNAC derivative of salicylic acid was synthesised. For the forward reaction salicylic acid (K_M = 91μM, k_{cat} = 150 min^{-1}) is the closest alternative substrate (with respect to kinetic parameters) to 2,3-Dhb. Nevertheless, even with the salicylate-SNAC, no increase in ATP production was seen compared with that of the control reaction .

4.4.3 Discussion

Initially the use of the forward reaction to monitor the A-domain reaction was only investigated as a test system to provide preliminary data to confirm details, such as if luciferase would be active in the presence of the EntE reaction buffer and reagents. In this respect, it was successful. However, the assay suffered when it came to applying it to measuring the turnover kinetics of the A-domain since, as the detected turnover rate was probably equal to the rate at which the acyl adenylate is hydrolysed or lost by the adenylation domain and not (as the aim had been) to create an assay which includes the PCP off-loading step. As such this assay was subject to the same criticism as the subsequently published malachite green assay[239] as discussed in the introduction of this chapter. In light of this result, work proceeded onto the synthesis and assay of SNAC product mimics.

The results of the ATP-generation assay can only be described as perplexing. The reason that the method only works for wild-type and no result can be seen for any other substrates is not immediately obvious. The literature suggests that the hydrolysis of the acyl adenylate within the active site and its subsequent release may serve as a proof-reading mechanism to minimise the incorporation of non-cognate substrates. [278, 279] This implies that the problem that is occurring is that the parameter measured by the reverse-reaction assay is not just that of the formation ATP, but also the equilibrium between this and the hydrolysis of the adenylate. In the case of all

the substrates except 2,3-Dhb, this means that, while the recorded background was low, once the activated adenylate intermediate was made then it preferably underwent hydrolysis rather than formation of ATP. In light of the fact that the leak rate of salicyl adenylate is over ten times greater than that of the 2,3-Dhb adenylate, this suggests that any remaining ATP that is generated is hydrolysed before reaching detectable levels.

4.5 A coupled assay for the monitoring of AMP production

A long standing method for assaying ADP and AMP production has been the coupling to lactate dehydrogenase to allow for the monitoring of NADH consumption by either UV absorbance or fluorescence. To assay ADP production, lactate dehydrogenase (LDH) is coupled to pyruvate kinase, which converts phosphoenolpyruvate and ADP into pyruvate and ATP. The pyruvate is then reduced by LDH forming NAD^+ from NADH. In nature, these enzymes are responsible for the final two steps of anaerobic respiration. In the case of A-domain activation, the resultant product is AMP, not ADP; therefore a third coupled enzyme is required, myokinase. Myokinase converts one molecule each of ATP and AMP into two molecules of ADP.

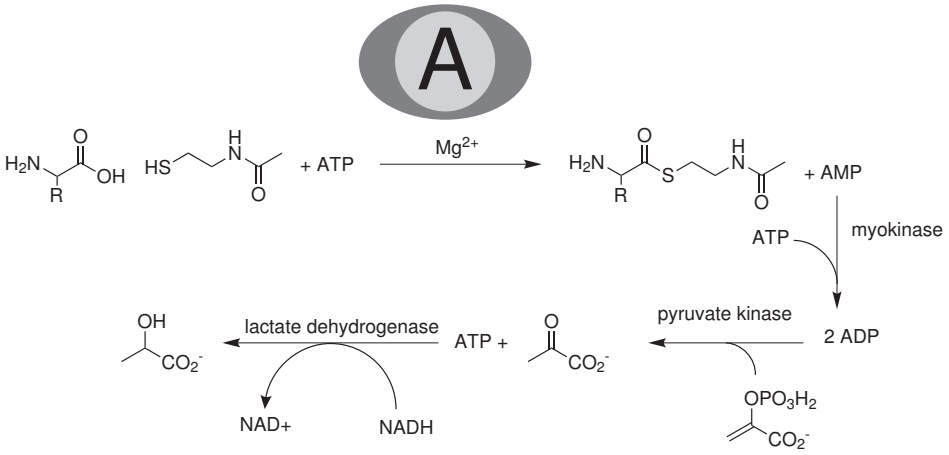


Figure 4.19: Schematic view of the NADH coupled assay.

Like the ATP/$^{32}PP_i$ exchange assay, the coupling to NADH by LDH and pyruvate kinase has a long standing history and has been used to characterise a wide range of systems. The coupled system also maintains several advantages over the traditional assay, by avoiding the requirement of specialist equipment such as a 96-well scintillation counter and the reaction can be continuously monitored using a 96-well format rather than just an end point value.

Although widely used in other systems, it was surprising that very few studies have used this coupled assay for the investigation of A-domains, and none of these used the method in a high-throughput manner. One such previous study investigated the interaction of EntB and EntE.[280] Importantly, the study showed that mutations in EntE could have drastically different effects on the measured enzyme activity depending on whether the ATP/PPi exchange or the coupled assay were used. This allowed a model describing which residues were important to the EntB/EntE docking interaction to be created.

4.5.1 Scaling down of the coupled assay

The aim was to perform the assay in a reaction volume of 100 μL, as required for high throughput on a standard plate reader using either absorbance or fluorescence. In the case of this study fluorescence of NADH was used, as it provided higher sensitivity. Nonetheless the use of the NADH absorbance band would still work for the majority of this study if a fluorescence reader were not available. The plates used had a total volume of 200 μL per well which would allow room for variation of conditions, if required for alternative systems, and for the mechanical shaking of the plate before reading to encourage mixing. Finally, EntB was also replaced by SNAC in the assay to provide a cheap alternative acyl acceptor and DTT was replaced by TCEP so as not to interfere with the A-domain second-half reaction.

4.5.2 Kinetic measurements

The initial turnover reactions were undertaken with all enzymatic substrates at ten times the concentration of K_m to ensure rate limiting conditions and that a linear rate was maintained during the time scale of the screening assay. The only exception was the inclusion of SNAC in far greater excess (94 mM) as it was found not to cause any inhibitory effects. To confirm the validity of the assay, control reactions were carried out that showed that the concentration of EntE was directly proportional to the rate of NADH consumption, confirming that the reaction rate was limited by the turnover rate of the A-domain and not that of any of the coupled enzymes. When other substrates were tested the relative turnover rates compared to that of 2,3-Dhb (after subtraction of the background rate observed in a substrate-free control) were: salicylate (0.62), 2,4-Dhb (0.07) and 2,3,4-Thb (0.02). The results correlated to the initial characterisation of EntE,[234] where the slowest reported k_{cat} was that for 2,3,4-Thb. Also the overall range of rates detected was 50-fold which was a promising start, even though as wide a range as possible would be helpful for a directed evolution study.

With the application of the assay for high-throughput screening confirmed, it was next decided to measure the kinetic parameters for the EntE A-domain and see if SNAC provides a successful alternative substrate in place of the EntB/PCP or if there would there be too great a reduction in the overall rate of reaction. The analysis of initial rates for EntE with varying concentrations of 2,3-Dhb gave a reasonable fit to the Michaelis-Menten saturation curve (see figure 4.20). The reported turnover rate of wild-type EntE, with its native PCP is $k_{cat} = 356\ \text{min}^{-1}$, which is 10–15 times faster than the rate observed here in the presence of SNAC. Interestingly, as with earlier studies to monitor the release of AMP from the A-domain to study kinetics, the production of AMP seen in the absence of a thiol donor was negligible,[280] which is in contrast to studies monitoring the PP_i released or my own results from monitoring

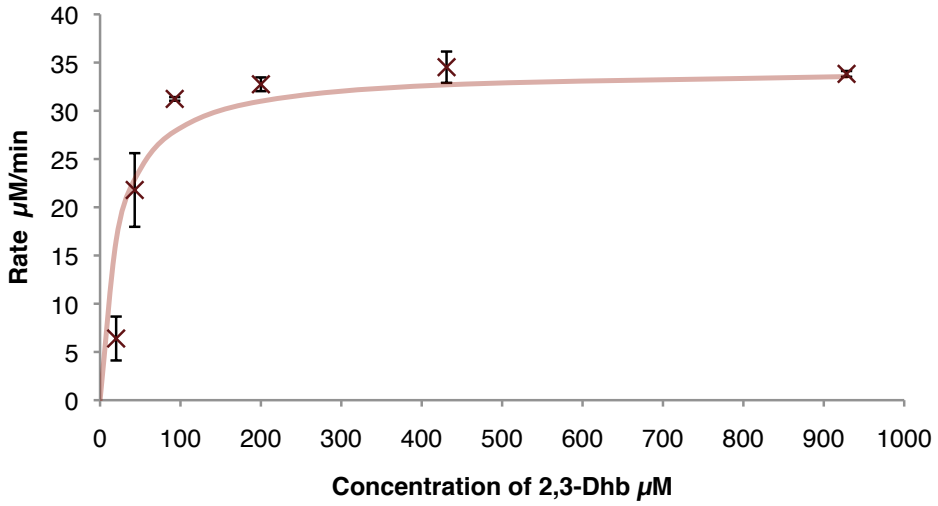


Figure 4.20: Michaelis-Menten saturation curve for EntE with 2,3-Dhb (K_M = 21.5 μM, k_{cat} = 27.78 min^{-1}).

the consumption of ATP with luciferase.

In order to confirm that kinetics could be calculated for a wide range of substrates EntE was incubated with varying concentrations of 2,4-Dhb and 2,3,4-Thb and data collected. The results are shown in figure 4.21. The results showed that data could be generated for substrates displaying a k_{cat}/K_M over 130-fold lower than that of the natural substrate whilst maintaining an excellent fit. This wide range is essential for the screening of possibly low initial activities in the directed evolution study, but the generation of high-quality kinetic data also allows for a profiling of the effects of mutations on the active sites.

4.5.3 Isolation and purification of EntB/PCP

In the hope of being able to compare results of activity assays using SNAC with those of the native EntB/PCP, and to see if they could be used within the luciferase assay to improve the rate of formation of ATP over that of the leak rate, it was planned to clone and express the EntB/PCP domain. As with EntE, the appropriate PCP was

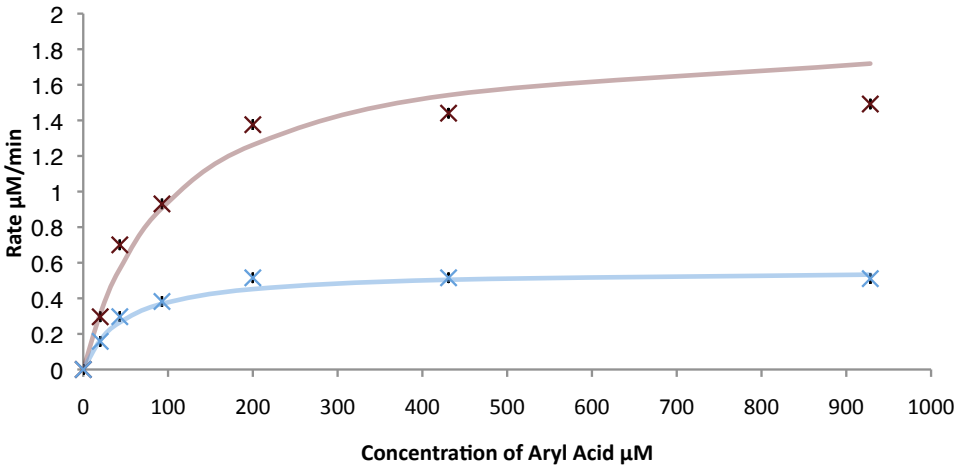


Figure 4.21: Michaelis-Menten saturation curve for EntE with 2,4-Dhb (red, K_M = 102.5 μM, k_{cat} =1.55 min^{-1}) and 2,3,4-Thb (blue, K_M = 48.3 μM, k_{cat} = 0.454 min^{-1}).

cloned from *E. coli* K12. Since EntB is a multidomain enzyme and both domains are not needed (the N-terminal domain being involved in the biosynthetic pathway from chorismate to 2,3-dihydroxybenzoate), it was decided that the C-terminal PCP domain should be expressed on its own without the rest of the protein. To identify the boundaries of the PCP domain, the NRPS-PKS knowledge-based resource[281] was used (the crystal structure is shown in figure 4.22).

The result suggested a cut within one of the alpha helices, hence it was decided to move the cut point into that flexible region between the two domains. To find an alternative cut point the sequence was aligned with the third PCP domain of *B. brevis* tyrocidine synthetase 3 (TycC) as this had previously been well characterised and was known to be soluble.[282]. Using the information gathered a cut site was then chosen preceding residue V200. The corresponding fragment of the gene was amplified out of the host organism *via* PCR, using primers to include restriction sites at the start and end of the DNA fragment produced.

This fragment was then ligated into a pET28-a(+) vector and transformed into Nova Blue competent cells. Multiple colonies from the transformation of the ligation

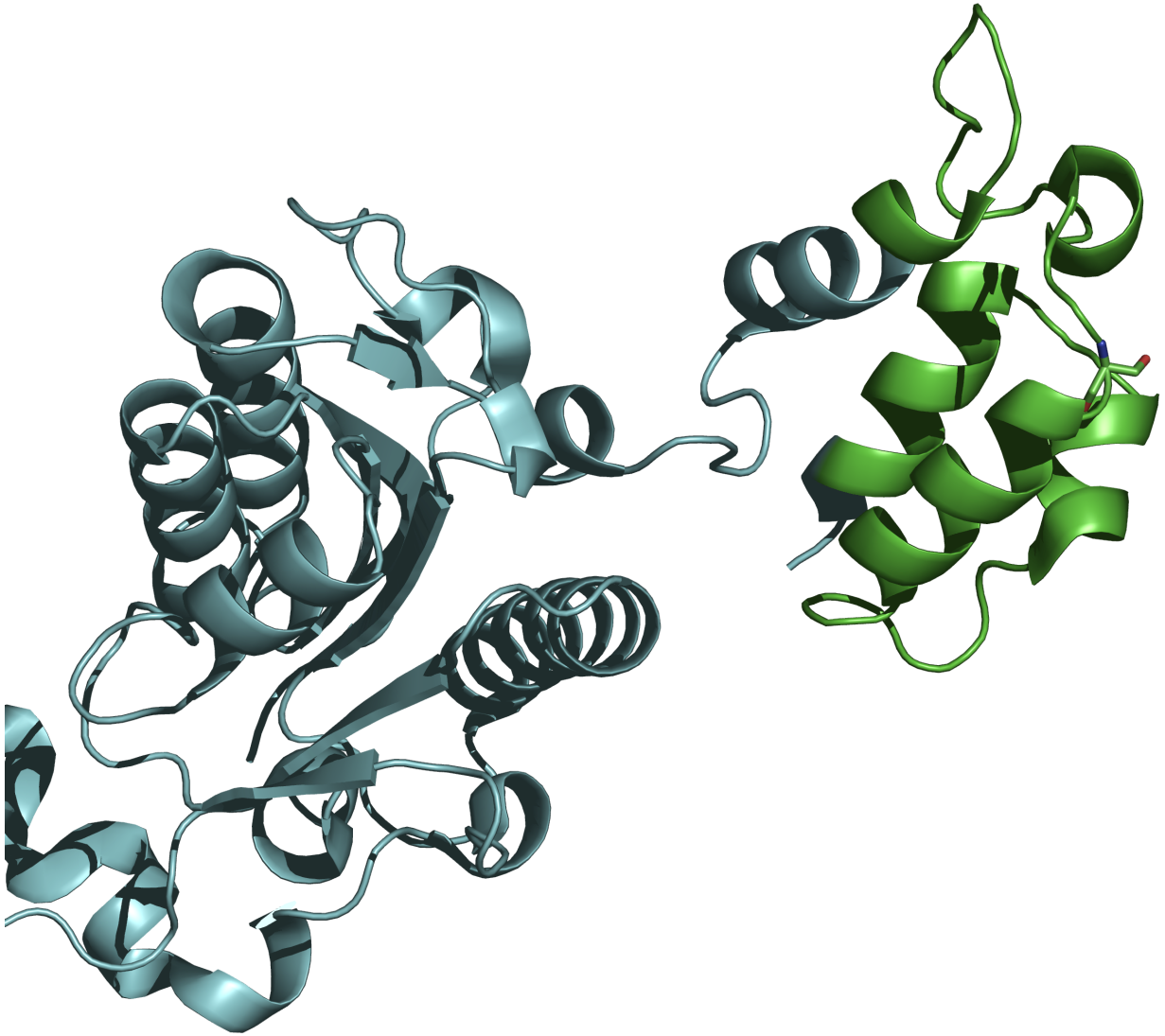


Figure 4.22: An X-ray crystal structure of EntB with PCP as identified by the NRPS-PKS resource domain highlighted in green. The serine that provides the anchor for the prosthetic pantethine arm is displayed on the right of the structure. (PDB:2FQ1).

mixture were then selected and sequenced to check that the correct insert had been incorporated and that no mutations had occurred (see figure 4.23).

MGSSHHHHHH SSGLVPRGSH MVMTEELLPA PIPASKAALR EVILPLLDES
DEPFDDDNLI DYGLDSVRMM ALAARWRKVH GDIDFVMLAK NPTIDAWWKL
LSREVK

Figure 4.23: EntB-PCP fragment sequence including His-tag.

Once a successful ligation product with no mutations was isolated, the vector was then transformed into an expression strain of *E. coli* BL21 and the gene over-expressed and the protein purified. The resultant protein was analysed by LCMS (see section 5.6.1) to determine whether the product was in the *apo*- or *holo*-form. The LCMS showed that about 20% of the enzyme produced was of the active *holo*-form. Both the *apo*- or *holo*-form were found to be missing the terminal methionine (see figure 4.24). This amount of active enzyme was considered high enough for an initial test of activity, as it was believed that the *apo*-form would not interact with the system.

The assays run with the mixed *apo*- and *holo*-form protein displayed an increased background rate that dominated the signal recorded. On investigation, this background rate was found to be dependent on the concentration of the PCP and thus suggested a co-eluted enzyme with ATPase activity, despite no other enzyme being visible by PAGE-Gel. In addition, a single assay ($K_{M,EntB}$ = 23.3 μM)[280] required approximately 100 times the concentration of the PCP over the A-domain in solution for steady state conditions. Despite multiple attempts to purify the enzyme its low yield and the high quantity required for a single assay made it difficult to obtain the purity and quantity required for any analytical study. It has since come to light that

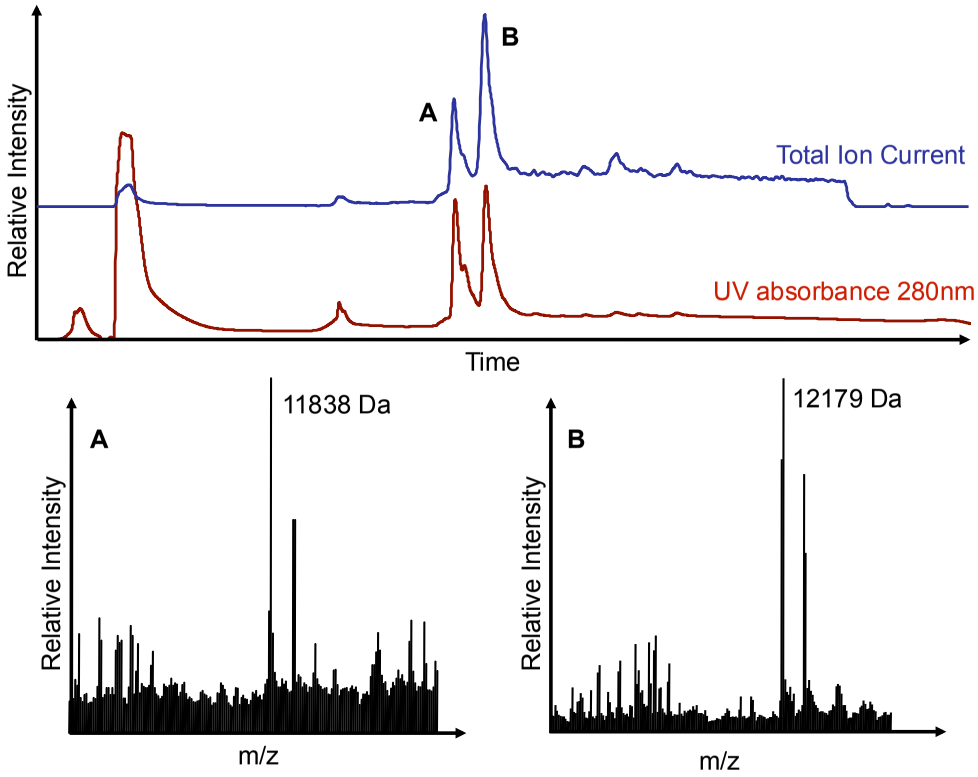


Figure 4.24: The total ion current and UV absorbance from the analysis of the purified protein by LCMS. The peaks representing both the *apo*- and *holo*-forms of the EntB PCP domain are labelled as **A** and **B** respectively. The *apo*-form's observed mass was11838 Da (11840.5 Da calculated) and the *holo*-form's observed mass was 12179 Da, displaying an increase of 339 Da to the calculated mass indicating phosphantetheinylation.

the difficulties of expression of PCP/ACP domains are not uncommon.[283] Therefore, while traditional methods of purification suffer poor yields the use of freeze-thaw selective release may well offer a route to obtaining the isolated PCP domain. Interestingly, in the initial example of purification of an ACP by this method the protein obtained in the crude supernatant was both in much higher concentration and purity than that which had been previously purified.[283]

4.5.4 Expression and protein purification in 96-well format

Originally it was hoped that it may be possible to use cell-free lysate in place of purified enzyme, however NADH was rapidly consumed. As with the traditional exchange assay it was confirmed that the cell lysis reagent was not responsible for this activity therefore it is presumed that it was due to the presence of ATPase activity within the cell free culture.[246] Thus the protein would have to be purified in a 96-well format as part of the screening process. The method used was based on the work published by Otten *et al.*[236] which had been successful for the isolation of the enzyme used in their study. Single colonies of *E. coli* harbouring the vector containing the His_6-tagged EntE were used to inoculate single wells containing LB/kan media within a 96 deep-well plate and were grown to an approximate OD_{600} of 0.8 before induction with IPTG. After 24 hours, the cell lysate was purified by use of a TALON 96-well plate. The eluted protein concentration of approximately 1 μM (200 μL) was consistent with previous work.[236] Using 20 μL of this solution in a 100 μL assay would result in an enzyme concentration of about 200 nM, a quarter of the concentration used in the preliminary studies. This reduced quantity of enzyme caused no problem in the assay. Previously in the case of the natural substrate all the NADH was consumed within the first 5 minutes of the total 30 minutes of the assay so the lowered rate due to lower enzyme concentration still allowed the reaction to be followed to completion in the same time-frame.

4.5.5 Detection of activity in a selection of active site mutants

To demonstrate the capability of the assay we sought to create several active-site mutants to screen. As there was no crystal structure for EntE to provide structural information, the active site was modelled using Phyre[284] and the DhbE[285] from *Bacillus subtilis* (PDB:1md9), which shares a 47% sequence identity with EntE. The residues highlighted in the active site structure (shown in figure 4.26) are those found within 5 Å of the substrate molecule. This is located within the active site and are nearly identically positioned to those within DhbE.

```
dhbE            SVYLAALPTAHNYPLSSPGVLGTLYAGGRVVLSPTPSPDDCFPLIEKERVTITALVPPLA 283
pchD            SVYLVALPMAHNFPMCCPGFIGTFSVGGRVVLSPSPSPEVCFELIERQGVTHTALVPPLA 290
entE            TRYLCAIPAAHNYAMSSPGSLGVFLAGGTVVLAADPSATLCFPLIEKHQVNVTALVPPAV 283
vbeE            TRYLCVLPAAHNFPLSSPGALGVFWAGGCVVLSQDASPQHAFKLIEQHKITVTALVPPLA 287
ytbE            SVYLAVLPVAHNFPLACPGILGTLACGGKVVLTDSASCDEVMPLIAQERVTHVALVPALA 270
mbtA            DVYLAVLSAGHNFPLACPGLLGAMTVGATTVFGTDPSPEAAFATIARHGVTVTALVPALA 298
                  ** .:. .**:.:..** :*.:  *. .*:   .*    :  * :. :. .****. .

dhbE            MVWMDASGS--RSDDLSSLQVLQVGGAKFSAEAARRVKSAFGCTLQQVFGMAEGLVNYTR 341
pchD            LVWLEAAQA--RGRGLAPLQLLQVGGAKLSYEAARRIEPVLGCRLQQVFGMAEGLICYTD 348
entE            SLWLQALIEGESRAQLASLKLLQVGGARLSATLAARIPAEIGCQLQQVFGMAEGLVNYTR 343
vbeE            LLWMDHAEK--STYDLSSLHFVQVGGAKFSEAAARRLPKALGCQLQQVFGMAEGLVNYTR 345
ytbE            QLWVQAREW--EDSDLSSLRVIQAGGARLDPTLAEQVIATFDCTLQQVFGMAEGLLCFTR 328
mbtA            KLWAQACEWE--DNPPKSLRLLQVGGAKLEADDARVIRSAL------------------- 337
                 :* :            .*:.:*.***::.   *  :   :
```

Figure 4.25: Alignment of the EntE sequence with other aryl acid A-domains. Green - Within 5 Å of the bound substrate. Blue - Part of the NRPS code. Yellow - Both within 5 Å of the bound substrate and in the NRPS code

The same information as shown in figure 4.26 is shown in the alignment of the EntE protein sequence with that of other aryl-acid A-domains (see figure 4.25). Of the amino acids within 5 Å of the aryl acid binding site, only 4 correspond (a conserved lysine is not shown in alignment) to the residues identified as being within the NRPS substrate specificity code. The lysine visible in figure 4.26 is also of little importance as it is conserved through all A-domains as it is responsible for the co-ordinating acid group that is activated. Taking this into account, it was concluded

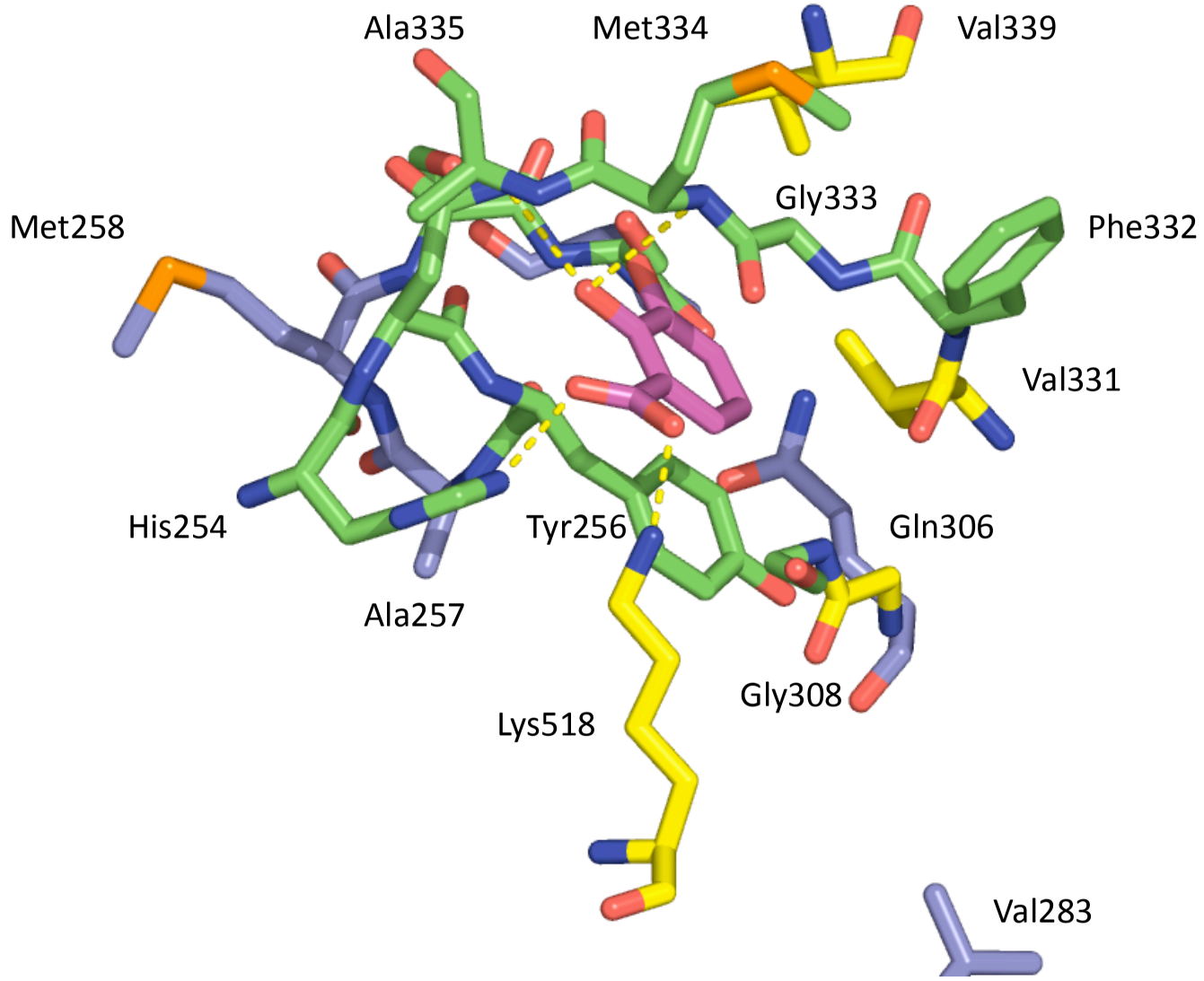


Figure 4.26: Predicted EntE active site structure. Green - Within 5 Å of the bound substrate. Blue - Part of the NRPS code. Yellow - Both within 5 Å of the bound substrate and in the NRPS code.

that, with the exception of the conserved lysine residue, the three amino acids, all of which were within the 5 Å radius of the active site and defined by the NRPS code as being responsible for substrate specificity would be most suitable ones to mutate in a test study.

Several primers were designed with an NNK (in the forward direction, where N = G, C, A or T and K = G or T) degenerate codon at the position of the amino acid to be mutated. The wild-type plasmid was then used as a template DNA for site directed mutagenesis PCR, giving a mixed product of template and degenerately mutated DNA. As the site directed mutagenesis protocol does not undergo an exponential reaction product formation the initial template DNA was digested by Dpn1 before transformation into super-competent cells. Ten colonies were picked from each plate and sequenced to confirm that randomisation had occurred at the desired codon. In the first round of PCR, this was found not to be the case, with a large preference for the wild-type sequence. The solution was to alter the PCR reaction parameters, such as binding temperature, to those given in the experimental. From the sequenced library of mutants, 2 examples of mutants at each mutation site were chosen for screening at random and both an initial high-throughput screen and the Michaelis-Menten kinetics for each mutant were investigated. Of the six mutants assayed, three

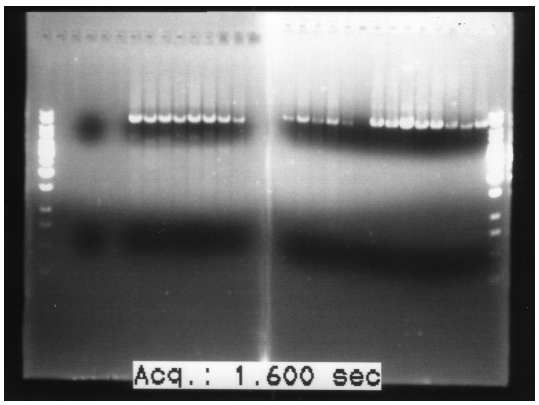


Figure 4.27: PCR products from saturation mutagenesis reaction.

Table 4.1: Residues within the EntE active site

Position	Mutation	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M	HT Screen
WT	–	30.7	24.1	0.78	100
G308	Tyrosine	–	–	–	0.5
G308	Arginine	–	–	–	0.2
V331	Leucine	42.1	1.4	0.033	1.15
V331	Glycine	–	–	–	0.4
V339	Methionine	231	3.8	0.016	13.5
V339	Lysine	30.5	1.2	0.039	6.15

showed minimal activity that could only be detected at the higher concentrations of substrate (up to 2000 μM). For these mutants the data obtained was inadequate for determination of the kinetic parameters.

While the data set is too small to make strong conclusions, both mutants at the 339 position are active which is in agreement with it being the only one of the three residues to show variation at this position in the other aryl acid homologues. As both positions 308 and 331 are conserved between the aryl acid A-domains, a lack of flexibility is to be expected, but in the light of the large differences in the bulk of the residue sidechain, for example glycine to tyrosine, no strong conclusions can be made. What the data do show is that the activity of a subsection of mutants could be detected over a 500-fold range, something that is required in situations such as directed evolution studies where very small activity is often observed.

4.5.6 Discussion

The aim of this work was to develop an alternative activity assay to those currently in the literature, with the aim of being as flexible as possible and eliminating a requirement for radio-label or specialist equipment. It was noted in the introduction that

methods such as phage-display provide an interesting and very rapid screening process, but that it is very difficult to select phage on the basis of their catalytic activity. The published phage-display work on A-domains is promising and will hopefully lead to interesting results but it was not considered a suitable solution for this project due to the limitations on the choice of substrate. In comparison, the ATP/$^{32}PP_i$ high-throughput assay has many disadvantages: a single 96-well plate must be monitored by a scintillation counter for 48 hours to gain accurate results, and the higher throughput method of fluid activity counting has a 10-100 times lower dynamic range when compared to the charcoal counting method. Where the procedure does excel is in its ability to measure, *via* the charcoal counting method, over a 1000-fold dynamic range.

In comparison the coupled assay developed and used in this chapter was able to detect activity over a 500-fold range (see table 4.1). While lower than that of the more sensitive ATP/$^{32}PP_i$ charcoal counting method, activity was measured in 30 minutes rather than 48 hours. A second advantage of the coupled assay is that the data is continuously measured, rather than just at some end point, allowing the user to clearly follow the reaction course if required. In summary, while the ATP/$^{32}PP_i$ exchange assay has a greater sensitivity, it is estimated that only ten plates could be screened this way per week, whereas we believe that it would be feasible to achieve ten a day *via* the method proposed here. Therefore, even if it did transpire that the initial rounds of mutations were not detectable by the coupled assay, a switch of assay choice once activity is high enough for detection by more rapid screening processes would allow for a rapid acceleration in the selection of active mutants.

One drawback of directed evolution-based work is that the result of the study is very dependent on the assay. Using the example of the malachite green assay, if one aimed to optimise the parameters detected, the result would undoubtedly be an enzyme that rapidly releases PP_i, although, it may well not be active within the NRPS

due to the rapid loss of the acyl-adenylate before it can be captured by the PCP. It has previously been acknowledged that similar difficulties occur with the traditional assay as it does not take into account the loading of the acyl group onto the phosphopantetheinyl arm of the carrier domain, which means that, while the selection has been for an enzyme that rapidly forms the adenylate, there is no confirmation that the phosphopantetheinyl arm is able to still interact with the substrate. In the past it has been suggested that this can be tested in the context of a recombinant A-PCP didomains construct before cloning into the whole NRPS.[236] However, this method could lead to a large number of inactive enzymes after a long evolution study. Instead, it would seem logical to use the independent measurement of AMP release by this coupled assay and the measurement of PP_i release or ATP consumption, either *via* the use of a malachite green reagent or luciferase. This would provide both the rate of the loading onto the thioester and the hydrolysis of the adenylated intermediate. This is highly advantageous over the use of a single assay as the ratio of formation of the thioester to the background leak rate would seem to be the critical parameter, not either part on its own, and thus a more desirable variable on which to exert evolutionary pressure. Such a method would have a greater chance of producing mutant A-domains with activity within the NRPS.

4.6 Substrate specificity of the first A-domain of the teicoplanin NRPS

4.6.1 Heterologous expression of TeiA module 1

A particularly interesting feature of the first module of the teicoplanin NRPS is that, as with the chloroeremomycin NRPS, it results in incorporation of D-*p*-hydroxyphenylglycine (Hpg) but the enzyme that makes Hpg (see section 1.10.1) is known to make the L-enantiomer and the module has no epimerisation domain. It is not known whether the A-domain accepts D- or L-Hpg, nor which enzyme catalyses the

epimerisation (or direct formation of the D-Hpg). Accordingly, by cloning the A-domain alone into *E. coli*, the heterologously expressed domain could be assayed with D- and L-Hpg to reveal its substrate preference. This type of experiment has previously been described for the first A-domain of the chloroeremomycin NRPS,[61] which has the same uncertainty over whether the substrate is D- or L-leucine. Interestingly L-leucine was preferred but only by a factor of 6:1. This previous work only monitored the selectivity of the activation of the amino acid, however, and would not have detected any selectivity in loading the amino acyl group onto the PCP. The methods we have proposed would detect the overall selectivity of both steps.

Under the supervision of Dr. J. Andexer and myself, P. Rombach undertook the heterologous expression and characterisation of the first module from the teicoplanin NRPS (TeiA-A_1) as part of the research for her diploma thesis.[286] Using the knowledge based prediction tool NRPS-PKS,[281] the potential domain residues of the proposed TeiA-A_1 were identified as the first 503 amino acids of the teiA enzyme. To confirm that this gave a product that was consistent with previous isolated A-domains, a structural model was generated using Phyre[284] from the primary sequence that was planned to be cloned. The resultant protein structure generated for TeiA-A_1 was modelled from the structure of the closest homologue of known structure, the phenylalanine activating domain from the first module within the gramicidin S synthetase[28] (PDB: 1amu). We also compared this to the structure of DhbE[285] (PDB: 1mbd) as it had provided soluble protein in previous work using EntE. As the models generated in both cases gave similar predictions, there was no reason to shift the domain bounds from those initially predicted.

Two primers were designed by P. Rombach for the amplification of the potential teiA-A_1 domain, incorporating *XhoI* and *NcoI* restriction sites at the 3'- and 5'-end respectively to allow for the ligation into previously digested pET28a(+). Sequencing of the resultant vector confirmed the presence of the gene within the vector.

Test expression of the transformed vector in *E. coli* BL21 (DE3) cells was carried out. Analysis of the soluble protein fraction after lysis showed a 70 kDa band on SDS-PAGE gel that was not present in the negative control. On comparison of the insoluble protein fraction the same band was present and it is estimated that only 25% of the expressed protein was soluble; however, sufficient protein was isolated from a 1 L culture for initial screening of the protein.

The AMP detection assay was carried out, using the same method as for the trials on EntE, on the purified enzyme and the results are displayed in figure 4.28.

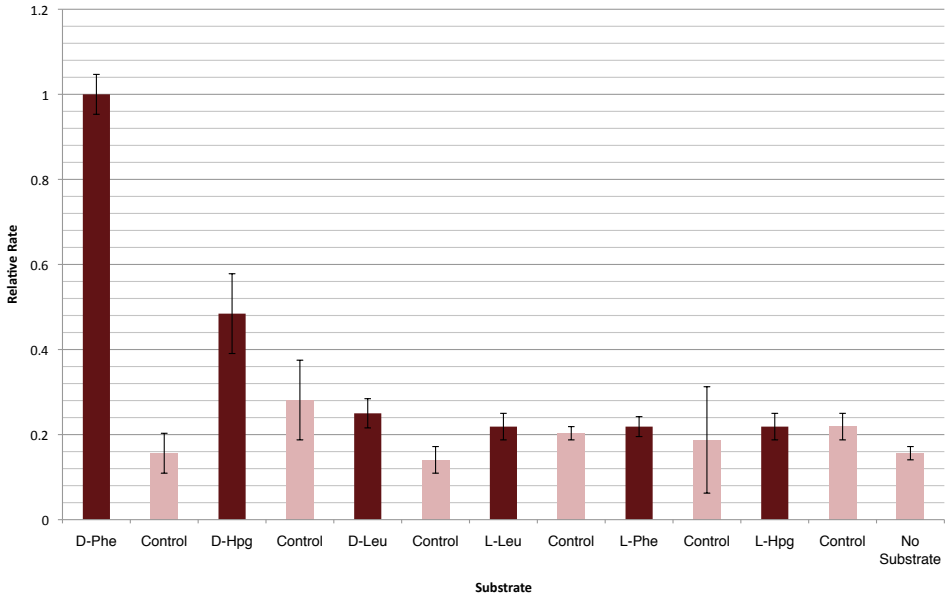


Figure 4.28: Enzymatic activity of teiA-A_1. Amino acid with teiA-A_1 as listed with enzyme-free control to its right.

4.6.2 Discussion

The values obtained from the initial screen were all repeated in triplicate at the time of the initial screening; however, due to the timescale of the diploma project and difficulties with the stability of the protein, it was not possible to repeat this procedure

to confirm the values obtained. The preliminary result obtained here shows a clear preference of the D-amino acid in contrast to the A-domain studies of the equivalent domain in chloroeremomycin (CepA1), [61] which preferentially activates the L-amino acid leucine. In addition, CepA1 only appears to load D- and L-leucine whereas TeiA-A_1 appears to show higher activity for D-Phe and D-Hpg in preference to any L-enantiomer. This is probably due to the increased size of the side chain of Hpg over Leu, resulting in greater stereoselective pressure in TeiA-A_1. There is also the secondary effect that the greater volume required for the Hpg side chain allows for the accommodation of other substrates more easily.

These results raise several important questions: As it is believed that Hpg is synthesised by the host organism as the L-enantiomer, at what stage does epimerisation occur? Furthermore, is it a process that is catalysed by a promiscuous enzyme within the organism? In both cases, TeiA-A_1 and CepA, we have results that show the loading of an amino acid, which is not the amino acid found in the final peptide: is the selectivity for the correct enantiomer in CepA1 and the Hpg over Phe in TeiA-A_1 carried out by the C-domain? This seems highly likely in the case of CepA if we recall that epimerisation domains are unable to exert stereochemical control and it is therefore the C-domain that creates the selectivity. However, as previously noted, in the TeiA-A_1 system no epimerisation domain is present. Is the difference in results between the two systems in fact due to selective loading of the activated amino acyl group onto the phosphopantetheinyl side-chain by the A-domain? If the assay were carried out on the CepA1 domain would the D-amino acid be preferred?

One final speculation: all the A-domains within the glycopepetide gene clusters show, according phylogenetic analysis, more similarity to the A-domain catalysing the activation at the same position in the glycopeptide, rather than those that activate the same amino acid.[287] This suggests an ancestral link between TeiA-A_1 and CepA1, and could imply that an epimerisation domain once existed adjacent to the

first module in both NRPSs; however, due to other epimerisation pathways within the cell, it was not maintained by evolutionary pressure before the two classes of glycopeptide diverged.

4.7 Future work

The work of this research focused on the first A-domain within the NRPSs for simplicity. Alteration of the substrate specificity of subsequent A-domains within an NRPS could be compromised by the substrate specificity of the upstream C-domain when the NRPS is reconstituted; and this would consequently reduce the yield of product from the system. Previous work has shown that the C-domain on EntF will turn over *in vitro* to produce dipeptides using SNAC thioesters of the nucleophilic amino acid,[288] and so the thioester linkage in the donor SNAC or PCP will be converted to the free-thiol as the amino acid is loaded onto the nucleophile. This newly generated thiol can then be detected by reaction with the colourless 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) to give yellow 5-mercapto-2-nitrobenzoic acid. Once an A-domain of altered substrate specificity is obtained, it would be of interest to use this assay to determine how well the new substrates are accepted by the downstream C-domain.

It has previously been shown that C-domains can also use SNAC thioesters as the upstream acyl donor.[289] If it proves possible to use the above assay with SNAC thioesters as both the donor and acceptor, we are then equipped with a method of high-throughput screening that could be used in directed evolution of C-domains. In the future, the combination of directed evolution of A-domains and their upstream C-domains would potentially allow any amino acid in any NRPS-produced peptide to be changed and would be a very powerful technique in combinatorial biosynthesis.

5

Materials and Methods

5.1 General materials and instruments

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich Chemical Company and were used as received. All chemicals used for the production of buffers were of analytical grade. All solvents and chemicals used were of HPLC or analytical grade. All gases were supplied by BOC with the exception of 19% oxygen-18 (98% isotopic purity) in high purity nitrogen which was purchased from CK Gas Products. Aqueous solutions and buffers were prepared using sterile deionized water from Milli-Q Plus systems (Millipore) equipped with a 0.22 μm filter at 20 $^{\circ}$C. Luciferin was from Apollo Scientific Ltd and ^{18}O-tyrosine from Isotech, USA.

The plasmid containing the gene for 4-hydroxymandelic acid synthetase was kindly donated by Dr. F. Huang, the plasmid containing the gene for EntE cloned from *E. coli* K12 was donated by Dr. N. M. Llewellyn, and the plasmid containing the gene for the SA1 mutant was obtained from Shilo Dickens at the Deparment of Biochemistry, Cambridge. The genomic DNA for *E. coli* K12 was donated by Dr. N. M. Llewellyn. The His_6-Luciferase-$\times$5 plasmid was donated by Dr. E. Law at the Deparment of Biotechnology, University of Cambridge.

5.1.1 Instrumentation

To maintain a sterilised environment, all disposable equipment (micro-centrifuge tubes, pipette tips, etc.) were autoclaved at a temperature of 120 °C for 30 minutes and the work was carried out within a MDH Microflow laminar flow hood.

Balances : AND HF-300G and OHAUS®Analytical Standard; BS Services
pH Meter: Corning 200 pH meter, Mettler Toledo pH electrode
Vortexer: Vortex Genie 2; Scientific Industries
Heat block; Dri-Block®DB-2D; Techne
Centrifuges: Biofuge *pico*, Biofuge Stratos, Sorvall®RC 5C Plus; Heraeus Instruments
UV-Visible spectrophotometer: CARY 100 Bio; Varian
Circular Dichroism Spectrometer: Applied Photophysics Chirascan
Plate Reader: FLASHScan 450; Flashscan, Sprectramax Plus; Molecular Biosciences
FPLC: Akta FPLC Explorer; GE Healthcare
Sonicator: Vibra-Cell^{TM}; Sonics & Materials, Inc.
Homogeniser: Micro Modulyo; Edwards
Electroporator: MicroPulser^{TM}; Bio-Rad Corporation
High Resolution Mass Spectrometer: LTQ Orbitrap Discovery^{TM}; Thermo Fisher Scientific (operated by Dr. E. Stephens and Dr. M. Tosin)

5.2 Synthetic chemistry methods

5.2.1 Synthesis of (*S*)- or (*R*)-methyl mandelate

To a solution of conc. HCl(aq) in MeOH (1:9 v/v) (2.5 ml) was added D- or L-mandelic acid (500 mg, 3.3 mmol). After 2 hours at room temperature toluene as an azeotrope was added and the solution evaporated either at reduced pressure on a

Buchi rotary evaporator or under a stream of nitrogen to yield the esters (530-540 mg, 96–99%). Spectral data were consistent with those reported.[147] ^{1}H NMR (400 MHz, $CDCl_3$) δ 3.3 (1H, d, J = 7.5 Hz), 3.7 (3H, s), 5.2 (1H, d, J = 7.5 Hz), 7.4 (5H, m). $[\alpha]_D$ = +179.2° ($CHCl_3$, 25 °C), (S) and $[\alpha]_D$ = -182.1° ($CHCl_3$, 25 °C), (R).

5.2.2 Synthesis of NTBC

HOBt-6-carboxamidomethylpolystyrene (66.8 mg), PyBop (62.3 mg) and diisopropylethylamine (31.7 μL) were dissolved in dry DCM (2 mL) and allowed to equilibrate overnight. The resin was then washed with dry DCM on a sinter funnel before being re-suspended in DCM. To this suspension 1,3-cyclohexadione (6.8 mg), triethylamine (6.06 mg) and 2-nitro-4-(trifluoromethyl)benzoic acid (286 mg) were added and the mixture was stirred overnight at room temperature. The resin was removed by filtration on a sinter funnel and discarded. The filtrate was purified by flash chromatography on silica gel (ethyl acetate-DCM 1:1) and one major product (120 mg, 32%) was isolated. The ^{1}H NMR spectrum of this product showed that it was not NTBC, it was the isomer coupled *via* the oxygen on the diketone. ^{1}H NMR (500 MHz, $CDCl_3$) δ: 2.15 (2H, qn, J = 7.1 Hz), 2.45 (2H, t, J = 7.1 Hz), 2.68 (2H, t, J = 7.1 Hz), 6.07 (1H, s), 7.93 (1H, d, J = 7.4), 8.06 (1H, d, J = 7.4), 8.4 (1H, s).

5.2.3 Synthesis of 1,3,6,8-tetrahydroxynaphthalene (THN)[217]

Synthesis of methyl 3,5-dibenzyloxyphenylacetic acid: Potassium carbonate (20.6 mg) and benzyl bromide (1.5 mL) were added to an anhydrous acetone (8 mL) solution of methyl 3,5-dihydroxyphenylacetate (1 g). The mixture was heated at reflux for 12 hours. The precipitate was removed by filtration and the filtrate was evaporated. The residue was then purified by flash chromatography on silica gel (chloroform-hexane 4:1) to give the dibenzyl ether as a white powder. Spectroscopic data were consistent with those reported).[217] ^{1}H NMR (500 MHz, $CDCl_3$) δ: 3.56

(–C$\underline{H}_2$COOCH$_3$, 2H, s), 3.68 (–OCH$_3$, 3H, s), 5.02 (2 × –C$\underline{H}_2$C$_6$H$_5$, 4H, s), 6.54 (C$_6$H$_3$, 3H, s), 7.3–7.4 (2 × C$_6$H$_5$, 10H, m).

Synthesis of methyl curvulinate: Methyl 3-5-dibenzyloxyphenylacetic acid (50 mg) was dissolved in acetic acid (60 μL) at 0 °C. Trifluoroacetic anhydride (350 μL) was added to the mixture and it was allowed to warm to room temperature. The mixture was then left to stir for 1 hour 45 mins. Saturated aqueous NaHCO$_3$ (1 mL) was then added and the mixture was extracted with ethyl acetate. The combined organic layers were dried over MgSO$_4$ and evaporated under reduced pressure to give a white powder. This was dissolved in ethanol (3 mL), 10% Pd/C (10 mg) was added and the mixture was stirred under an atmosphere of hydrogen for 15 minutes. The Pd/C was removed by filtration through Celite and the solvent removed *in vacuo* to give a white powder. Spectroscopic data were consistent with those reported.[217] ^{1}H NMR (500 MHz, acetone-d_3) δ: 2.53 (CH$_3$CO–, 3H, s), 3.63 (OCH$_3$, 3H, s), 3.79 (–C$\underline{H}_2$COOCH$_3$, 2H, s), 6.33 (4-H, 1H, d, J =2 .3 Hz), 6.37 (6-H, 1H, d, J = 2.3 Hz), 9–10 (-OH, 1H, br s). ^{13}C-NMR(125 MHz, Acetone-d_3) δ: 32.6, 40.9, 52.4, 103.2, 112.8, 119.2, 138.3, 161.9, 172.6, 204.0.

Synthesis of 1,3,6,8-tetrahydroxynaphthalene Metallic sodium was reacted with d_4-methanol (2 mL) to form a solution of CD$_3$ONa (2.0 mmol) in CD$_3$OD. Methyl curvulinate (50 mg) was added and the mixture was stirred at 60 °C under argon for 15 minutes. 0.6 mL of the product solution was then submitted for NMR measurements in a sealed tube under an atmosphere of argon. Spectroscopic data were consistent with those reported).[217] ^{1}H NMR (500 MHz, CD$_3$OD-d_6) δ 6.23 (2-H and 7-H, 2H, d, J = 2.1), 6.45 (4-H and 5-H, 2H, d, J = 2.1).

5.2.4 Synthesis of [2,3,5,6-$^{2}H_4$,^{18}O]4-hydroxymandelic acid

[2,3,4,5,6-$^{2}H_5$]Benezenediazonium tetrafluoroborate: 98% [U-$^{2}H_5$]aniline (0.91 mL, 9.98 mmol, Cambridge Isotopes) was dissolved in a mixture of distilled water (5 ml) and 50% aqueous fluoroboric acid (3.4 mL, Avocado Chemicals). The solution was cooled to 0 °C and a solution of sodium nitrite (0.69 g, 10 mmol) in water (1.5 mL) was added dropwise. The suspension was stirred, keeping the temperature between 0 and 4 °C for 30 min, and then filtered. The residue was recrystallised from acetone/diethyl ether to give the benzenediazonium salt as a white powder (1.69 g, 86%).

Synthesis of [2,3,4,5,6-$^{2}H_5$,^{18}O]phenol:[290] To a stirred paste of [$^{2}H_5$]benzenediazonium tetrafluoroborate (1.0 g, 5.1 mmol) in 97% [^{18}O]water (1 mL, Cambridge Isotopes), concentrated sulphuric acid (125 μL) was added. The mixture was then heated to 65 °C until evolution of nitrogen ceased. The solution was extracted with diethyl ether multiple times. The ethereal layer was washed with 1 M HCl and brine, dried over magnesium sulphate, and concentrated under reduced pressure. The crystalline residue was purified by sublimation to give the labelled phenol as colourless crystals (134 mg, 26%). The final oxygen-18 enrichment was found by GC-MS to be at least 80%.

Synthesis of [2,3,5,6-$^{2}H_4$,^{18}O]4-hydroxymandelic acid: [$^{2}H_5$,^{18}O]Phenol (94.1 mg, 0.93 mmol) was added to a solution of sodium hydroxide (8 mg) in distilled water (0.8 mL), followed by the dropwise addition of a solution of glyoxylic acid (92 mg, 1.24 mmol) in distilled water (100 μL). The mixture was stirred for 5 h at 30 °C, then adjusted to pH 6.0 by the addition of hydrochloric acid and washed with toluene to remove unreacted phenol. The aqueous layer was then adjusted to pH 1.5 and the product was extracted into diethyl ether by the use of continuous

liquid/liquid extraction overnight. Evaporation of the solvent under reduced pressure gave the labelled 4-hydroxymandelic acid (146 mg, 90%) as a white powder. The oxygen-18 enrichment was found by mass spectrometry to be 87%. HRMS (ESI) calculated for $C_8D_4H_4O_3{}^{18}O^{32}Na$ ([M+Na]$^+$): 197.0604. Found: 197.0608 (+2.0 ppm).^{1}H NMR (500 MHz, D_2O): δ 5.20 (s). ^{13}C NMR (500 MHz, D_2O): δ: 176.9 (C=O), 155.8 (C-OH), 130.1 (C), 128.4 (CD, t, J_{C-D} = 24 Hz), 115.3 (CD, t, J_{C-D} = 24 Hz), 72.5 (CH).

[2,3,5,6-^{2}H]4-hydroxymandelic acid was produced by the same method, except distilled water was used in place of [^{18}O]H_2O during the formation of the labelled phenol.

5.2.5 Synthesis of *N*-acetylcysteamine thioesters of benzoic acids

***S*-2-Acetamidoethyl 2,3-Dihydroxybenzothioate (2,3DHB-SNAC):** 2,3-Dihydroxybenzoic acid (308 mg) and hydroxybenzotriazole hydrate (HOBt) (306 mg) were dissolved in 4 mL of dry tetrahydrofuran (THF). *N*-Acetylcysteamine (SNAC) (266 μL) was added to the stirred solution, followed by *N,N'*-dicyclohexylcarbodiimide (DCC) (413 mg). After 1 hour at 23 °C, solid potassium carbonate (132 mg) was added and stirring was continued for another hour. Ethyl acetate (5 mL) was added, the mixture filtered and the filtrate concentrated to dryness under reduced pressure. The residue was dissolved in ethyl acetate, washed twice with 10% sodium bicarbonate, dried over sodium sulphate, and concentrated to a white solid. The product was purified by dissolving in DMSO followed by preparative HPLC (20–40% acetonitrile in water over 20 min, 15 mL/min). The final purified product was isolated as a white solid (yield = 61%). ^{1}H NMR (500 MHz, DMSO) δ: 1.80 (3H, s), 3.08 (2H, t, J = 6.4 Hz), 3.27 (2H, q, J = 6.2 Hz), 6.77 (1H, t, J = 7.8 Hz), 7.03 (1H, dd, J = 8.0 and 1.4 Hz), 7.23 (1H, dd, J = 8.0 and 1.4 Hz), 8.11 (1H, t, J = 5.3 Hz), 9.68 (1H, s), 10.20 (s, 1H). HRMS calculated for $C_{11}H_{13}NNaO_4S^+$ ([M+Na]$^+$): 278.063. Found:

278.0464 (0.4 ppm).

***S*-2-Acetamidoethyl benzothioate:** The same procedure was used as with 2,3DHB-SNAC, yield = 21%. ^{1}H NMR (500 MHz, DMSO-d_6) δ: 1.81 (3H, s), 3.13 (2H, t, *J* = 6.5 Hz), 3.29 (2H, td, *J* = 6.5 Hz and 5.0 Hz), 7.29 (2H, dd, *J* = 7.5 and 1.5 Hz), 7.57 (2H, t, *J* = 7.5 Hz), 7.706 (1H, tt, *J* = 1.5 and 7.5 Hz), 8.11 (1H, t, *J* = 5.0 Hz). HRMS calculated for $C_{11}H_{14}NO_2S^+$ ($[M+H]^+$): 224.0748. Found: 224.0745 (1.3 ppm).

***S*-2-Acetamidoethyl 2,4-dihydroxybenzothioate (2,4DHB-SNAC):** The same procedure was used as with 2,3DHB-SNAC, yield = 44%. ^{1}H NMR (500 MHz, DMSO-d_6) δ:1.86 (1H, s), 3.11 (2H, t, *J* = 6.5 Hz), 3.30 (2H, q, *J* = 6.4 Hz), 6.36 (1H, d, *J* = 2 Hz), 6.45 (dd, *J* = 2.5 and 9.0 Hz), 7.74 (1H, d, *J* = 9.0 Hz), 8.14 (1H, t, *J* = 5.5 Hz), 10.63 (1H, s), 10.961 (1H, s). HRMS calculated for $C_{11}H_{13}NNaO_4S^+$ ($[M+Na]^+$): 278.0466 Found: 278.0464 (0.4 ppm).

***S*-2-Acetamidoethyl 2,5-dihydroxybenzothioate (2,4DHB-SNAC):** The same procedure was used as with 2,3DHB-SNAC, yield = 37%. ^{1}H NMR (500 MHz, DMSO-d_6) δ: 1.81 (3H, s), 3.06 (2H, t, *J* = 7.0 Hz), 3.27 (2H, q, *J* = 6.5 Hz), 6.83 (1H, D, *J* = 9 Hz), 6.94 (1H, dd, *J* = 3.5 and 9 Hz), 7.15 (1H, D, *J* = 3.5 Hz), 8.09 (1H, t, *J* = 5.5), 9.25 (1H, s), 10.05 (1H,s) HRMS calculated for $C_{11}H_{13}NNaO_4S^+$ ($[M+Na]^+$): 278.0466. Found: 278.0463 (1.1 ppm).

***S*-2-Acetamidoethyl 2,3,4-trihydroxybenzothioate (2,3,4-THB-SNAC):** The same procedure was used as with 2,3DHB-SNAC, yield = 27%. ^{1}H NMR (500 MHz, DMSO-d_6) δ: 1.81 (3 H, s), 3.07 (2H, t, *J* = 7.5 Hz), 3.27 (2H, q (dt), *J* = 5.5 Hz), 6.55 (1H, d, *J* = 9.1 Hz), 7.25 (1H, d, *J* = 9.1 Hz), 8.97 (1H, t, *J* = 5.5 Hz). HRMS calculated for $C_{11}H_{14}NO_5S^+$ ($[M+H]^+$): 272.0587. Found: 272.0589 (0.7 ppm).

***S*-2-Acetamidoethyl 2-hydroxybenzothioate:** The same procedure was used as with 2,3DHB-SNAC, yield = 26%. ^{1}H NMR (500 MHz, DMSO-d_6) δ: 1.91 (3H, s), 3.37 (2H, dt, *J* = 5.5 and 6.0 Hz), 3.37 (2H, t, *J* = 6.0 Hz). 7.09 (1H, dd, *J* = 1.0 and 8 Hz), 7.5 (1H, ddd, *J* = 8.0 and 7.5 and 1.5 Hz), 7.63 (1H, ddd, *J* = 8.0 and 7.5 and 1 Hz), 8.2 (1H, t, *J* = 5.5 Hz). HRMS calculated for $C_{11}H_{14}NO_3S^+$ ($[M+H]^+$): 242.2982. Found: 242.2984 (0.8 ppm).

5.3 Analysis of synthesised labelled compounds

5.3.1 GC-MS analysis of labelled compounds

Analysis of labelled phenol was performed on a Perkin Elmer Autosystem XL GC connected to a Perkin Elmer Turbomass Spectrometer. Helium was used as a carrier gas and detection was carried out via electron impact ionisation The column used was a Perkin Elmer Elite-Pe-5MS. The method used was to start at 40 °C, which was held for 1 minute before heating to 230 °C at a rate of 10 °C min^{-1}. The column was then held at 230 °C for 5 min.

5.4 Molecular biology methods

5.4.1 Library screening in microplates of HmaS/HppD mutants

Libaries were transformed into *Escherichia coli* BL21-CodonPlus®-RP (Stratagene). Cultures were grown in 1 ml LB-tyrosine in Deepwell microplates (Fisher). LB-tyrosine was prepared from LB (1 litre) by the addition of 1 g of tyrosine, dissolved in 20 ml of 1 M HCl, followed by 100 ml of 1 M disodium orthophosphate and 2 ml of 30 mg/ml kanamycin sulphate. After 48 hours growth at 30 °C cells were harvested by centrifugation and 100 μl of the supernatant was transferred to a UV-transparent microplate (Grenier). The absorbance was measured at 280–400 nm, at 10 nm intervals using a Spectramax Plus spectrophotometer running Softmax

(Molecular Biosciences). The phenotypes of the clones were classified as HmaS wild type-like ($A_{330} > 2.0$, $A_{430} < 0.1$, no visible pigmentation), Hma producer ($A_{330} > 0.3$, $A_{430} < 0.1$, no pigmentation), HppD wild type-like ($A_{330} > 2.0$, $A_{430} \approx 0.1$, dark pigmentation), Homogentisate producer ($A_{330} > 0.3$, $A_{430} > 0.1$, brown pigmentation) or inactive ($A_{330} < 0.3$). Plasmids of the genes of interest were prepared from cell pellet by standard methods. To sample the range of amino acid substitutions that are tolerated by the enzymes, the genes of six wild type-like clones from each single-residue library were sequenced. Only one gene contained a PCR-induced mutation outside the active site: HmaS H106R T214V.

5.4.2 Expression and purification of His_6-tagged proteins

A single colony of *E. coli* BL21 (DE3) harbouring the desired recombinant plasmid was picked from the LB/agar plate and inoculated into 10 mL of LB/Kan (LB supplemented with 50 μg/ml kanamycin). The culture was incubated overnight at 37 °C with shaking at 250 rpm. A 1 mL aliquot of the culture was diluted into 1 L liquid LB/Kan and incubated at 37 °C with shaking at 250 rpm until the observed OD_{600} reached 0.6-1.0. The culture was induced by the addition of 200 μL of 1 M IPTG stock solution and incubated overnight at 16 °C with shaking at 250 rpm.

All purification steps were performed at 4 °C. Induced cells were pelleted by centrifugation at 7,000 rpm for 10 minutes. Cell pellets were resuspended in 30 mL of binding buffer and lysed by either ultrasonication or homogenisation. Cell lysates were centrifuged at 18,000 rpm for 45 minutes to pellet the insoluble fraction. Clear cell lysates were passed through a column (1 cm diameter) containing 1 mL His-Bind^{TM} resin (Novagen), which was previously washed with 10 mL deionized water, charged with 10 mL of 50 mM $NiCl_2$ soluton, and equilibrated with 15 mL of binding buffer. Bound protein was washed with 10 mL binding buffer followed by 5 mL washing buffer and eluted with 4 mL elution buffer. Eluted protein was

exchanged into the appropriate storage buffer by ultrafiltration with an Amicon Ultra-4 concentrator. Proteins were stored at -20 °C.

Binding Buffer:	20 mM Tris/HCl, 0.5 M NaCl, 5 mM imidazole, 10% glycerol, pH 7.9.
Washing Buffer:	20 mM Tris/HCl, 0.5 M NaCl, 60 mM imidazole, 10% glycerol, pH7.9
Elution Buffer:	20 mM Tris/HCl, 0.5 M NaCl, 200 mM imidazole, 20% glycerol, pH7.9
Storage Buffer (HmaS and SA1):	200 mM Tris, 50 mM NaCl, pH 7.0, 20% glycerol
Storage Buffer (EntE):	500 mM Tris, pH 7.4, 50% glycerol.
Storage Buffer (Sfp):	50 mM HEPES, 200 mM NCl, 50% glycerol, pH 7.6

5.4.3 FPLC protein purification

Proteins were futher purified by affinity chromatography, as required, by the use of fast protein liquid chromatography. The eluted protein from the His-Bind™ resin (Novagen) was purified on an AKTA Explorer (GE Healthcare) attached to an FPLC HR 16/60 HiLoad Superdex 200 prep grade column at pH 6.9. The column was pre-equilibrated with 20 mM Tris buffer at pH 7.5 containing 100 mM KCl and the elution of protein from the column monitored at 280 nm. The eluate was fractionated automatically into 1 mL samples. The presence of the target protein in fractions was confirmed by SDS-PAGE gel, and these were then combined. The protein was either immediately used for crystallisation trials or combined with glycerol to give a final concentration of 50% and stored at -20 °C until required.

5.4.4 Sodium dodecylsulphate polyacrylamide gel electrophoresis

Proteins were analysed for purity by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE) on a vertical polyacrylamide gel prepared on a Hoefer Mighty Small 245 Dual gel caster according to standard protocols.[291] Stacking and resolving gels were prepared with final acrylamide concentration of 5% and 10% respectively. Protein samples were mixed 1:1 with SDS loading buffer and incubated at 95 °C for 5 minutes prior to loading on the gels. Gels were run in Tris-glycine buffer at 100 V, 20 mA (per gel), for 50 minutes. Protein molecular weight markers (MBI-Fermentas) contained the following reference standards: β-galactosidase (116.0 kDa); bovine serum albumin (66.2 kDa); ovalbumin (45.0 kDa); lactate dehydrogenase (35.0 kDa); Bsp98I (25.0 kDa); β-lactoglobulin (18.4 kDa); lysozyme (14.4 kDa). Run gels were stained for 1 hour with staining solution, followed by 1 hour in de-staining solution, and imaged on a UVI-Doc ultraviolet transilluminator.

Tris-glycine buffer:	25 mM Tris, 250 mM glycine, 0.1% SDS, pH 8.3
SDS Loading buffer:	50 mM Tris/HCL, 100 mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol, pH 6.8
Staining solution:	2.5 g/L Coomassie Brilliant Blue in MeOH:H_2O:acetic acid (4.5:4.5:1)
Destaining solution:	MeOH:H_2O:Acetic acid (4.5:4.5:1)

5.4.5 Determination of protein concentration

Protein concentration was determined spectrophotometrically by the Bradford method.[292] Bradford reagent (900 μL, Bio-Rad) was mixed with 1-10 μL of protein solution in a total volume of 1 mL and incubated at room temperature for 5 minutes. The absorbance at 595 nm was observed and compared against a standard curve generated

using bovine serum albumin to determine the concentration of the protein.

5.4.6 Growth of *A. orientalis*

A 250 mL flask with spring containing common streptomyces media (CSM) (50 mL), was inoculated with *Amycolatopsis orientalis* (A82846.2) from a frozen permanent stock and grown for 3 days at 30 °C with shaking at 300 rpm. 5 mL of this starter culture was then transferred to FermB media (50 mL, 30 °C, 300 rpm). The culture was fed [$^{2}H_{4}{}^{18}O$]4-hydroxymandelic acid (10 mg) in water (10 mL) adjusted to pH 7.0 with NaOH (0.1 M) in 4 portions over a period between 36-48 hours of growth. The culture was harvested at 110 hours and purified as described below.

CSM: 30 g/L tryptic soy broth, 3 g/L bacto-yeast extract, 2 g/L magnesium sulphate, 5 g/L glucose, anhydrous, 4g maltose, pH 7.0

FermB: 40 g/L anhydrous glucose, 80 g/L potato dextrin, 40 g/L bacteriological peptone, 8 g/L beet molasses, and 4 g/L calcium carbonate, pH 7.0

5.4.7 Isolation and purification of glycopeptides

The glycopeptide-containing media was purified by the use of a D-alanine-D-alanine-based affinity column. The column was synthesised as follows: CH-Sepharose 4B (2 g) was added to hydrochloric acid (100 mL, 1 mM) and allowed to swell before washing on a glass sinter for 10 minutes with hydrochloric acid (400 mL, 1 mM). The D-alanine-D-alanine (63.8 mg, 0.40 mmol) dipeptide was dissolved into coupling buffer before being combined with the activated resin and rotated end-over-end for 90 minutes. The resin was extracted by filtration and washed again with coupling buffer (200 mL), then re-suspended in aqueous ethanolamine hydrochloride (25 mL, 1 M, pH 8.0). After 60 minutes, the suspension was filtered and washed with 5

cycles of 0.1 M sodium acetate/0.5 M NaCl (150 mL, pH 4.0) followed by 0.1 M Tris HCl/0.5 M NaCl (150 mL, pH 8.0) alternately, and finally left to equilibrate with sodium phosphate buffer (0.02 M, pH 7.0) before use.

The production culture was pelleted by centrifugation (6470 rpm, 5000 g) and the cell-free supernatant was filtered through a GF/A glass fiber filter under suction. The filtrate was adjusted to pH 7.0 (where necessary) and loaded onto the D-alanine-D-alanine column using a peristaltic pump at a flow rate of 0.5 mL min^{-1}. The column was washed sequentially with aqueous sodium phosphate (40 mL, 0.2 M, pH 7.0), aqueous ammonium acetate (40 mL, 0.4 M, pH 7.8), and then 10% acetonitrile in water (40 mL). The glycopeptide was eluted with 0.1 M NH_4OH:acetonitrile (1:1; 40 mL), and the eluate was lyophilised.

The column was regenerated by washing with 0.4 M aqueous sodium carbonate/30% acetonitrile (80 mL, pH 11.0), and was then re-equilibrated with sodium phosphate buffer (50 mL, 0.02 M, pH 7.0).

Coupling buffer: 0.1 M $NaHCO_3$/ 0.5 M NaCl, 12 mL, pH 8.0

5.4.8 Polymerase chain reaction from *E. coli* gDNA

PCR reactions were carried out in a Minicycler™ thermal cycler (MJ Research). KOD Hot Start DNA Polymerase (Novagen) was used for amplification of target DNA. A 50 μL reaction was prepared containing 1μL *E. coli* DH10B gDNA. 1.25 μL each of primers ENTB_PCP and ENTB_PCP_REV, 5 μL of 10$\times$ KOD buffer, 5 μL of 2 mM dNTPs, 3 μL of 25 mM $MgCl_2$, 3 μL of DMSO, and 1 μL of KOD polymerase. The PCR reaction was performed under the following program: 2 minutes at 95 °C followed by 15 seconds at 95 °C; 15 seconds at 95 °C; 30 seconds at 63 °C and 5.5 minutes at 72 °C for 30 cycles; then a final extension for 10 minutes at 72 °C; hold at 4 °C.

ENTB_PCP: GGA CGT TCT GGC CAT ATG GTG ATG ACT GAA G

ENTB_PCP_REV: CCG CTG AAT TCC AT TAT TTC ACC TCG CG

5.4.9 Agarose electrophoresis gels

Electrophoresis-grade agarose was dissolved in TAE buffer to a concentration of 1% with microwave heating. The agarose gel was poured into a horizontal caster, and electrophoresis carried out in TAE buffer in a Hoefer mini horizontal submarine unit (Amersham Pharmacia Biotech) at room temperature. DNA samples were mixed 1:1 with 2× DNA loading dye (MBI-Fermentas) prior to loading on the gel. Gels were run at 100 V, 500mA for 1 hour. GeneRuler™ 1 kb DNA ladder (MBI-fermentas) contained fragments of the following sizes: 0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, 8.0 and 10.0 kb. Resolved gels were imaged on a UVI-Doc ultraviolet transilluminator. For DNA purification, the desired band was excised and extracted using the AccuPrep® Gel Purification Kit (Bioneer) according to the manufacture's protocol)

TAE Buffer: 40 mM Tris-acetate, 1mM EDTA

5.4.10 Construction of EntB-PCP gene within pET-28a(+)-Based Plasmid DNA

The purified 95 bp PCR product (20 μL) was digested simultaneously by the restriction enzymes NdeI and EcoR1 (New England Biolabs) in 30 μL reaction containing 3 μL 10× EcoR1 Buffer, 1 μL NdeI and 1 μL EcoR1 and the volume made up with MQ water. The reaction was incubated at 37 °C for 4 hours and purified by agarose gel electrophoresis. The *E. coli* vector pET-28a(+) (Novagen) was digested by the same restriction enzymes and purified by agarose gel electrophoresis. The digested PCR fragment (2 μL) and digested pET28a(+) (1 μL) were ligated in a 20 μL reaction containing 2 μL of 10× T4 DNA ligase buffer (New England Biolabs), 2 μL of 1 mM

ATP, and 1 μL of t4 DNA ligase incubated overnight at 4 °C.

5.4.11 Transformation and isolation of ligated DNA

E. coli NovaBlue competent cells (Novagen) were thawed on ice. The NovaBlue cells were allowed not to defrost for longer than 5 minutes on ice after being removed from storage at -80 °C. To a 20 μL suspension of cells was added 10 μL of plasmid stock solution. The solution was kept on ice for 10 minutes, heat-shocked at 42 °C for 40 s and subsequently returned to ice for 5 minutes. Transformed cells were suspended in SOC medium (300 μL) and incubated at 37 °C for 1 hour prior to spreading on LB/agar/Kan plates (LB/agar supplemented with 50 μg/ml kanamycin). After overnight incubation at 37 °C, single colonies were picked from the plate into 5 mL liquid LB/Kan and incubated overnight at 37 °C with shaking at 250 rpm. The recombinant plasmid was purified from the culture using AccuPrep® Plasmid Mini Extraction Kit (Bioneer) according to the manufacturer's protocol.

SOC Medium: 20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 20 mM glucose, pH 7.0

5.4.12 Transformation of and isolation of pET-28a(+)-based plasmid DNA

E. coli competent cells (Novagen) were thawed on ice. To a 10 μL suspension of cells was added 0.5-1 μL of plasmid stock solution. The mixture was kept on ice for 10 minutes, heat-shocked at 42 °C for 45 seconds, and returned to ice for 5 minutes. Transformed cells were suspended in 100 μL of SOC medium and incubated at 37 °C for 1 hour prior to spreading on LB/Agar/Kan plates.

5.4.13 Expression and purification of proteins in 96-well format

E. coli BL21 cells containing the required vector were grown overnight in a 96-well

plate (Nunc) at 37 °C and 250 rpm in 1 mL LB/Kan. 150 μL of each overnight culture was used to inoculate 1 mL of LB/Kan in a 96-well plate. The culture was then grown in the same conditions as overnight culture until it reached an OD_600 = 0.6–1.0 before inducing with 200 μL LB containing IPTG to give a final concentration of 0.5 mM. Expression was carried out overnight at 16 °C and 250 rpm.

Cells were lysed using BugBuster Master Mix as directed by the user manual. The product was centrifuged (10 min, 4000 rpm, 4 °C) to obtained the cleared cells lysate.

The cell lystate was purified using a 96-well His-bind column (TALON, Clontech) using a vacuum manifold. 1.2 mL of cell lysate was purified as directed by the user manual. The protein was eluted with elution buffer (see section) and the resin regenerated.

5.4.14 Site-directed saturation mutagenesis

Site-directed mutagenesis reactions were carried out in a Minicycler™ thermal cycler (MJ Research). KOD Hot Start DNA Polymerase (Novagen) was used to catalyse polymerisation. A 50 μL reaction was prepared containing 1 μL of the pET28a(+) vector containing the EntE gene, 10 μL of both degenerate primers (Operon), 5 μL 10× KOD buffer, 5 μL 2 mM dNTPs, 3 μL 25 mM $MgCl_2$, 3 μL DMSO and 1 μL KOD polymerase. The PCR reaction was performed under the following program: 2 minutes at 94 °C; then 30 seconds at 94 °C, 30 seconds at 60 °C, and 15 minutes at 72 °C, for 30 cycles. then a final extension of 10 minutes at 72 °C,; then hold at 4 °C.

Primers:

ENTE_G308:	GCTGAAACTGTTACAGGTCNNKGGCGCACGTCTTTCTGCC
ENTE_G308_R:	GGCAGAAAGACGTGCGCCMNNGACCTGTAACAGTTTCAGC
ENTE_V331:	GGCTGTCAGTTGCAGCAGNNKTTTGGCATGGCGGAAGGGCTG
ENTE_V331_R:	CAGCCCTTCCGCCATGCCAAAMNNCTGCTGCAACTGACAGCC
ENTE_V339:	GGCATGGCGGAAGGGCTGNNKAACTACACCCGACTTGATG
ENTE_V339_R:	CATCAAGTCGGGTGTAGTTMNNCAGCCCTTCCGCCATGCC

5.4.15 Transformation and isolation of DNA after site-directed mutagenesis

Template DNA in the site-directed mutagenesis reaction was digested by incubation with 1μL of Dpn1 (10U/μL, New England Biolabs) at 37 °C for 1 hour. The product was then transformed and plasmid DNA isolated by the method in section 5.4.11. The purified DNA was then sequenced to confirm the presence of the desired mutation.

5.5 Bioinformatics methods

5.5.1 Protein database homology searches

Protein database homology searches and conserved domain homology searches were performed using the BLAST2 algorithm hosted at the NCBI website.

5.5.2 Protein structural prediction

Protein structural prediction was performed by the use of the Phyre2 algorithm[284] hosted at the Imperial College London website (http://www.sbg.bio.ic.ac.uk/phyre/).

5.6 Analysis of enzymatic activities

5.6.1 LC-ESI-MS analysis of proteins

Proteins were analysed by LC-ESI-MS (helium collision gas) using a Finnigan LCQ ion-trap mass spectrometer (Thermo Electron) coupled to an HP1100 HPLC sytem (Agilent). Proteins were separated on a 2 $\times$ 250 mm Jupiter 5μM C4 column (Phenomenex) with the following gradient at a flow rate of 0.2 mL/min at 40 °C: 0 – 5 minutes, 5% B; 5 – 6 minutes, 5% – 35%; 6 – 31 minutes; 35% – 65% B, 31 – 36 minutes, 65% B; 35 – 41 minutes, 65% – 95% B (buffer A: 0.1% TFA in H_2O; buffer B: 0.1% TFA in acetonitrile). Absorbance was monitored at 214 nm and 254 nm. Mass spectra were acquired between 200 and 2000 Da.

5.6.2 HmaS/SA1 activity assay

The activity of HmaS/SA1 was assayed by incubating 10 μL enzyme (100 μg) in a 500 μL reaction containing: 2.0 μL of 1.0 M $FeSO_4$; 20.0 μL of 520 mM ascorbic acid pH 7; 8.0 μL of 1 M DTT; 3.0 μL of 600 mM substrate (4-hydroxyphenylpyruvic acid) in Buffer C. The reaction was incubated for 2 hours at 30 °C. The enzyme was then precipitated out with 300 μL of 1M HCl (aq.), spun down with a microcentrifuge and the supernatant extracted with ether. The organic extract was then dried under flow of nitrogen to give a white residue and then derivatised by dissolving in methanol:conc. HCl(aq) (9:1 v/v) and standing for 2 hours. The reaction mixture was then evaporated under a stream of nitrogen and dissolved in ether for analysis.

Analysis of the derivatised assay products was performed by GC-MS. The column temperatures used were 40 °C for 2 minutes, 40-250 °C at 10 °C/min, 250-300 °C at 25 °C/min, 300 °C for 5 minutes. This method was altered for the chiral column so that it was capped at a temperature of 220°C as at higher temperatures the column is damaged, but otherwise remained the same.

Buffer C: 250 mM Tris-HCl/maleic acid, pH 6.9

5.6.3 LC-ESI-MS analysis of glycopeptides

Initial LC-ESI-MS (helium collision gas) using a Finnigan LCQ ion-trap mass spectrometer (Thermo Electron) coupled to an HP1100 HPLC system (Agilent): samples were loaded onto a Phenomenex Luna C18(2) column (250 x 2 mm, 5 μm), eluting with a linear gradient of 0 to 60% acetonitrile (0.1% trifluoroacetic acid) in water (0.1% trifluoroacetic acid) over 30 min with a flow-rate of 0.3 mL min^{-1}. High resolution HPLC-ESI-MS was performed on a Thermo Electron LTQ-Orbitrap. Samples were injected onto a Dionex Acclaim PepMap 100 column (C18, 3μM, 100 Å), eluting with a linear gradient of 0 to 100% acetonitrile (0.1% formic acid) in water (0.1% formic acid) over 28 min with a flow rate of 50 μL min^{-1}. Chloroeremomcyin, eremomycin and dechloroeremomycin were all identified by high-resolution accurate mass within 3 ppm of the predicted mass.

5.6.4 Luciferase analysis of the forward EntE reaction

The turnover of ATP by EntE was assayed in a 100 μL reaction containing: 100 μM ATP; 27 μM substrate, 500 nM EntE in TEM buffer. The reaction was incubated for 1.5 hours at 37 °C. The concentration of ATP was then measured by the photon emission from the reaction with luciferase (as in section 5.6.6).

TEM Buffer: 0.1 M Tris/Acetate, 10 mM $MgSO_4$, 2mM EDTA, pH 7.8

5.6.5 Luciferase analysis of the reverse EntE reaction

The generation of ATP by EntE from incubation with its products was assayed in a 100 μL reaction containing 100 μM ATP, 100 μM PP_i, 23 μM substrate-SNAC, 500 nM EntE in TEM buffer and incubated 37 °C for 1.5 hours. The concentration of

ATP was then measured by the photon emission from the reaction with luciferase (as in section 5.6.6).

5.6.6 Measurement of ATP concentration with luciferase

The measurement of ATP concentration in an assay sample was carried out on 20 μL of the assay reaction mixture combined with 20 μL of luciferase (500 nM) in TEM Buffer. This mixture was then analysed by a photometric assay using a B-systems Luminoskan Ascent Luminometer. The device injected 200 μL of 200 μM D-Luciferin in TEM buffer, immediately before detection. The emission of light was then recorded by photo-multipler tube (PMT). The PMT voltage was set to give the largest signal strength without saturating the detector.

5.6.7 Measurement of AMP usage by coupling to NADH

A-domain enzymatic activity was determined by the monitoring of NADH consumption. The reactions were carried out in 96 well microtitre plate and NADH consumption was measured by fluorescence using a FLASHScan 550 plate reader (Analytick Jena), controlled by a computer running WinFLASH®Version 1.6. At the start of each cycle the plate was shaken rotationally. The recorded spectra were from 200–600 nm and the excitation wave length filter was set to 365 nm. The fluorescence was measured at 4 points per well and at the maximum rate achievable by the plate reader. The final signal was then intergrated from 450 to 500 nm to give a relative fluorescence signal that was calibrated against a NADH standard. The reaction mixture analysed contained 50 mM HEPES buffer (pH 7.5), 5 mM $MgCl_2$, 1 mM TCEP, 1 mM ATP, 2 mM substrate, 3 mM PEP, 4 U myokinase, 4.5 U pyruvate kinase, 6.5 U lactate dehydrogenase, 94 mM NAC and 0.25 mM NADH. The final concentration of A-domain was varied as required.

6

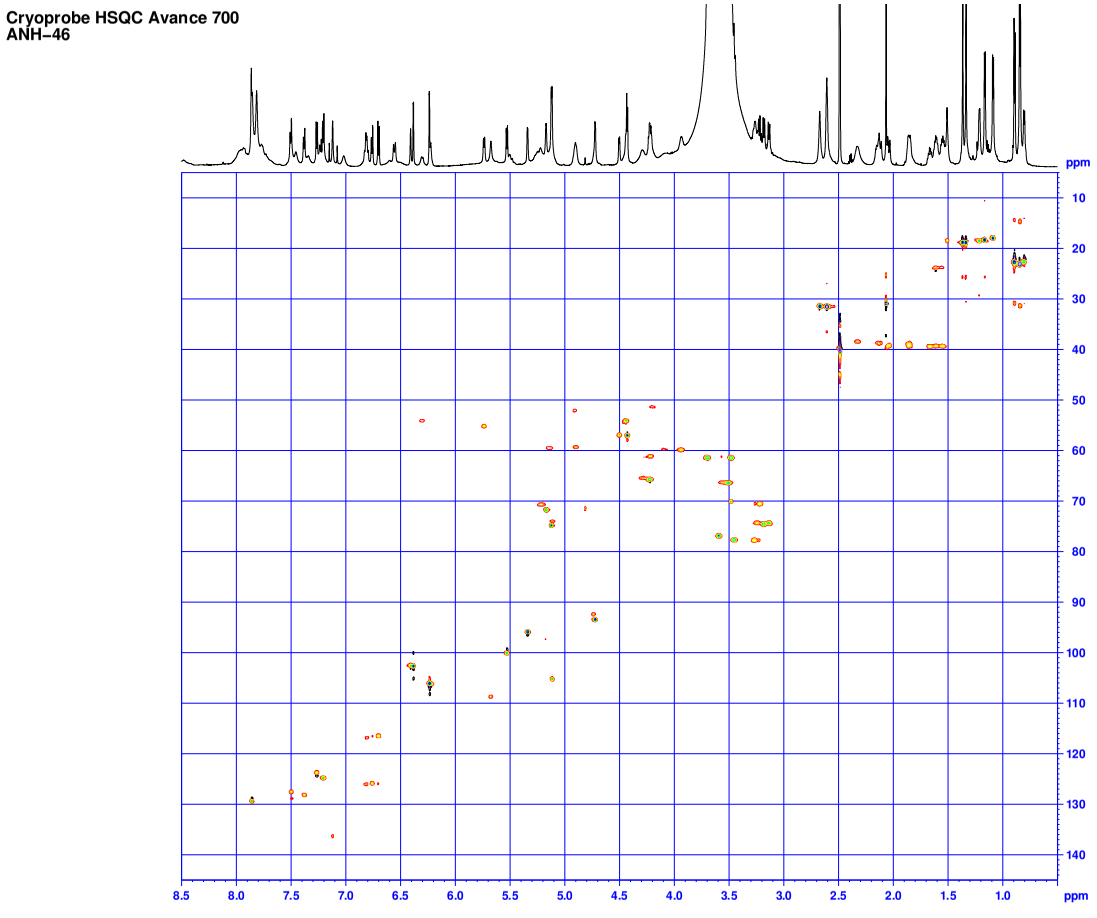
Supplementary

6.1 2-D NMR Spectra of chloroeremomycin

The partial assignment of the chloroeremomycin NMR is given in table 6.1. For convenience the HMQC and HMBC Spectra are also provided in figure 6.1 and 6.2 respectively.

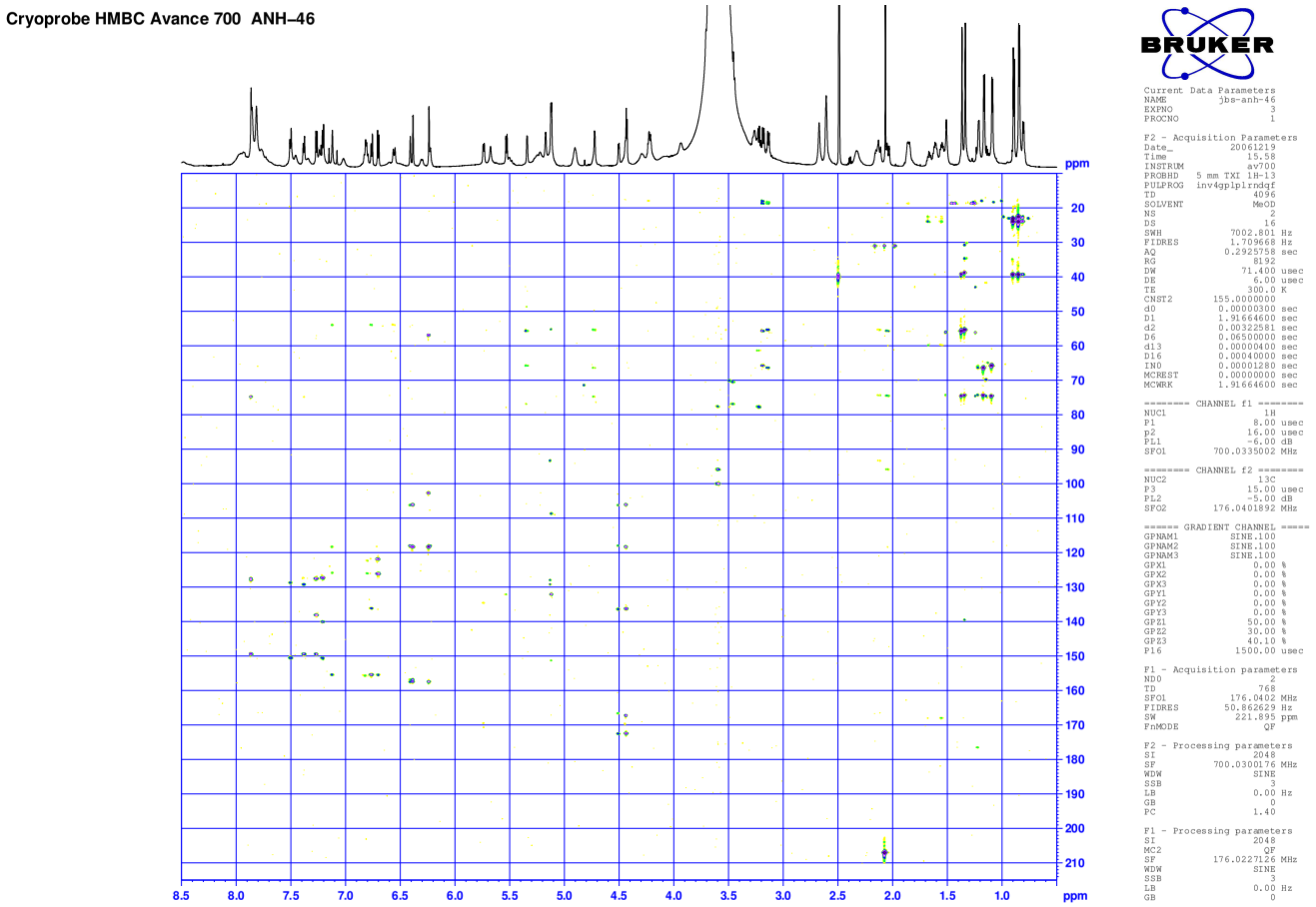
Table 6.1: Partial NMR assignment of chloroereomycin

Position	^{1}H	^{13}C	Position	^{1}H	^{13}C	Position	^{1}H	^{13}C
N-Me	2.61	31.2	3γ	-	174.4	5-C=O	-	171.8
1-NH		-	3γ-NH_2		-	6α		-
1α	3.93	60	3-C=O	-	171.29	6β	5.47	75.3
1β	1.55/1.65	39.6	4-NH		-	6a	-	132.5
1γ	1.6	23.7	4α	6.53	55.21	6b	7.85	129.5
1δ	0.87	22.3	4a		134.92	6c	7.06	123.55
1δ’	0.92	22.8	4b	5.63	107.7	6d	-	149.5
1-C=O			4c	-	157.21	6e	7.37	128
2-NH	5.45	-	4d	-	134.22	6f	7.25	123.4
2α	4.88	59.5	4e	-	153.34	6-C=O	-	167.88
2β	5.12	71.9	4f	5.12	104.8	7-NH		
2a	-	130.7	4-C=O	-	172.11	7α	4.75	59.46
2b	7.39	128.35	5-NH	6.57	-	7a	-	137.92
2c	-	150.76	5α	4.56	55.34	7b	-	118.27
2d	-	124.56	5a	-	126.56	7c	-	155.68
2e	7.2	124.9	5b	7.16	136.31	7d	6.61	103.87
2f	7.5	127.5	5c	-	122.17	7e	-	157.21
3-NH		-	5d	-	155.08	7f	6.23	106.5
3α	4.94	53.1	5e	6.7	117.6	7-C=O	-	176.4
3β	2.67	37	5f	6.77	125.9			




```
Current Data Parameters
NAME                 jbs-anh-46
EXPNO                         2
PROCNO                        1

F2 - Acquisition Parameters
Date_              20061219
Time                   14.49
INSTRUM                av700
PROBHD    5 mm TXI 1H-13
PULPROG             inv1etgp
TD                      2048
SOLVENT                 DMSO
NS                         2
DS                        16
SWH                7002.801 Hz
FIDRES             3.419337 Hz
AQ                 0.1463486 sec
RG                      8192
DW                   71.400 usec
DE                    6.00 usec
TE                    300.0 K
CNST2             155.0000000
d0                0.00000300 sec
D1                2.00000000 sec
d4                0.00161290 sec
d11               0.03000000 sec
d13               0.00000400 sec
D16               0.00040000 sec
DELTA             0.00192200 sec
DELTA1            0.00010490 sec
IN0               0.00001710 sec
MCREST            0.00000000 sec
MCWRK             0.40000001 sec
STICNT                   128

======== CHANNEL f1 ========
NUC1                     1H
P1                    8.00 usec
p2                   16.00 usec
P28                   0.10 usec
PL1                  -6.00 dB
SFO1           700.0335002 MHz

======== CHANNEL f2 ========
CPDPRG2                garp
NUC2                    13C
P3                   15.00 usec
p4                   30.00 usec
PCPD2                55.00 usec
PL2                  -5.00 dB
PL12                  5.50 dB
SFO2           176.0367658 MHz

====== GRADIENT CHANNEL =====
GPNAM1             SINE.100
GPNAM2             SINE.100
GPX1                  0.00 %
GPX2                  0.00 %
GPY1                  0.00 %
GPY2                  0.00 %
GPZ1                 80.00 %
GPZ2                 20.10 %
P16                1500.00 usec

F1 - Acquisition parameters
ND0                       2
TD                      640
SFO1           176.0368 MHz
FIDRES          45.687134 Hz
SW                166.100 ppm
FnMODE     Echo-Antiecho

F2 - Processing parameters
SI                      2048
SF             700.0300197 MHz
WDW                    QSINE
SSB                       2
LB                     0.00 Hz
GB                        0
PC                     1.00

F1 - Processing parameters
SI                      2048
MC2            echo-antiecho
SF             176.0227129 MHz
WDW                    QSINE
SSB                       2
```


Figure 6.1: HSQC 2D NMR spectrum of chloroeremomycin



BRUKER

```
Current Data Parameters
NAME                 jbs-anh-46
EXPNO                         3
PROCNO                        1

F2 - Acquisition Parameters
Date_              20061219
Time                   15.58
INSTRUM                av700
PROBHD    5 mm TXI 1H-13
PULPROG   inv4gplplrndqf
TD                      4096
SOLVENT                 MeOD
NS                         2
DS                        16
SWH                7002.801 Hz
FIDRES             1.709668 Hz
AQ                 0.2925758 sec
RG                      8192
DW                   71.400 usec
DE                    6.00 usec
TE                   300.0 K
CNST2            155.0000000
d0                0.00000300 sec
D1                1.91664600 sec
d2                0.00322581 sec
D6                0.06500000 sec
d13               0.00000400 sec
D16               0.00040000 sec
IN0               0.00001280 sec
MCREST            0.00000000 sec
MCWRK             1.91664600 sec

======== CHANNEL f1 ========
NUC1                      1H
P1                    8.00 usec
p2                   16.00 usec
PL1                  -6.00 dB
SFO1            700.0335002 MHz

======== CHANNEL f2 ========
NUC2                     13C
P3                   15.00 usec
PL2                  -5.00 dB
SFO2            176.0401892 MHz

====== GRADIENT CHANNEL =====
GPNAM1               SINE.100
GPNAM2               SINE.100
GPNAM3               SINE.100
GPX1                    0.00 %
GPX2                    0.00 %
GPX3                    0.00 %
GPY1                    0.00 %
GPY2                    0.00 %
GPY3                    0.00 %
GPZ1                   50.00 %
GPZ2                   30.00 %
GPZ3                   40.10 %
P16                 1500.00 usec

F1 - Acquisition parameters
ND0                       2
TD                      768
SFO1            176.0402 MHz
FIDRES          50.862629 Hz
SW                221.895 ppm
FnMODE                    QF

F2 - Processing parameters
SI                      2048
SF              700.0300176 MHz
WDW                     SINE
SSB                        3
LB                    0.00 Hz
GB                         0
PC                     1.40

F1 - Processing parameters
SI                      2048
MC2                       QF
SF              176.0227126 MHz
WDW                     SINE
SSB                        3
LB                    0.00 Hz
GB                         0
```

Figure 6.2: HSBC 2D NMR spectrum of chloroeremomycin

6.2 Mass spectra data for labelling study

The data is shown in table 6.2. As the [$^{2}H_4$,^{18}O]Hma that was fed was only 87 % enriched in ^{18}O the expected ratio for [$^{2}H_5$,^{18}O]:[$^{2}H_5$,$^{18}O_2$] is (2 $\times$ 0.87 $\times$ 0.13):(0.87 $\times$ 0.87). This equates to 1:3.35. The observed ratio based on the deconvolution above is 1:3.5. All data are shown as percent relative abundance.

Table 6.2: Deconvolution of the mass spectral data for dechloroeremomycin after feeding of [$^{2}H_4$,^{18}O]Hma

	Fed	Natural abundance	Difference	[$^{2}H_2$]	[$^{2}H_3$]	[$^{2}H_2$,^{18}O]	[$^{2}H_5$], [$^{2}H_3$,^{18}O]	[$^{2}H_5$,^{18}O]	[$^{2}H_5$,$^{1}8O_2$]	Calculated Total
M	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
M+1	82.3	83.8	-1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
M+2	36.5	36.0	0.5	0.5	0.0	0.0	0.0	0.0	0.0	0.5
M+3	12.1	11.1	1.0	0.4	0.6	0.0	0.0	0.0	0.0	1.0
M+4	21.1	2.2	18.9	0.2	0.5	18.3	0.0	0.0	0.0	18.9
M+5	27.6	0.0	27.6	0.1	0.2	15.3	12.1	0.0	0.0	27.7
M+6	16.9	0.0	16.9	0.0	0.1	6.6	10.1	0.0	0.0	16.8
M+7	9.6	0.0	9.6	0.0	0.0	2.0	4.3	5.0	0.0	11.3
M+8	7.7	0.0	7.7	0.0	0.0	0.4	1.3	4.2	0.0	5.9
M+9	20.7	0.0	20.7	0.0	0.0	0.0	0.3	1.8	17.5	19.6
M+10	14.4	0.0	14.4	0.0	0.0	0.0	0.0	0.6	14.7	15.3
M+11	5.4	0.0	5.4	0.0	0.0	0.0	0.0	0.1	6.3	6.4

7

Bibliography

1. Veiga-Crespo, P., Ageitos, J. M., Poza, M., and Villa, T. (2007) Enzybiotics: a look to the future, recalling the past. *J. Pharm. Sci.*, **96**, 1917–24.

2. Lister, J. (1867) On the antiseptic principle in the practice of surgery. *Lancet*, **90**, 353–356.

3. Pasteur, L., Camberland, C.-E., and Roux, E. (1881) Compte rendu sommaire des experiences faites a Pouilly-le-Fort, pres Melun, sur la vaccination charbonneuse. *Comptes Rendus des seances De L'Academie des Sciences*, **92**, 1378–83.

4. Kaufmann, S. H. and Schaible., U. E. (2005) 100th anniversary of Robert Koch's Nobel prize for the discovery of the *Tubercle bacillus*. *Trends Microbiol.*, **13**, 469–475.

5. Fleming, A. (1929) *Br. J. Exp. Pathol.*, **10**, 226–236.

6. Waksman, S. A., Bugie, E., and Schatz, A. (1944) *Proc. Staff Meetings Mayo Clinic*, **19**, 537–548.

7. Brotzu, G. (1948) Research on a new antibiotic. *Publications of the Cagliari Inistitute of Hygiene*, pp. 5–19.

8. Dax, S. L. (1997) *Antibacterial chemotherapeutic Agents*. Blackie A & P, 1st edn.

9. Levy, S. B. (1992) *The Antibiotic Paradox: How Miracle Drugs Are Destroying the Miracle*. Da Capo Press Inc.

10. Palumbi, S. R. (2001) Humans as the world's greatest evolutionary force. *Science*, **293**, 1786–1790.

11. Neu, H. C. (1994) Emerging trends in antimicrobial resistance in surgical infections. a review. *Eur. J. Surg. Suppl.*, **573**, 7–18.

12. Hiramatsu, K., Cui, L., Kuroda, M., and Ito, T. (2001) The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol.*, **9**, 486–493.

13. Walsh, C. T. (2003) *Antibiotics, Actions, Origins, Resistance*. ASM Press.

14. Brotz, H. and Sahl, H.-G. (2000) New insights into the mechanism of action of lantibiotics – diverse biological effects by binding to the same molecular target. *J. Antimicob. Chemother.*, **46**, 1–6.

15. Walsh, C. T., Fisher, S. L., Park, S. I., Prahalad, M., and Wu, Z. (2001) Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. *Chem. Biol.*, **3**, 21–28.

16. Bugg, T. D., Wright, G. D., Dutka-Malen, S., M. Arthur, P. C., and Walsh, C. T. (1991) Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins vanh and vana. *Biochemistry*, **30**, 10408–10415.

17. Lattanzi, A. (1970) Mechanism of action of synthetic antibiotics and synthetic antibiotics resistance. *Omnia Media et Therapeutica*, **48**, 175–184.

18. Moellering, R. C. and Weinberg, A. N. (1971) Studies on antibiotic synergism against enterococci ii. effect of various antibiotics on the uptake of 14C-labeled streptomycin by enterococci. *J. Clin. Invest.*, **50**, 2580–2584.

19. Uttley, A. H., Collins, C. H., Naidoo, J., and George, R. C. (1988) Vancomycin-resistant enterococci. *Lancet*, **1**, 57–58.

20. Smith, T. L., et al. (1999) Emergence of vancomycin resistance in *Staphylococcus aureus. New Engl. J. of Med.*, **340**, 493–501.

21. Sieber, S. A. and Marahiel, M. A. (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. *Chem. Rev.*, **105**, 715–738.

22. Chen, C. L., Chang, L. K., Chang, Y. S., Liu, S. T., and Tschen, J. S. (2005) Transposon mutagenesis and cloning of the genes encoding the enzymes of fengycin biosynthesis in *Bacillus subtilus*. *Mol. Gen. Genet.*, **248**, 121–125.

23. Peypoux, F., Bonmatin, J. M., and Wallach, J. (1999) Recent trends in micro-biochemistry of surfactin. *Appl. Microbiol. Biotechnol.*, **51**, 553–63.

24. Steller, S. and Vater, J. (2000) Purification of fengycin synthetase multienzyme system from *Bacillus subtilus*. *J. Chromatogr. B Biomed. Sci. Appl.*, **271**, 267–75.

25. Li, T.-L., Huang, F., Haydock, S. F., Mironenko, T., Leadlay, P. F., and Spencer, J. B. (2004) Biosynthetic gene cluster of the glycopeptide antibiotic teicoplanin. Characterization of two glycosyltranferases and the key acyltransferase. *Chem. Biol.*, **11**, 107–119.

26. Guenzi, E., Galli, G., Grgurina, I., Gross, D. C., and Grandi, G. (1998) Characterization of the syringomycin synthetase gene cluster. a link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.*, **273**, 32857–32863.

27. Weber, G., Schrgendorfer, K., Schneider-Scherzer, E., and Leitner, E. (1994) The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame. *Curr. Genet.*, **26**, 120–125.

28. Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of Gramicidin S. *EMBO J.*, **16**, 4174–4183.

29. Stachelhouse, T., Mootz, H. D., and Marahiel, M. A. (1999) Specificity-conferring code of adenylation domains in non-ribosomal biosynthesis of gramicidin. *Chem. Biol.*, **6**, 493–505.

30. Challis, G., Ravel, J., and Townsend, C. A. (2000) Predictive, structure based model of amino acid recognition by nonribosomal peptide synthetases. *Chem. Biol.*, **7**, 211–224.

31. Gruenewald, J., Sieber, S. A., and Marahiel, M. A. (2004) Chemo- and regioselective peptide cyclization triggered by the N-terminal fatty acid chain length: The recombinant cyclase of the calcium-dependent antibiotic from *Streptomyces coelicolor*. *Biochemistry*, **43**, 2915–2925.

32. Clugston, S. L., Sieber, S. A., Marahiel, M. A., and Walsh, C. T. (2003) Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: The C5 domain of tyrocidine synthetase is a DCL catalyst. *Biochemistry*, **42**, 12095–12104.

33. Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., and Walsh, C. T. (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry*, **37**, 1585–1595.

34. Belshaw, P. J., Walsh, C. T., and Stachelhaus, T. (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, **284**, 486–489.

35. Doekel, S. and Marahiel, M. A. (2000) Dipeptide formation on engineered hybrid peptide synthetase. *Chem. Biol.*, **7**, 373–384.

36. Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2000) Construction of hybrid peptide synthetases by module and domain fusions. *Proc. Natl. Acad. Sci. USA*, **97**, 5848–5853.

37. Linne, U., Stein, D. B., Mootz, H. D., and Marahiel, M. A. (2003) Systematic and quantitative analysis of protein-protein recognition between nonribosomal peptide synthetases investigated in the tyrocidine biosynthetic template. *Biochemistry*, **42**, 5114–5124.

38. Mootz, H. D., Kessler, N., Linne, U., Eppelmann, K., Schwarzer, D., and Marahiel, M. A. (2002) Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J. Am. Chem. Soc.*, **124**, 10980–10981.

39. Schwarzer, D., Mootz, H. D., Linne, U., and Marahiel, M. A. (2002) Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci. USA*, **99**, 14083–14088.

40. Li, J. and Jensen, S. E. (2008) Nonribosomal biosynthesis of fusaricidins by *Paenibacillus polymyxa* PKB1 involves direct activation of D-amino acid. *Chem. Biol.*, **15**, 118–127.

41. Hahn, M. and Stachelhaus, T. (2004) Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc. Natl. Acad. Sci. USA.*, **101**, 15585–15590.

42. Lautru, S., Deeth, R. J., Bailey, L. M., and Challis, G. L. (2005) Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat. Chem. Biol.*, **1**, 265–269.

43. Süssmuth, R. D. and Wohlleben, W. (2004) The biosynthesis of glycopeptide antibiotics – a model for complex, non-ribosomally synthesized, peptidic secondary metabolites. *Appl. Microbiol. Biotechnol.*, **63**, 344–350.

44. Poulakou, G. and Giamarellou, H. (2008) Oritavancin: a new promising agent in the treatment of infections due to Gram-positive pathogens. *Expert Opin. Invest. Drugs*, **17**, 225–243.

45. van Wageningen, A. M., Kirkpatrick, P. N., Williams, D. H., Harris, B. R., Kershaw, J. K., Lennard, N. J., Jones, M., Jones, S. J. M., and Solenberg, P. J. (1998) Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem. Biol.*, **5**, 155–162.

46. Kare, M. and Keady, D. (2003) Antimicrobial therapy of methicillin resistant staphylococcus aureus infection. *Expert Opin. Pharm.*, **4**, 165–177.

47. Nicolaou, K. C., Boddy, C. N. C., Brase, S., and Winssinger, N. (1999) Chemistry, biology, and medicine of the glycopeptide antibiotics. *Angew. Chem. Int. Ed. Engl.*, **38**, 2096–2159.

48. Solenberg, P. J., Matsushima, P., Stack, D. R., Wilklie, S. C., Thompson, R. C., and Baltz, R. H. (1997) Production of hybrid glycopeptide antibiotics in vitro and in *Streptomyces toyocaensis*. *Chem. Biol.*, **4**, 195–202.

49. Hubbard, B. K., Thomas, M. G., and Walsh, C. T. (2000) Biosynthesis of L- p-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics. *Chem. Biol.*, **7**, 931–941.

50. Hammond, S. J., Williamson, M. P., Williams, D. H., Boeck, L. D., and Marconi, G. G. (1982) On the biosynthesis of the antibiotic vancomycin. *J. Chem. Soc., Chem. Commun.*, pp. 344–346.

51. Pfeifer, V., Nicholason, G. J., Ries, J., Recktenwald, J., Schefer, A. B., Shawky, R. M., Schroder, J., Wohlleben, W., and Pelzer, S. (2001) A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (*S*)-3,5-dihydroxyphenylglycine. *Biol. Chem.*, **236**, 38390–38377.

52. Chen, H., Tseng, C. C., K., B., and Walsh, C. T. (2001) Glycopeptide biosynthesis: enzymatic assembly of the dedicated amino acid monomer (*S*)-3,5-dihydroxyphenylglycine. *Proc. Natl. Acad. Sci. USA*, **98**, 14901–14905.

53. Li, T. L., Choroba, O. W., Hong, H., Williams, D. . H., and Spencer, J. B. (2001) Biosynthesis of the vancomycin group of antibiotics: characterisation of a type III polyketide synthase in the pathway to (*S*)-3,5-dihydroxyphenylglycine. *Chem. Commun.*, pp. 2156–2157.

54. Widboom, P. F., Fielding, E. N., Liu, Y., and Bruner, S. D. (2007) Structural basis for cofactor-independent dioyxgenation in vancomycin biosynthesis. *Nature*, **447**, 342–345.

55. Puk, O., D., Kittel, C., Pelzer, S., Weist, S., Stegmann, E., Süssmuth, R. D., and Wohlleben, W. (2004) Biosynthesis of chloro-beta-hydroxytyrosine, a non-

proteinogenic amino acid of the peptidic backbone of glycopeptide antibiotics. *J. Bacteriol*, **186**, 6093–6100.

56. He, J., Magarvey, N., Piraee, M., and Vining, L. C. (2001) The gene cluster for chloramphenicol biosynthesis in *Streptomyces venezuelae* ISP5230 includes novel shikimate pathway homologues and a monomodular non-ribosomal peptide synthetase gene. *Microbiol.*, **147**, 2817–2829.

57. Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2002) Ways of assembling complex natural products on modular nonribosomal peptide sythetases. *ChemBioChem*, **3**, 490–504.

58. Konz, D. and Marahiel, M. A. (1999) How do peptide synthetases generate structural diversity? *Chem. Biol.*, **6**, R39–R48.

59. Schwarzer, D., Finking, R., and Marahiel, M. A. (2002) Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.*, **20**, 275–287.

60. Stachelhouse, T., Mootz, M. D., and Marahiel, M. A. (1999) How do peptide synthetases generate structure diversity? *Chem. Biol.*, **6**, 293–505.

61. Trauger, J. W. and Walsh, C. T. (2000) Heterologous expression in *Escherichia coli* of the first module of the nonribosomal peptide synthetase for chloroeremomycin, a vancomycin-type glycopeptide antibiotic. *Proc. Natl. Acad. Sci. USA*, **97**, 3112–3117.

62. Balibar, C. J., Vaillancourt, F. H., and Walsh, C. T. (2005) Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem. Biol.*, **12**, 1189–1200.

63. Rausch, C., Hoof, I., Weber, T., Wohlleben, W., and Huson, D. H. (2007) Phylogentic analysis of condensation domains in NRPS sheds light on their function evolution. *BMC Evol. Biol.*, **7**, 78.

64. Hillson, N. J. and Walsh, C. T. (2003) Dimeric structure of the six-domain VibF subunit of ribriobactin synthetase: Mutant domain activity regain and ultracentrifugation studies. *Biochemistry*, **42**, 966–775.

65. Gribble, G. W. (1998) Naturally occuring organohalogen compounds. *Acc. Chem. Res.*, **31**, 141–152.

66. Nagarajan, S. R. (1993) Structure-activity relationships of vancomycin-type glycopeptide antibiotcs. *J. Antibiot.*, **46**, 1181–1195.

67. Harris, C. M., Kannan, R., Kopecka, H., and Harris, T. M. (1985) The role of chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono- and didechlorovancomycin. *J. Am. Chem. Soc.*, **107**, 6652–6658.

68. Gerhard, U., Mackay, J. P., Maplestone, R. A., and Williams, D. H. (1993) The role of sugar and chlorine substituents in the dimerization of the antibiotic vancomycin antibiotics. *J. Am. Chem. Soc.*, **115**, 232–237.

69. Bischoff, D., Pelzer, S., Bister, B., Nicholson, G. J., Stockert, S., Schirle, M., Wohlleben, W., Jung, G., and Süssmuth, R. D. (2001) The biosynthesis of vancomycin-type glycopeptide antibiotics: the order of cyclisation steps. *Angew. Chem. Int. Ed. Engl.*, **40**, 4688–4691.

70. Williams, D. H. and Bardsely, B. (1999) The vancomycin group of antibiotics and fight against bacteria. *Angew. Chem. Int. Ed. Engl.*, **38**, 1173–1193.

71. Bischoff, D., Pelzer, S., Holtzel, A., Nicholson, G. J., Stockert, S., Wohlleben, W., Jung, G., and Süssmuth, R. D. (2001) The biosynthesis of vancomycin-type glycopeptide antibiotics: new insights into cyclisation steps. *Angew. Chem. Int. Ed. Engl.*, **40**, 1693–1696.

72. Süssmuth, R. D., Pelzer, S., Nicholson, G., Walk, T., Wohlleben, W., and Jung, G. (1999) New advances in the biosynthesis of glcopeptide antibiotics of the vancomycin type from *Amycolaptopsis mediterranei*. *Angew. Chem. Int. Ed. Engl.*, **38**, 1976–1979.

73. Lehmann, C., Bunkoczi, G., Vertesy, L., and Sheldrick, G. M. (2002) Structures of glycopeptide antibiotics with peptides that model bacterial cell-wall precursors. *J. Mol. Biol.*, **318**, 723.

74. Zerbe, K., Withe, K., Li, D. B., Vitali, F., Bigler, L., and Robinson, J. A. (2004) An oxidative phenol coupling reaction catalyzed by OxyB, a cytochrome P450 from the vancomycin producing organism. *Angew. Chem. Int. Ed. Engl.*, **43**, 6709–6713.

75. Snyder, N. J., Cooper, R. D. G., Briggs, B. S., Zmijewski, M., Mullen, D. L., Kaiser, R. E., and Nicas, T. I. (1998) Enzymatic deacylation of teicoplanin followed by reductive alkylation: synthase and antibacterial activity of new glycopeptides. *J. Antibiot.*, **51**, 945–951.

76. Shlüzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A., and Franceschi, F. (2001) Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature*, **413**, 814–821.

77. Li, Y., Llewellyn, N. M., Giri, R., Huang, G., and Spencer, J. B. (2005) Biosynthesis of the amino acid side chain of butirosin: Possible protective-group

chemistry in an acyl carrier protein-mediated pathway. *Chem. Biol.*, **12**, 665–675.

78. Mulichak, A. M., Losey, H. C., Lu, W., Wawrazak, Z., Walsh, C. T., and Gararvito, R. M. (2003) Structure of the TDP- epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc. Natl. Acad. Sci. USA*, **100**, 9238–9243.

79. Ha, S., Walker, D., Shi, Y., and Walker, S. (2000) The 1.9 Å crystal structure of *Escherichia coli*, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.*, **9**, 1305–1314.

80. Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M., , and Walker, S. (2004) Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc. Natl. Acad. Sci. USA*, **100**, 845–849.

81. Vrielink, A., Ruger, W., Driessen, H. P., and Freemont, P. S. (1994) Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J.*, **13**, 3413–3433.

82. Morra, S., Imberty, A., Aschke-Sonnenborn, U., Rüger, W., and Freemont, P. S. (1999) T4 phage beta-glucosyltransferase: substrate binding and proposed catalytic mechanism. *J. Mol. Biol.*, **3**, 717–730.

83. Chen, H., Thomas, M. G., Hubbard, B. K., Losey, H. C., Walsh, C. T., and Burkart, M. D. (2000) Deoxysugars in glycopeptide antibiotics: Enzymatic synthesis of TDP-L-epivancosamine in chloroeremomycin biosynthesis. *Proc. Natl. Acad. Sci. USA*, **97**, 11942–11947.

84. Perkins, H. R. (1969) Specificity of combination between mucopeptide precursors and vancomycin or ristocetin. *Biochem. J.*, **111**, 195–200.

85. Williams, D. H., Williamson, M. P., Butcher, D. W., and Hammond, S. J. (1983) Detailed binding sites of the antibiotics vancomycin and ristocetin a: determination of intermolecular distances in antibiotic/substrate complexes by use of the time-dependaent NOE. *J. Am. Chem. Soc.*, **105**, 1332–1339.

86. Schoutena, J. A., Baggaa, S., Lloydb, A. J., de Pascaleab, G., Dowsonb, C. G., Roperb, D. I., and Bugg, T. D. H. (2006) Fluorescent reagents for *in vitro* studies of lipid-linked steps of bacterial peptidoglycan biosynthesis: derivatives of UDPMurNAc-pentapeptide containing D-cysteine at position 4 or 5. *Mol. BioSyst.*, **2**, 484–491.

87. Loll, P. J., Bevivino, A. E., Korty, B. D., and Axelsen, P. H. (1997) Simultaneous recognition of a carboxylate-containing ligand and an intramolecular surrogate ligand in the crystal structure of an asymmetric vancomycin dimer. *J. Am. Chem. Soc.*, pp. 1516–1522.

88. SchŁfer, M., Schneider, T. R., and Sheldrick, G. M. (1996) Crystal structure of vancomycin. *Structure*, **4**, 1509–1515.

89. Waltho, J. P. and Williams, D. H. (1989) Aspects of molecular recognition: solvent exclusion and dimerization of the antibiotic ristocetin when bound to a model bacterial cell-wall precursor. *J. Am. Chem. Soc.*, **111**, 2475–2480.

90. Mackay, J. P., Gerhard, U., Beauregard, D. A., Maplestone, R. A., and Williams, D. H. (1994) Dissection of the contributions toward dimerization of glycopeptide antibiotics. *J. Am. Chem. Soc.*, **116**, 4573–4580.

91. V. M. Good, M. N. G. and Knowles, D. J. C. (1990) MM 45289, a potent glycopeptide antibiotic which interacts weakly with diacetyl-L-lysyl-D-alanyl-D-alanine. *J. AntiBiot.*, **43**, 550–550.

92. Beauregard, D. A., Williams, D. H., Gwynn, M. N., and Knowles, D. J. (1995) Dimerization and membrane anchors in extracellular targeting of vancomycin group antibiotics. *Antimicrob. Agents Chemother.*, **39**, 781–785.

93. Nicolaou, K. C., Mitchell, H. J., Jain, N. F., Bando, T., andN. Winssinger, R. H., Natarajan, S., and Koumbis, A. E. (1999) Total synthesis of vancomycin - part 4: Attachment of the sugar moieties and completion of the synthesis. *Chem.-Eur J.*, **5**, 2648–2667.

94. Noble, W. C., Virani, Z., and Cree, R. G. A. (1992) Co-transfer of vancomycin and other resistance genes from *Enterococcus faecalis* NCTC12201 to *Staphylococcus aureus*. *FEMS Microb. Lett.*, **93**, 195.

95. Hiramatsu, K. (1998) Vancomycin resistance in staphylococci. *Drug Resist. Updat.*, **1**, 135–150.

96. Bambeke, F. V. (2004) Glycopeptide in clinical development: pharmacological profile and clinical perspectives. *Curr. Opin. Pharm.*, p. 471.

97. Witte, W. (2004) Glycopeptide resistant Staphylococcus. *J. Vet. Med. B Infect. Dis. Vet. Public Health*, **51**, 370–373.

98. Pootoolal, J., Neu, J., and Wright, G. D. (2002) Glycopeptide antibiotic resistance. *Annu. Rev. Pharmacol. Toxicol*, **42**, 381–408.

99. Gehring, A. M., Mori, I., and Walsh, C. T. (1998) Reconstitution and characterization of the *Eshcerichia coli* enterobactin synthetase for EntB, EntE and EntF. *Biochemisty*, **37**, 2648–2659.

100. Williams, D. H. (1996) The glycopeptide story – how to kill the deadly 'superbugs'. *Nat. Prod. Rep.*, **13**, 469–477.

101. Newman, D. J. and Cragg, G. M. (2007) Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.*, **70**, 461–477.

102. Dobson, C. M. (2004) Chemical space and biology. *Nature*, **432**, 824–828.

103. Feher, M. and Schmidt, J. M. (2003) Property distributions: differences between drugs, natural products and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.*, **43**, 218–277.

104. Ritter, T. K., Mong, K. K., Liu, H., Nakatani, T., and Wong, C. H. (2003) A programmable one-pot oligosaccharide synthesis for diversifying the sugar domains of natural products: A case study of vancomycin. *Angew. Chem. Int. Ed. Engl.*, **42**, 4657–4660.

105. Powell, A., Amir-Heidari, M. B. B., Neary, J. M., Thirlway, J., Wilkinson, B., Smith, C. P., and Micklefield, J. (2007) Engineered biosynthesis of nonribosomal lipopeptides with modified fatty acid side chains. *J. Am. Chem. Soc.*, **129**, 15182–15191.

106. Butz, D., T. Schmiederer, B. H., Wohlleben, W., Weber, T., and Süssmuth, R. D. (2008) Module extension of a non-ribosomal peptide synthetase of the glycopeptide antibiotic balhimycin produced by *Amycolatopsis balhimycina*. *ChemBioChem*, **9**, 1195–1200.

107. Bugg, T. D. and Ramaswamy, S. (2008) Non-heme iron-dependent dioxygenases: unravelling catalytic mechanisms for complex enzymatic oxidations. *Curr. Opin. Chem. Biol.*, **12**, 134–140.

108. Bugg, T. D. (2003) Dioxygenase enzymes: catalytic mechanisms and chemical models. *Tetrahedron*, **59**, 7075–7101.

109. Feig, A. L. and Lippard, S. J. (1994) Reactions of non-heme iron(II) centers with dioxygen in biology and chemistry. *Chem. Rev.*, **94**, 759–805.

110. Kivirikko, K. I., Myllyla, R., and Pihlajaniemi, T. (1989) Protein hydroxylation: prolyl 4-hydroxylase, an enzyme with four cosubstrates and a multifunctional subunit. *FASEB J.*, **3**, 1609–1617.

111. Hanauske-Abel, H. M. and Günzler, V. (1982) A stereochemical concept for the catalytic mechanism of prolylhydroxylase : Applicability to classification and design of inhibitors. *J. Theor. Biol.*, **94**, 421 – 455.

112. Salowe, S. P., Marsh, E. N., and Townsend, C. A. (1990) Purification and characterization of clavaminate synthase from *Streptomyces clavuligerus*: an unusual oxidative enzyme in natural product biosynthesis. *Biochemistry*, **29**, 6499–6508.

113. Bladwin, J. E. and Abraham, S. E. (1988) The biosynthesis of penicillins and cephalosporins. *Nat. Prod. Rep.*, **5**, 129–145.

114. Bradley, F. C., Lindstedt, S., Lipscomb, J. D., Que, J., Lawrence, Roe, A. L., and Rundgren, M. (1986) 4-hydroxyphenylpyruvate dioxygenase is an iron-tyrosinate protein. *J. Biol. Chem.*, **261**, 11694–11696.

115. Gibson, D. T. (ed.) (1984) *Microbial Degradation of Organic Molecules*. Marcel Dekker.

116. Que, L. and Ho, R. Y. N. (1996) Dioxygen activation by enzymes with mononuclear non-heme iron active sites. *Chemical Reviews*, **96**, 2607–2624.

117. Koeberle, A., Zettl, H., Greiner, C., Wurglics, M., Schubert-Zsilavecz, M., and Werz, O. (2008) Pirinixic acid derivatives as novel dual inhibitors of microsomal prostaglandin E2 synthase-1 and 5-lipoxygenase. *J. Med. Chem.*, **51**, 8068–8076.

118. Purpero, V., , and Moran, G. R. (2007) The diverse and pervasive chemistries of the α-keto acid dependent enzymes. *J. Biol. Inorg. Chem.*, **12**, 587–601.

119. Neidig, M. L., Kavana, M., Moran, G. R., and Solomon, E. I. (2004) CD and MCD studies of the non-heme ferrous active site in (4-hydroxyphenyl)pyruvate dioxygenase: Correlation between oxygen activation in the extradiol and α-kg-dependent dioxygenases. *J. Am. Chem. Soc.*, **126**, 4486–4487.

120. Neidig, M. L., Decker, A., Choroba, O. W., Huang, F., Kavana, M., Moran, G. R., Spencer, J. B., and Solomon, E. I. (2006) Spectroscopic and electronic structure studies of aromatic electrophilic attack and hydrogen-atom abstraction by non-heme iron enzymes. *Proc. Natl. Acad. Sci. USA*, **103**, 12966–12973.

121. Davis, M. I., Wasinger, E. C., Decker, A., Pau, M. Y. M., Vaillancourt, F. H., Bolin, J. T., Eltis, L. D., Hedman, B., Hodgson, K. O., and Solomon, E. I. (2003) Spectroscopic and electronic structure studies of 2,3-dihydroxybiphenyl 1,2-dioxygenase: O2 reactivity of the non-heme ferrous site in extradiol dioxygenases. *J. Am. Chem. Soc.*, **125**, 11214–11227.

122. Eltis, L. D. and Bolin, J. T. (1996) Evolutionary relationships among extradiol dioxygenases. *J. Bacteriol.*, **178**, 5930–5937.

123. Moran, G. R. (2005) 4-Hydroxyphenylpyruvate dioxygenase. *Arch. Biochem. Biophys.*, **433**, 117–128.

124. Serre, L., Sailland, A., Sy, D., Boudec, P., Rolland, A., Pebay-Peyroula, E., and Cohen-Addad, C. (1999) Crystal structure of *Pseudomonas fluorescens* 4-hydroxyphenylpyruvate dioxygenase: an enzyme involved in the tyrosine degradation pathway. *Structure*, **7**, 977–988.

125. Costas, M., Mehn, M. P., Jensen, M. P., and Que, L. (2004) Dioxygen activation at mononuclear nonheme iron active sites: Enzymes, models, and intermediates. *Chem. Rev.*, **104**, 939–986.

126. Lindblad, B., Lindstedt, G., and Lindstedt, S. (1970) Mechanism of enzymic formation of homogentisate from p-hydroxyphenylpyruvate. *J. Am. Chem. Soc.*, **92**, 7446–7449.

127. Counts, D. F., Cardinale, G. J., and Udenfriend, S. (1978) Prolyl hydroxylase half reaction: Peptidyl prolyl-independent decarboxylation of-ketoglutarate. *Proc. Natl. Acad. Sci. USA*, **75**, 2145–2149.

128. Cardinale, G. J. and Udenfriend, S. (1974) Prolyl hydroxylase. *Adv. Enzymol.*, **41**, 24–300.

129. Jefford, C. W. and Cadby, P. A. (1981) Evaluation of models for the mechanism of action of 4-hydroxyphenylpyruvate dioxygenase. *Experientia*, **37**, 1134–1137.

130. Thornburg, L. D., Lai, M. T., Wishniok, J. S., and Stubbe, J. (1993) A non-heme iron protein with heme tendencies: An investigation of the substrate specificity of thymine hydroxylase. *Biochemistry*, **32**, 14023–14033.

131. Pascal, R. A., Oliver, M. A., and Chen, Y. C. J. (1985) Alternate substrates and inhibitors of bacterial 4-hydroxyphenylpyruvate dioxygenase. *Biochemistry*, **24**, 3158–3165.

132. Crouch, N. P., Adlington, R. M., Baldwin, J. E., Lee, M. H., and MacKinnon, C. H. (1997) A mechanistic rationalisation for the substrate specificity of recombinant mammalian 4-hydroxyphenylpyruvate dioxygenase (4-HPPD). *Tetrahedron*, **53**, 6993–7010.

133. Hammond, S. J., Williams, D. H., and Nielsen, R. V. (1982) The biosynthesis of ristocetin. *J. Chem. Soc., Chem. Commun.*, pp. 116–117.

134. Choroba, O. W. (2001) *Enzymes involved in crosslinking and biosynthesis of the rare acids in glycopeptide antibiotics*. Ph.D. thesis, University of Cambridge, Cambridge, UK.

135. O'Hare, H. M., Huang, F., Holding, A., Choroba, O. W., and Spencer, J. B. (2006) Conversion of hydroxyphenylpyruvate dioxygenases into hydroxymandelate synthases by directed evolution. *FEBS Letters*, **580**, 3445–3450.

136. Mitchell, G., Bartlett, D. W., Fraser, T. E. M., Hawkes, T. R., Holt, D. C., Townson, J. K., and Wichert, R. A. (2001) Mesotrione: a new selective herbicide for use in maize. *Pest. Manage. Sci*, **57**, 120–128.

137. Lindstedt, S., Holme, E., Lock, E. A., Hjalmarson, O., and Strandvik, B. (1992) Treatment of hereditary tyrosinaemia typ1 by inihibition of 4-hydroxyphenyl dixoygenase. *Lancet*, **340**, 813–817.

138. Gunsior, M., Ravel, J., Challis, G. L., and Townsend, C. A. (2004) Engineering p-hydroxyphenylpyruvate dioxygenase to a p-hydroxymandelate synthase and evidence for the proposed benzene oxide intermediate in homogentisate formation. *Biochemistry*, **4**, 644–674.

139. Spencer, J. B., Choroba, O. W., and Williams, D. H. (2001) Biosynthesis of some unusual amino acids found in clinically important antibiotics. *Abstracts*

of Papers, 221st ACS National Meeting, San Diego, CA, United States April 1-5.

140. Serre, L., Sailland, A., Sy, D., Boudec, P., Rolland, A., Pebay-Peyroula, E., and Cohen-Addad, C. (1999) Crystal structure of *pseudomonas fluorescens* 4-hydroxyphenylpyruvate dioxygenase: an enzyme involved in the tyrosine degradation pathway. *Structure*, **7**, 977–988.

141. Fritze, I. M., Linden, L., Freigang, J., Auerbach, G., Huber, R., and Steinbacher, S. (2004) The crystal structures of zea mays and arabidopsis 4-hydroxyphenylpyruvate dioxygenase. *Plant Physiol.*, **134**, 188–1400.

142. Brownlee, J. M., Johnson-Winters, K., Harrison, D. H. T., and Moran, G. R. (2004) Structure of the ferrous form of (4-hydroxyphenyl)pyruvate dioxygenase from *Streptomyces avermitilis* in complex with the therapeutic herbicide, NTBC. *Biochemistry*, **43**, 6370–6377.

143. Yang, C., Pflugrath, J. W., Camper, D. L., Foster, M. L., Pernich, D. J., and Walsh, T. A. (2004) Structural basis for herbicidal inhibitor selectivity revealed by comparison of crystal structures of plant and mammalian 4-hydroxyphenylpyruvate dioxygenase. *Biochemistry*, **43**, 10414–10423.

144. Neidhardt, F. C. (1987) *Escherichia Coli and Salmonella typherium: Cellular and Molecular Biology*. Am. Soc. Microbiol.

145. Denoya, C., Skinner, D., and Morgenstern, M. (1994) A *Streptomyces avermitilis* gene encoding a 4-hydroxyphenylpyruvic acid dioxygenase-like protein that directs the production of homogentisic acid and an ochronotic pigment in *Escherichia coli*. *J. Bacteriol.*, **176**, 5312–5319.

146. Neidig, M. L. and Spencer, J. B. (2006) (personal communication).

147. Defoin, A., Pires, J., Tissot, I., Tschamber, T., Bur, D., Zehnder, D., and Streith, J. (1991) Single and double asymmetric induction in Diels-Alder cycloadditions with chiral acylnitroso dienophile. *Tetrahedron: Asymmetry*, **2**, 1209.

148. Ellis, M. K., Lindsteadt, S. T., Markstedt, M. E. H., Mutter, L. C., and Prisbylla, P. M. (1993) Preparation of 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione as a drug for treatment of tyrosinemia. *Application: WO*.

149. Brownlee, J. M., He, P., Moran, G. R., and Harrison, D. H. T. (2008) Two roads diverged: the structure of hydroxymandelate synthase from *Amycolatopsis orientalis* in complex with 4-hydroxymandelate. *Biochemistry*, **47**, 2002–2013.

150. Stegmann, E., Pelzer, S., D., Puk, O., Stockert, S., Butz, D., Zerbe, K., J. Robinson, R. D. S., and Wohlleben, W. (2006) Genetic analysis of the balhimycin (vancomycin-type) oxygenase genes. *J. Biotechnol.*, **124**, 640–653.

151. Müller, H. G., Schunk, W. H., and Kärgel, E. (1991) Cytochromes P-450 in alkane-assimilating yeasts. Ruckpaul, K. and Rein, H. (eds.), *Frontiers in Biotransformation*, vol. 4, pp. 87–126, Taylor and Francis.

152. Klingenberg, M. (1958) Pigments of rat liver microsomes. *Arch. Biochem. Biophys.*, **75**, 376–386.

153. Garfinkel, D. (1958) Studies on pig liver microsomes i. enzymic and pigment composition of different microsomal fractions. *Arch. Biochem. Biophys.*, **77**, 493–509.

154. Omura, T. and Sato, R. (1964) The carbon monoxide-binding pigment of liver microsomes i. evidence for its hemoprotein nature. *J. Biol. Chem.*, **239**, 2370–2378.

155. Omura, T. and Sato, R. (1964) The carbon monoxide-binding pigment of liver microsomes ii. solubilization, purification, and properties. *J. Biol. Chem.*, **239**, 2379–2385.

156. Mason, H. S., North, J. C., and Vanneste, M. (1965) Microsomal mixed-function oxidations: the metabolism of xenobiotics. *Fed. Proc.*, **24**, 1172–1180.

157. Nebert, D. W. and Gonzalez, F. J. (1987) P450 genes: structure, evolution, and regulation. *Ann. Rev. Biochem.*, **56**, 945–993.

158. Sligar, S. G. and Murray, R. I. (1966) Bacterial cytochrome P450s. Ortiz de Montellano, P. R. (ed.), *Cytochrome P450: Structure, Mechanism and Biochemistry*, pp. 429–503, Plenum Press, 1^{st} edn.

159. Peterson, J. A. and Graham-Lorence, S. E. (1995) Bacterial P450s: Structure similarities and function differences. Ortiz de Montellano, P. R. (ed.), *Cytochrome P450: Structure, Mechanism and Biochemistry*, pp. 151–182, Plenum Press, 2^{nd} edn.

160. Black, S. D. and Coon, M. J. (1987) P-450 cytochromes: structure and function. *Adv. Enzymol. Realt. Areas. Mol. Biol.*, **60**, 1172–1180.

161. Akhtar, M. and Wright, J. N. (1991) A unified mechanistic view of oxidative reactions catalysed by P-450 and related Fe-containing enzymes. *Nat. Prod. Rep.*, **8**, 527–551.

162. Black, S. D. and Coon, M. J. (1986) *Cytochrome P-450, Structure, Mechanism and Biochemistry*. Plenum Press New York.

163. Graf, H., Ruf, H. H., and Ullrich, V. (1983) Prostacyclin synthase, a cytochrome P450 enzyme. *Angew. Chem. Int. Ed. Engl.*, **6**, 478–488.

164. Peterson, J. A., Lorence, M. C., and Amarneh, J. (1990) Putidaredoxin reductase and putidaredoxin. Cloning, sequence determination, and heterologous expression of the proteins. *J. Biol. Chem*, **265**, 6066–6073.

165. Hintz, M. J. and Pederson, T. C. (1981) The kinetics of reduction of cytochrome P-450cam by reduced putidaredoxin. *J. Biol. Chem*, **256**, 6721–6728.

166. Raag, R. and Poulos, T. L. (1991) Crystal structure of cytochome P-450cam complexed with camphane, thiocamphor, and adamantane: factors controlling P-450 substrate hydroxylation. *Biochemistry*, **30**, 2674–2684.

167. Ravichandran, K. G., Boddupalli, S. S., Hasermann, C. A., Peterson, J. A., and Deisenhofer, J. (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science*, **261**, 731–736.

168. Cupp-Vickery, J., Anderson, R., and Hatziris, Z. (2000) Crystal structures of ligand complexes of P450eryF exhibiting homotropic cooperativity. *Proc. Natl. Acad. Sci. USA*, **97**, 3050–3055.

169. I. C. Gunsalus, S. G. S. (1978) Oxygen reduction by the P450 monoxygenase systems. *Adv. Enzymol. Relat. Areas. Mol. Biol.*, **47**, 1–44.

170. Lipscomb, J. D. (1980) Electron paramagnetic resonance detectable states of cytochrome P-450cam. *Biochemistry*, **19**, 3590–3599.

171. Champion, P. M., Stallard, B. R., Wagner, G. C., and Gunsalus, I. C. (1982) Resonance Raman detection of an iron-sulfur bond in cytochrome P450cam. *J. Am. Chem. Soc.*, **104**, 5469–5472.

172. Champion, P. M., Gunsalus, I. C., and Wagner, G. C. (1978) Resonance Raman investigations of cytochrome P450cam from *Pseudomonas putida*. *J. Am. Chem. Soc.*, **100**, 3743–3751.

173. Ozaki, Y., Kitagawa, T., Kyogoku, Y., Imai, Y., Hashimoto-Yusudo, C., and Sato, R. (1978) Resonance Raman studies of hepatic microsomal cytochromes P-450: evidence for strong .pi. basicity of the fifth ligand in the reduced and carbonyl complex forms. *Biochemistry*, **17**, 5826–5831.

174. O'Keefe, D. H., Ebel, R. E., Peterson, J. A., Maxwell, J. C., and Caughey, W. S. (1978) An infrared spectroscopic study of carbon monoxide bonding to ferrous cytochrome P-450. *Biochemistry*, **17**, 5845–5852.

175. Jung, C., Hoa, G. H. B., Schroeder, K. L., Simon, M., and Doucet, J. P. (1992) Substrate analog induced changes of the carbon-oxygen-stretching mode in the cytochrome P450cam-carbon monoxide complex. *Biochemistry*, **31**, 12855–12862.

176. Mouro, C., Jung, C., Bondon, A., and Simonneaux, G. (1997) Comparative Fourier transform infrared studies of the secondary structure and the CO heme ligand environment in cytochrome P-450cam and cytochrome P-420cam. *Biochemistry*, **36**, 8125–8134.

177. Keller, R. M., Wüthrich, K., and Debrunner, P. G. (1972) Proton magnetic resonance reveals high-spin iron (ii) in ferrous cytochrome P450cam from *Pseudomonas putida*. *Proc. Natl. Acad. Sci. USA*, **69**, 2073–2075.

178. Banci, L., Bertini, I., Marconi, S., Pieratelli, R., and Sligar, S. G. (1994) Cytochrome P450 and aromatic bases: A 1H NMR study. *J. Am. Chem. Soc.*, **116**, 4866–4873.

179. Philson, S. B., Debrunner, P. G., Schmidt, P. G., and Gunsalus, I. C. (1979) The effect of cytochrome P-450cam on the NMR relaxation rate of water protons. *J. Biol. Chem.*, **254**, 10173–10179.

180. Hahn, J. E., Hodgson, K. O., Andersson, L. A., and Dawson, J. H. (1982) Endogenous cysteine ligation in ferric and ferrous cytochrome P-450. Direct evidence from x-ray absorption spectroscopy. *J. Biol. Chem.*, **257**, 10934–10941.

181. Dawson, J. H., Trudell, J. R., Linder, R. E., Barth, G., Bunnenberg, E., and Djerassi, C. (1978) Magnetic circular dichroism of purified forms of rabbit liver cytochromes P-450 and P-420. *Biochemistry*, **17**, 33–42.

182. Thomann, H., Bernardo, M., Goldfarb, D., Kroneck, P. M. H., and Ullrich, V. (1995) Evidence for water binding to the Fe center in cytochrome P450cam obtained by 17O electron spin-echo envelope modulation spectroscopy. *J. Am. Chem. Soc.*, **117**, 8243–8251.

183. Sligar, S. G. (1976) Coupling of spin, substrate, and redox equilibriums in cytochrome P450. *Biochemistry*, **15**, 5399–5406.

184. Loew, G. H. and Harris, D. L. (2000) Role of the heme active site and protein environment in structure, spectra, and function of the cytochrome P450s. *Chem. Rev.*, **100**, 406–420.

185. Harris, D. and Loew, G. (1993) Determinants of the spin state of the resting state of cytochrome P450cam. *J. Am. Chem. Soc.*, **115**, 8775–8779.

186. Aissaoui, H., Bachmann, R., Schweiger, A., and Woggon, W. D. (1998) On the origin of the low-spin character of cytochrome P450cam in the resting state : Investigations of enzyme models with pulse EPR and ENDOR spectroscopy. *Angew. Chem. Int. Ed. Engl.*, **37**, 2998–3002.

187. Poulus, T. L., Finzel, B. C., and Howard, A. J. (1987) High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.*, **195**, 687–700.

188. Aissaoui, H., Ghirlanda, S., Gmur, C., and Woggon, W. D. (1996) The synthesis of a new active-site analogue of cytochrome P450 carrying substrate recognition sites and athiolate ligand. *J. Mol. Catal. A: Chem.*, **113**, 393–402.

189. Gunsalus, I. C., Meeks, J. R., and Lipscomb, J. D. (1973) Cytochrome P-450cam substrate and effector interactions. *Ann. N. Y. Acad. Sci*, **212**, 107–121.

190. Honeychurch, M. J., Hill, H. A. O., and Wong, L. L. (1999) The thermodynamics and kinetics of electron transfer in the cytochrome P450cam enzyme system. *FEBS Lett.*, **451**, 351–353.

191. Fulop, V., Phizackerley, R. P., Soltis, S. M., Clifton, I. J., Wakatsuki, S., Erman, J., Hajdu, J., and Edwards, S. L. (1994) Laue diffraction study on the structure of cytochrome c peroxidase compound i. *Structure*, **2**, 201–208.

192. Gustaffson, J. A., Rohndahl, L., and Bergman, J. (1979) Iodosylbenzene derivatives as oxygen donors in cytochrome P-450 catalyzed steroid hydroxylations. *Biochemistry*, **18**, 865–870.

193. Ortiz de Montellano, P. R. and Stearns, R. A. (1987) Timing of the radical recombination step in cytochrome P-450 catalysis with ring-strained probes. *J. Am. Chem. Soc.*, **109**, 3415–3420.

194. Hartford-Cross, C. F., Carmichael, A. B., Allan, F. K., England, P., Rouch, A. D., and Wong, L. L. (2000) Protein engineering of cytochrome P450cam (CYP101) for the oxidation of polycyclic aromatic hydrocarbons. *Prot. Engng.*, **13**, 121–128.

195. Guengerich, F. P. and MacDonald, T. L. (1990) Mechanisms of cytochrome P-450 catalysis. *FASEB J.*, **4**, 2453–2459.

196. Groves, J. T. (1985) Key elements of the chemistry of cytochrome P-450: The oxygen rebound mechanism. *J. Chem. Educ.*, **62**, 928.

197. Newcomb, M., Tadic, M. H. L., Putt, D. A., and Hollenberg, F. P. (1995) An incredibly fast apparent oxygen rebound rate constant for hydrocarbon hydroxylation by cytochrome P-450 enzymes. *J. Am. Chem. Soc.*, **177**, 3312–3313.

198. Newcomb, M., Biadatti, M. H. L. T., Chestney, D. L., Roberts, E. S., and Hollenberg, P. F. (1995) A nonsynchronous concerted mechanism for cytochrome P-450 catalyzed hydroxylation. *J. Am. Chem. Soc.*, **117**, 12085–12091.

199. Collman, J. P., Chien, A. S., Eberspacher, T. A., and Brauman, J. I. (1998) Multiple active oxidants in cytochrome P-450 model oxidations. *J. Am. Chem. Soc.*, **1202**, 11098–11100.

200. Newcomb, M. and Toy, P. H. (2000) Hypersensitive radical probes and the mechanisms of cytochrome P450-catalyzed hydroxylation reactions. *Acc. Chem. Res.*, **122**, 449–455.

201. Barton, D. H. R., Boar, R. B., and Widdowson, D. A. (1970) Phenol oxidation and biosynthesis. Part XX. The mechanism of erysodienone formation in vitro. *J. Chem. Soc, C*, pp. 1208–1213.

202. Zhao, B., et al. (2005) Binding of two flaviolin substrate molecules, oxidative coupling, and crystal structure of *Streptomyces coelicolor* A3(2) cytochrome P450 158A2. *J. Biol. Chem.*, **280**, 11599–11607.

203. Zerbe, K., et al. (2002) Crystal structure of oxyb, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. *J. Biol. Chem*, **277**, 47476–47485.

204. Pylypenko, O., Vatali, F., Zerbe, K., and Robinson, J. A. (2003) Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C-C coupling reaction during vancomycin biosynthesis. *J. Biol. Chem.*, **278**, 46727–46733.

205. Woithe, K., Geib, N., Li, D. B., Heck, M., Fournier-Rousset, S., Meyer, O., Vitali, F., Matoba, N., Abou-Hadeed, K., and Robinson, J. A. (2007) Oxidative phenol coupling reactions catalyzed by OxyB: A cytochrome P450 from the vancomycin producing organism. implications for vancomycin biosynthesis. *J. Am. Chem. Soc*, **129**, 6887–6895.

206. Geib, N., Woithe, K., Zerbe, K., Li, D. B., and Robinson, J. A. (2008) New insights into the first oxidative phenol coupling reaction during vancomycin biosynthesis. *Biorg. Med. Chem. Lett.*, **18**, 3081–3084.

207. Cane, D. E., Liang, T.-C., and Hasler, H. (1982) Polyether biosynthesis. 2. Origin of the oxygen atoms of monensin a. *J. Am. Chem. Soc.*, **104**, 7274–7281.

208. Kirkpatrick, P. (2000) *The biosynthesis of chloroeremomycin*. Ph.D. thesis, University of Cambridge, Cambridge, UK.

209. Holding, A. N. and Spencer, J. B. (2008) Investigation into the mechanism of phenolic couplings during the biosynthesis of glycopeptide antibiotics. *ChemBioChem*, **9**, 2209–2214.

210. Boger, D. L. and Yohannes, D. (1989) Total synthesis of K-13. *J. Org. Chem.*, **54**, 2498–2502.

211. Funa, N., Funabashi, M., Ohnishi, Y., and Horinouchi, S. (2005) Biosynthesis of hexahydroxyperlenequione melanin via oxidative aryl coupling by cytochrome P-450 in *Streptoomyces griseus*. *J. Bacteriol.*, **187**, 8149–8155.

212. Staunton, J. and Weissman, K. J. (2001) Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.*, **18**, 380–416.

213. Tuan, J. S., Weber, J. M., Staver, M. J., Leung, J. O., Donadio, S., and Katz, L. (1990) Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete-*Escherichia coli* cosmid. *Gene*, **90**, 21–29.

214. Bevitt, D. J., Corts, J., Haydock, S. F., and Leadlay, P. F. (1992) 6-deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea*. cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur. J. Biochem.*, **204**, 39–49.

215. Baker, P. M. and Bycroft, B. W. (1968) A convenient route to some naturally occurring hydroxynaphthaquinones. *Chem. Comm.*, pp. 71–72.

216. Bycroft, B. W., Cashyap, M. M., and Leung, T. K. (1974) A biomimetic synthesis of (+-)-scyltalone (3,6,8-triydroxytetralone). *L. Chem. Soc., Chem. Commun.*, pp. 443–444.

217. Ichinose, K., Ebizuka, Y., and Sankawa, U. (2001) Mechanistic studies on the biomimetic reduction of tetrahydroxynaphthalene, a key intermediate of melanin biosynthesis. *Chem. Pharm. Bull.*, **49**, 192–169.

218. Valls, N., Lopez-Canet, M., Vallribera, M., and Bonjoch, J. (2001) First total syntheses of aeruginosin 298-A and aeruginosin 298-B, based on a stereocontrolled route to the new amino acid 6-hydroxyoctahydroindole-2-carboxylic acid. *Chemistry*, **7**, 3446–3460.

219. Wright, J. N. and Akhtar, M. (1990) Studies on estrogen biosynthesis using radioactive and stable isotopes. *Steroids*, **55**, 142–151.

220. Shibata, N., Nakanishi, Y., Fukuoka, M., Yamanishi, M., Yasuoka, N., and Toraya, T. (2003) Structural Rationalization for the Lack of Stereospecificity in Coenzyme B12-dependent Diol Dehydratase. *J. Biol. Chem.*, **278**, 22717–22725.

221. Finking, R. and Marahiel, M. A. (2004) Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.*, **58**, 453–88.

222. Mootz, H. and Marahiel, M. (1997) The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.*, **179**, 6843–6850.

223. Tang, G.-L., Cheng, Y.-Q., and Shen, B. (2007) Chain initiation in the leinamycin-producing hybrid nonribosomal peptide/polyketide synthetase from *Streptomyces atroolivaceus* S-140: Discrete, monofunctional adenylation enzyme and peptidyl carrier protein that directly load D-alanine. *J. Biol. Chem.*, **282**, 20273–20282.

224. de Ferra, F., Rodriguez, F., Tortora, O., and Tosi, C. (1997) Engineering of peptide synthetases, key role of the thioesterase-like domain for efficient production of recombinant peptides. *J. Biol. Chem.*, **272**, 25304–25309.

225. Schneider, A., Stachelhous, T., and Marahiel, M. A. (1998) Targeted alteration of the substrate specificity of peptide synthetase by rational module swapping. *Mol. Gen. Gentet.*, **257**, 308–318.

226. Stachelhouse, T., Schneider, A., and Marahiel, M. A. (1995) Rational design of peptide antibiotics by target replacement of bacterial and fungal domains. *Science*, **269**, 69–72.

227. Fischback, M. A., Lai, J. R., Roche, E. D., Walsh, C. T., and Liu, D. R. (2007) Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc. Natl. Acad. Sci. USA*, **104**, 11951–11956.

228. Eppelmann, K., Stachelhous, T., and Marahiel, M. A. (2002) The exploitation of the selectivitiy-conferrring code of nonribosomal peptide synthetase for the rational design of novel peptide antibiotics. *Biochemistry*, **41**, 9718–9726.

229. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.*, **6**, 493–505.

230. Lombo, F., Velasco, A., Castro, A., de la Calle, F., Brana, A. F., Sanchez-Puelles, J. M., Mendez, C., and Salas, J. A. (2006) Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two *Streptomyces* species. *ChemBioChem*, **7**, 366–376.

231. O'Hare, H. M., Baerga-Ortiz, A., Popovic, B., Spencer, J. B., and Leadlay, P. F. (2006) High-throughput mutagenesis to evaluate models of stereochemical control in ketoreductase domains from the erythromycin polyketide synthase. *Chem. Biol.*, **13**, 287–296.

232. Lai, J. R., Fischbach, M. A., Liu, D. R., and Walsh, C. T. (2006) A protein interaction surface in nonribosmal peptide synthesis mapped by combinatorial mutagenesis and selection. *Proc. Natl. Acad. Sci. USA*, **103**, 5314–5319.

233. Lai, J. R., Fischbach, M. A., Liu, D. R., and Walsh, C. T. (2006) Localized protein interaction surfaces on the EntB carrier protein revealed by combinatorial mutagenesis and selection. *J. Am. Chem. Soc.*, **128**, 11002–11003.

234. Rusnak, F., Fraci, W. S., and Walsh, C. T. (1989) Subcloning, expression, and purification of the enterobactin biosynthetic enzyme 2,3-

dihydroxybezoate-AMP ligase: Demostration of enzymes-bound (2,3-dihydroxybenzoyl)adenylate product. *Biochemistry*, **28**, 6827–6835.

235. Dorrestein, P. C., et al. (2006) Activity screening of carrier domains within non-ribosomal peptide synthetases using complex substrate mixtures and large molecule mass spectrometry/. *Biochemistry*, **45**, 1537–1546.

236. Otten, L. G., Schaffer, M. L., Villiers, B. R. M., Stachelhaus, T., and Hollfelder, F. (2007) An optimized ATP/PP_i-exchange assay in 96-well format for screening of adenylation domains for applications in combinatorial biosynthesis. *Biotechnol. J*, **2**, 232–240.

237. Zou, Y. and Yin, J. (2008) Cu-free cycloaddition for identifying catalytic active adenylation domains of nonribosomal peptide synthetases by phage display. *Bioorg. Med. Chem. Lett.*, **18**, 5664–5667.

238. Neres, J., Wilson, D. J., Celia, L., Beck, B. J., and Aldrich, C. C. (2008) Aryl acid adenylating enzymes involved in siderophore biosynthesis: Fluorescence polarization assay, ligand specificity, discovery of non-nucleoside inhibitors via high-throughput screening. *Biochemistry*, **47**, 11735–11749.

239. McQuade, T. J., Shallop, A. D., Sheoran, A., DelProposto, J. E., Tsodikov, O. V., and Garneau-Tsodikova, S. (2009) A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Anal. Biochem.*, **386**, 244–250.

240. Reuter, K., Mofid, M. R., Marahiel, M. A., and Fincer, R. (1999) Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype 4'-phosphopantetheinyl transferase superfamily. *EMBO J.*, **18**, 6823–6831.

241. Glam, U., Dessoy, M. A., Schmidt, J., Wessjohann, L. A., and Heide, L. (2004) In vitro and In vivo production of new aminocoumarins by a combined biochemical, genetic and synthetic approach. *Chem. Biol.*, **11**, 173–183.

242. Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A., and Gerwick, W. H. (2004) Structure and biosynthesis of the jamaicaides, new mixed poly-ketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem. Biol.*, **11**, 817–833.

243. Kunst, F., Ogasawara, N., and Moser, I. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, **390**, 249–256.

244. Hofemeister, J., et al. (2004) Genetic analysis of the biosynthesis of non-ribosomal peptide- and polyketide-like antibiotics, iron uptake and biofilm formation by *Bacillus subtilis* A1/3. *Molecular Genetics and Genomics*, **272**, 363–378.

245. Gruenewald, S., Mootz, H. D., Stehmeier, P., and Stachelhaus, T. (2004) *In vivo* production of artificial nonribosomal peptide products in the heterogolous host *Escherichia coli*. *Appl. Environ. Microbiol.*, **70**, 2383–3291.

246. Lee, S. G. and Lipmann, F. (1975) Tyrocidine synthetase sytems. *Methods in Enzymol.*, **43**, 585–602.

247. Smith, G. P. and Petrenko, V. A. (1997) Phage display. *Chem. Rev.*, **97**, 391–410.

248. Baca, M., Scanlan, T. S., Stephenson, R. C., and Wells, J. A. (1997) Phage display of a catalytic antibody to optimize affinity for transition-state analog binding. *Proc. Natl. Acad. Sci. USA*, **94**, 10063–10068.

249. Avrantinis, S. K., Stafford, R. L., Tian, X., and Weiss, G. A. (2002) Dissecting the streptavidin-biotin interaction by phage-displayed shotgun scanning. *ChemBioChem*, **12**, 1229–1234.

250. Janda, K. D., Lo, L.-C., Lo, C.-H. L., Sim, M.-M., Wang, R., Wong, C.-H., and Lerner, R. A. (1997) Chemical selection for catalysis in combinatorial antibody libraries. *Science*, **275**, 945–948.

251. Xia, G., Chen, L., Sera, T., Fa, M., Schultz, P. G., and Romesberg, F. E. (2002) Directed evolution of novel polymerase activities: Mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci. USA*, **99**, 6597–6602.

252. Yin, J., Mills, J. H., and Schultz, P. G. (2004) A catalysis-based selection for peroxidase antibodies with increased activity. *J. Am. Chem. Soc.*, **126**, 3006–3007.

253. Agard, N. J., Baskin, J. M., Prescher, J. A., Lo, A., and Bertozzi, C. R. (2006) A comparative study of bioorthogonal reactions with azides. *ACS Chemical Biology*, **1**, 644–648.

254. Rostovtsev, V. V., Green, L. G., Fokin, V. V., and Sharpless, K. B. (2002) A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.*, **41**, 2596–2599.

255. Baskin, J. M., Prescher, J. A., Laughlin, S. T., Agard, N. J., Chang, P. V., Miller, I. A., Lo, A., Codelli, J. A., and Bertozzi, C. R. (2007) Copper-free click chemistry for dynamic *in vivo* imaging. *Proc. Natl. Acad. Sci. USA*, **104**, 16793–16797.

256. Geladopoulos, T. P., Sotiroudis, T. G., and Evangelopoulos, A. E. (1991) A malachite green colorimetric assay for protein phosphatase activity. *Anal. Biochem.*, **192**, 112 – 116.

257. Keating, T. A., Suo, Z., Ehmann, D. E., and Walsh, C. T. (2000) Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain. *Biochemistry*, **39**, 2297–2306.

258. Ehmann, D. E., Shaw-Reid, C. A., Losey, H. C., and Walsh, C. T. (2000) The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proc. Natl. Acad. Sci. USA*, **97**, 2509–2514.

259. Neilands, J. B. (1995) Siderophores: Structure and function of microbial iron transport compounds. *J. Biol. Chem.*, **270**, 26723–26726.

260. Kraemer, S. M. (2004) Iron oxide dissolution and solubility in the presence of siderophores. *Aquatic Sciences - Research Across Boundaries*, **66**, 3–18.

261. Miethke, M. and Marahiel, M. A. (2007) Siderophore-based iron acquisition and pathogen control. *Microbiol. Mol. Biol. Rev.*, **71**, 413–451.

262. Robins-Browne, R. M. and Prpic, J. K. (1985) Effects of iron and desferrioxamine on infections with *Yersinia enterocolitica. Infect. Immun.*, **47**, 774–779.

263. Griffiths, E. (1991) Iron and bacterial virulence: a brief overview. *BioMetals*, **4**, 7–13.

264. Bullen, J. J., Leigh, L. C., and Rogers, H. J. (1968) The effect of iron compounds on the virulence of *Escherichia coli* for guinea-pigs. *Immunology*, **15**, 581–588.

265. Pollack, J. R. and Neilands, J. B. (1970) Enterobactin, an iron transport compound from. *Biochem. Bioph. Res. Co.*, **38**, 989 – 992.

266. O'Bren, I. G. and Gibson, F. (1970) The structure of enterochelin and related 2,3-dihydroxy-*N*-benzoylserine conjugates from *Escherichia coli*. *Biochim Biophys Acta.*, **215**, 393–402.

267. Raymond, K. N., Dertz, E. A., and Kim, S. S. (2003) Enterobactin: An archetype for microbial iron transport. *Proc. Natl. Acad. Sci. USA*, **100**, 3584–3588.

268. Harris, W. R., Carrano, C. J., Cooper, S. R., Sofen, S. R., Avdeef, A. E., McArdle, J. V., and Raymond, K. N. (1979) Coordination chemistry of microbial iron transport compounds. 19. stability constants and electrochemical behavior of ferric enterobactin and model complexes. *J. Am. Chem. Soc.*, **101**, 6097–6104.

269. Shaw-Reid, C. A., Kelleher, N. L., Losey, H. C., Gehring, A. M., Berg, C., and Walsh, C. T. (1999) Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization. *Chem. Biol.*, **6**, 385 – 400.

270. Zhou, G.-H., Gotou, M., Kajiyama, T., and Kambara, H. (2005) MultiplexSNP typing by bioluminometric assay coupled with terminator incorporation (BATI). *Nucleic Acids Res.*, **33**, e333.

271. Baggett, B., Roy, R., Momen, S., Morgan, S., Tisi, L., Morse, D., and Gllies, R. J. (2004) Thermostability of firefly luciferases affects efficiency of detection by in vivo bioluminescence. *Mol. Imaging*, **3**, 324–332.

272. Law, G. H. E., Gandelman, O. A., Tisi, L. C., Lowe, C. R., and Murray, J. A. H. (2006) Mutagenesis of solvent-exposed amino acids in *Photoinus pyralis*

luciferase improves thermostability and pH-tolerance. *Biochem. J.*, **397**, 305–312.

273. White, P. J., Squirrell, D. J., Arnaud, P., Lowe, C. R., and Murray, J. A. (1996) Improved thermostability of the North American firefly luciferase: saturation mutagenesis at position 354. *Biochem. J.*, **319**, 343–350.

274. Seliger, H. H. and McElroy, W. D. (1964) The colors of firefly bioluminescence: Enzyme configuration and species specificity. *Proc. Natl. Acad. Sci. USA*, **52**, 75–81.

275. Seliger, H. H. and McElroy, W. D. (1960) Spectral emission and quantum yield of firefly bioluminescence. *Arch. Biochem. Biophys.*, **88**, 136–41.

276. Schall, O. F., Suzuki, I., Murray, C. L., Gordon, J. I., and Gokel, G. W. (1998) Characterization of Acyl Adenyl Anhydrides: Differences in the Hydrolytic Rates of Fatty Acyl-AMP and Aminoacyl-AMP Derivatives. *J. Org. Chem.*, **63**, 8661–8667.

277. Marshall, C. G., Burkart, M. D., Keating, T. A., , and Walsh, C. T. (2001) Heterocycle formation in vibriobactin biosynthesis: Alternative substrate utilization and identification of a condensed intermediate. *Biochemistry*, **40**, 10655–10663.

278. Pavela-Vrancic, M., Dieckmann, R., Dohren, H. V., and Kleinkauf, H. (1999) Editing of non-cognate aminoacyl adenylates by peptide synthetases. *Biochem. J.*, **342**, 715–719.

279. Kleinkauf, H. and vonDohren, H. (1997) Applications of peptide synthetases in synthesis of peptides analogues. *Acta. Biochim. Pol.*, **44**, 839–847.

280. Drake, E. J., Nicolai, D. A., and Gulick, A. M. (2006) Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chem. Biol.*, **13**, 409–419.

281. Ansari, M. Z., Yadav, G., Gokhale, R. S., and Mohanty, D. (2004) NRPS-PKS: a knowledge-based resourse for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.*, **32**, 405–513.

282. Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A., and Holak, T. A. (2000) Solution structure of PCP, a prototype for peptidyl carrier domains of modular peptide synthetases. *Structure*, **8**, 407–418.

283. Morris, S. A., Revill, W. P., Staunton, J., and Leadlay, P. F. (1993) Purification and separation of *holo*- and *apo*-forms of *Saccharopolyspora erythraea* acyl-carrier protein released from recombinant *Escherichia coli* by freezing and thawing. *Biochem. J.*, **294**, 521–527.

284. Kelley, L. A. and Sternberg, M. J. E. (2009) Protein structure prediction on the web: a case study using the *p*hyre server. *Nature Protocols*, **4**, 363–371.

285. Jürgen J. May, N. K., Marahiel, M. A., and Stubbs., M. T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. USA*, **99**, 12120–12125.

286. Rombach, P. (2009) *Studies of the non-ribosomal peptide synthetase (NRPS) of teicoplanin biosynthesis*. Master's thesis, Albert-Ludwigs-University, Freiburg, Germany.

287. Truman, A. W. (2009) (personal communication).

288. Roche, E. D. and Walsh, C. T. (2003) Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry*, **42**, 1334–1344.

289. Ehmann, D. E., Trauger, J. W., Stachelhaus, T., and Walsh, C. T. (2000) Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem. Biol.*, **7**, 765–772.

290. Kupczyk-Subotkowska, L., Saunders, W. H., and Shine, H. J. (1998) Claisen rearrangement of allyl phenyl ether: heavy-atom kinetic isotope effects and bond orders in the transition structure. *J. Am. Chem. Soc.*, **110**, 7153–7159.

291. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molcular cloning: a laboratory manual*. Cold Spring Harbor Laboratory.

292. Bradford, M. M. (1976) A rapid and sensitive method for the quantisation of microgram quantities of protein utilising the principle of protein-dye binding. *Anal. Biochem.*, **72**, 248–54.